DOCUMENT RESUME

ED 398 900 IR 018 081

AUTHOR	Swaak, Janine; And Others
TITLE	Support for Simulation-Based Learning; The Effects of Model Progression and Assignments on Learning about Oscillatory Motion.
INSTITUTION	Twente Univ., Enschede (Netherlands). Faculty of Educational Science and Technology.
REPORT NO	ISBN-90-365-08-436
PUB DATE	Mar 96
NOTE	41p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Achievement Gains; Assignments; College Students; *Computer Simulation; *Discovery Learning; Foreign Countries; Higher Education; Knowledge Level; *Models; *Physics; Predictor Variables; Pretests Posttests; Science Instruction; Testing; Undergraduate Students
IDENTIFIERS	Cognitive Load; Netherlands; Transaction Log Analysis

ABSTRACT

In this study, learners worked with a simulation of harmonic oscillation. Two supportive measures were introduced: model progression and assignments. In model progression, the model underlying the simulation is not offered in its full complexity from the start, but variables are gradually introduced. Assignments are small exercises that help the learner to define goals during discovery learning. Subjects were 63 first year students in physics at a university level. Three experimental conditions were created: in one, learners received a computer simulation together with model progression and assignments; in the second, only model progression was added; and in the third (control) condition, neither model progression nor assignments were available. For measuring learning results, three types of tests were used: a definitional test, measuring students' factual knowledge of the domain; an intuitive test, measuring students' insight; and a test measuring students' propositional knowledge. The definitional and intuitive test were used as pre- and post-test, and the propositional test was only used as a post-test. All student actions were logged and several aspects of cognitive load were measured with an electronic questionnaire. Results showed a small gain in definitional knowledge for all three conditions. The gain in intuitive knowledge was considerable, and differed across the experimental groups in favor of the conditions with assignments and/or model progression compared to the control condition. The cognitive load measure indicated that operating the environment did not interfere with the learning process. (Contains 50 references.) (Author/AEF)

universiteit twente

onderzoekcentrum

toegepaste

onderwijskunde

Janine Swaak
Wouter van Joolingen
Ton de Jong

Support for simulation-based learning; the effects of model progression and assignments on learning about oscillatory motion

OCTO

2

Support for simulation-based learning; the effects of model progression and assignments on learning about oscillatory motion

March 1996

JANINE SWAAK, WOUTER VAN JOOLINGEN, &
TON DE JONG

Faculty of Educational Science and Technology
Centre for Applied Research on Education
Department of Instructional Technology
University of Twente
The Netherlands

Janine Swaak, Wouter van Joolingen, & Ton de Jong

Support for simulation-based learning; the effects of model progression and assignments on learning about oscillatory motion
Published by Center for Applied Research on Education (OCTO), March 1996
Printed in the Netherlands by CopyPrint 2000, Enschede
ISBN: 90 365 08 436

University of Twente
PO-Box 217
7500 AE Enschede
Tel.: +31 53 4893588
Fax.: +31 53 4893023

Table of Contents

Support for simulation-based learning; the effects of model progression and assignments on learning about oscillatory motion

JANINE SWAAK, WOUTER VAN JOOLINGEN, & TON DE JONG

Faculty of Educational Science and Technology
Centre for Applied Research on Education
Department of Instructional Technology
University of Twente
The Netherlands

Abstract

Discovery learning with computer simulations is generally seen as a promising way of learning and instruction. Studies have shown that in many cases discovery learning with computer simulation leads to higher performance compared to more expository ways of teaching, but this advantage could not always be found. One of the possible reasons for not finding better results with discovery environments is that learners experience problems with one or more of the aspects of discovery learning. Solutions can be found in combining simulations with support for the discovery process. In the current study learners worked with a simulation from a physics domain (harmonic oscillation). Two supportive measures were introduced: model progression and assignments. In model progression the model underlying the simulation is not offered in its full complexity from the start, but variables are gradually introduced. Assignments are small exercises that help the learner to define goals during discovery learning. Subjects were 63 students in physics from a first year university level. Three experimental conditions were created, in one condition learners received a computer simulation together with model progression and assignments, in the second one only model progression was added, and in the third (control) condition, neither model progression nor assignments were available. For measuring learning results three types of tests were used. A 'definitional test', measuring students' factual knowledge of the domain, an 'intuitive test', called the WHAT-IF test, that was meant to measure the students' insight in the domain, and a test measuring the students' propositional knowledge. The definitional and intuitive test were used as pre- and post test, and the propositional test was only used as a post-test. For assessing the learning process all student actions were logged and several aspects of cognitive load were measured with an electronic questionnaire. The results showed a small gain in definitional knowledge for all three conditions. The gain in intuitive knowledge was considerable, and differed across the experimental groups in favour of the conditions with assignments and/or model progression compared to the control condition. The cognitive load measure indicated that operating the environment did not interfere with the learning process.

Address correspondence to: Janine Swaak, Faculty of Educational Science and Technology, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands. Email: swaak@edte.utwente.nl. This project was sponsored by the Institute for Educational Research in the Netherlands (SVO) under grant 95728. We would like to express our thanks to Dr. ir. Imme de Bruijn of the Faculty of Applied Physics for his advice and for his cooperation in recommending his students to participate in the experiment, to the students who were willing to do so, and to Elwin Savelsbergh for scoring the hypotheses lists.

1. INTRODUCTION

Discovery learning is a way of learning that offers opportunities for learners to engage in a process of active knowledge construction (e.g. Bruner, 1961; de Jong 1991; Shulman & Keisler, 1966; White, 1984). Computer simulation is one of the types of environment that are suited for discovery learning. In computer simulation learners have to infer properties of the model that underlie the simulation from varying the values of input variables and observing the values of output variables. De Jong and Van Joolingen (1996) present an overview of discovery learning with computer simulations. They list a number of studies that have compared simulation-based discovery learning with other modes of instruction (e.g. Carlsen & Andre, 1992; Chambers et al., 1994; Choi & Gennaro, 1987; De Jong, De Hoog, & De Vries, 1993; Grimes & Willey, 1990; Lewis, Stern, & Linn, 1993; Rieber, 1990; Rieber, Boyce, & Assad, 1990; Rieber & Parmley, in press; Rivers & Vockell, 1987; Shute & Glaser, 1990; White, 1993). The overall conclusion of these studies is that simulation based learning quite often is more effective than for example expository teaching, but still in a large number of cases the effectiveness is equal to expository teaching. According to De Jong & van Joolingen (1996) there are two main reasons why discovery learning is not more effective in some cases. The first reason is that learners may experience problems in the discovery learning process. The second one is that discovery learning with simulations is supposed to lead to a more 'intuitive', deeply rooted form of knowledge that is not measured adequately by the types of tests quite often used in the studies cited. A third related reason, added here, is that in simulation-based discovery environments learners may experience high cognitive load which may hinder learning effects to come about.

In the current study we have concentrated on one of the problems that students may have with discovery learning: *regulation* of the discovery process. In a computer simulation on a physics topic (harmonic oscillations) we introduced two instructional measures or 'cognitive tools' (Lajoie, 1993) that aimed at supporting the learner in regulating the process. The learning process and the result of learners working with a simulation environment that included these tools (model progression and assignments) were compared with the process and results of learners learning with a plain simulation. For measuring the results, we used a rather traditional 'definitional' test, asking learners about facts, and a test that could measure knowledge with a more 'intuitive' character. To find out whether the instructional measures lead to a change in cognitive load, we measured several aspects of cognitive load as experienced by the learners.

2. REGULATION IN DISCOVERY LEARNING

2.1 Discovery learning processes

Scientific discovery learning is a learning method of a complicated nature, putting a high responsibility for the learning process in the hands of the learner. Studies into discovery learning processes have identified a large number of subprocesses. Friedler, Nachmias, and Linn (1990), for example, say that scientific reasoning comprises the abilities to (a) define a scientific problem; (b) state a hypothesis; (c) design an experiment; (d) observe, collect, analyse, and interpret data; (e) apply the results; and (f) make predictions on the basis of the results." (p. 173). Njoo and de Jong (1993) make a main distinction between *transformative processes* (processes that directly yield knowledge) and *regulative processes* (processes that are necessary to manage the discovery process). For transformative processes they further distinguish: *Analysis* which is the process of identifying and relating variables in the model and indicating general properties of the model. *Hypotheses generation* which is the formulation of a relation between one or more variables (input and output) and parameters in the simulation model. A hypothesis is stated with the intention to test it. *Testing*, which refers to those activities that are necessary for furnishing data on which the learner expects to be able to accept or refute an hypothesis, or to create an hypothesis. Testing includes the processes 'designing an experiment', 'making predictions', and 'data interpretation'. And, finally, *evaluation*, in which results are put into a more general context. Regulative processes are subdivided into: *Planning*, which can take place at the level of the complete discovery process, or at the level of one of the transformative processes indicated above. *Verifying*, which is checking the correctness of actions and results at a conceptual level. And, finally, *monitoring*, in which the learner observes and keeps track of his/her own study process.

Both transformative and regulative learning processes can be problematic for a learner. For instance, a learner may have trouble stating a hypothesis, designing an experiment to test it or to interpret the results of the experiment (Njoo & De Jong, 1993). Problems with regulation in discovery learning are sometimes referred to as 'floundering' (Goodyear et al., 1991). Glaser, Schauble, Raghavan, and Zeitz (1992) analysed learners' behaviour in three different simulation environments and found that, compared to successful learners, unsuccessful ones used a more random strategy, were less systematic, concentrated on local decisions, and had trouble monitoring what they had done. Similar findings are reported by Lavoie and Good (1988), Shute & Glaser (1990), Veenman and Elshout (1995) and Klahr, Dunbar, and Fay (1991). For improving the effectiveness of discovery learning, it can be considered to help learners in their regulative processes by providing them with additional support next to the simulation.

2.2 Support for regulation

In De Jong and Van Joolingen (1996) many support measures that can be combined with computer simulation are listed. These support measures are intended to help the learner to succeed in the discovery process. De Jong and Van Joolingen (1996) mention support measures that help to gain access to prior knowledge, to assist in the generation of hypotheses, for the design of experiments, for making predictions, and for regulative processes. In the area of regulative processes they mention *planning support* and *model progression.*

Planning support takes away decisions from learners and in this way helps them in managing the learning process. Planning support This support for planning can be given in different ways. Already quite early in the use of simulations for scientific discovery learning, Showalter (1970) recommended to use *questions* as a way to guide the learner through the discovery process. His questions focused the learners attention to specific aspects of the simulation. White (1984) helped learners to set goals in a simulation of Newtonian mechanics by introducing *games*. Games, as White uses them, ask learners to reach a specific state of the simulation (e.g. to get a spaceship in the simulation around a corner without crashing into any walls, p. 78). In an experiment White found that learners who learned with a simulation that contained games, outperformed learners who worked with the pure simulation on a test of qualitative problems (asking questions of the form "What would happen if ..?", p. 81)). Also, in the ThinkerTools environment (White, 1993) games are used in a similar context as in White (1984). Along a similar line, in the SMISLE[1] learning environments regulative support is, given in the form of assignments (de Jong et al., 1994; De Jong & Van Joolingen, 1995). The idea of *assignments* is to provide the learner with short-term learning goals, such as discovering a part of the domain, or applying knowledge that has just been discovered.

The general idea of *model progression* is to keep the simulation environment manageable by not introducing too many new ideas at a time. White and Frederiksen (1990), from whom the idea of model progression stems, distinguish three kinds of ways to do this:

- *Simple to complex model progression*, that is starting with a simplified version of the model in which only a few variables are present, and gradually expanding the set of variables, by offering more and more complex versions of the model;
- *Changing the order of the model*, in changing the order of the model, models of increasing *precision* (see Van Joolingen and De Jong, 1993) are put in sequence. Typically, one will start with a qualitative model, in which only statements of the order true/false are made and end with fully specified quantitative models;
- *Changing perspective on the model*, often, models can be described from different viewpoints. For instance, models in physics can often be described from the viewpoint of state variables (e.g. positions and velocities, but also voltage and current) and from the viewpoint of energy flow. In this type of model progression, the word *progression* is used not correctly, as there is no sense of direction in switching between views.

In the present study we created a simulation learning environment that contained both assignments and model progression. Section 5.1 presents how these measures have been operationalised in our study. The way model progression is operationalised in the discovery environment of the present study conforms to the first type, the progression from a simple to more complex models.

[1] The discovery environment of the present study is created with the SMISLE authoring environment. The SMISLE project was partly funded by the European Commission as project D2007 in its Telematics programme. The SMISLE project is currently being continued in the SERVIVE project (project ET 1020).

3. ASSESSING THE PRODUCTS AND PROCESSES OF DISCOVERY LEARNING

In order to assess the effectiveness of the support measures given to the learner we need measures for both the *product* of the discovery process, i.e., the knowledge gained by learners as a result of working with the simulation environment, and the *process* of interaction, i.e., the way learners interact with the simulation. In addition, since discovery learning is a highly demanding way of learning, we wanted to measure the *cognitive load* that learners experience in the learning process. As product measures, we choose for knowledge tests, however, we argue that 'traditional' knowledge tests are not necessarily the best means of assessing the results of simulation-based discovery learning, because such tests neglect intuitive properties of knowledge. Therefore, next to a definitional knowledge test, we developed a test which intended to tap the intuitive knowledge acquired in interaction with simulations. For assessing the learning process we used logfiles, and for assessing the *cognitive load* of students we developed an on-line measuring device. In the next two sections, we elaborate upon the rationale behind the newly developed assessment measures.

3.1 Intuitive knowledge

An important premise of this work is that discovery learning with simulations may not lead to a kind of knowledge that can easily be measured by the types of tests normally used in the studies on the effects of learning methods. Instead, we think that the interaction with simulations may result in a type of knowledge which we can label as 'intuitive knowledge'. When literature on intuitive related knowledge is reviewed, we find that many authors have written about it, but that only few tried to assess intuitive knowledge (Swaak, 1995). Despite the underrepresentation of serious efforts to assess intuitive knowledge, literature on intuitive knowledge (e.g., Fischbein, 1987; diSessa, 1993) together with research on interacting with complex simulation systems (e.g., Berry & Broadbent, 1988; Broadbent, FitzGerald, & Broadbent, 1986; Hayes & Broadbent, 1988; Leutner, 1993) have provided us with at least three, more or less stable, notions on intuitive knowledge (for a complete review see Swaak, 1995).

The first is that the intuitive quality of knowledge is acquired after 'using' knowledge in perceptually rich, dynamic situations (see also Fischbein, 1987). Important assumptions are that the learning environments of the present study can be well described as rich, dynamic environments, and that the learners are actively engaged in the learning process. We infer that if knowledge is 'used' in rich contexts, in which perceptions play a critical role, *experiential* learning processes are elicited, and that those experiential learning processes lead to intuitive knowledge.

A second notion is the intuitive quality makes the knowledge difficult to verbalise. An important hypothesis is, indeed, that in the interaction with a simulation environment learners are invited to follow a learning mode – an implicit, experiential learning mode –which leads to knowledge that is hard to verbalise. Many studies involving the control of complex simulation systems suggest that there is more to knowledge than only the verbalisable part (e.g., Berry & Broadbent, 1984; 1988; 1990; Broadbent, FitzGerald, & Broadbent, 1986; Hayes & Broadbent, 1988; Leutner, 1993).

The third observation is that the access in memory of knowledge with an intuitive quality is different from the access in memory of knowledge without this quality (see also Fischbein, 1987). We speculate the differential access exists next to differences in verbalisation. We hypothesise that the experiential learning mechanisms tune the knowledge and give it an intuitive quality. However difficult to verbalise, the intuitive quality causes the access to the knowledge in memory to be more efficient. Van Berkum and De Jong (1991) illustrate that examples in the domain of chess (e.g., Chase & Simon, 1973) suggest that the phenomenon of knowledge tuning is not limited to operational knowledge (see Anderson, 1987, for this opinion), but that it also extends to more conceptual knowledge (see also Fischbein, 1987). In the authors' words,'...Chess masters mainly differ from novices in their 'direct perception' of complex, meaningful chess patterns, and much less in their basic problem solving procedures.' (1991, p. 313).

To summarise, low verbalisability, 'rich' situations, and speed are the three observations most frequently found in relation to intuitive quality of knowledge. A certain coherence can be stated between these findings. Several questions remain unanswered, however. So far, there is no agreement on the exact nature of the processes involved in the acquisition of the intuitive knowledge. Even more remains unclear about the precise presentation of intuitive conceptual knowledge. However, most researchers (e.g., Broadbent and colleagues, Brown, Van Berkum & De Jong) agree that, whatever the exact nature of the processes involved in the *acquisition* and whatever the precise presentation of intuitive conceptual knowledge, the processes involved in the *manifestation* of the intuitive quality of knowledge can be described as a 'the quick perception of meaningful situations'. As will become clear, this description of intuitive quality is central to the intuitive tests we developed. The test, that is called the 'WHAT-IF test' is described in Section 5.2.3.

3.2 Cognitive load

As is outlined above, learning involved in discovery environments, such as simulations, is supposed to be based on learning processes that are qualitatively different from the learning processes in more traditional instructional situations. In more traditional instruction an emphasis on the mere acquisition of knowledge is present, whereas in discovery environments learning processes that, first of all, make sense of the information presented – *transformative processes* (such as, for example, hypothesis generation)– are of uttermost importance (see section 2). Another aspect is that discovery environment give learners much freedom and thereby require learners to regulate their own study process – here, *regulative processes* are important– (see also Section 2). Both of these characteristics of discovery learning are assumed to be highly demanding. In similar vein, Cates (1992) talks about "the cognitive demands of the 'information age'(....), and about developers of hypermedia and multimedia instructional programs arguing for learner 'empowerment' and learner control" (1992, p. 1). The author makes the point that not only more information is offered, but also that the information changes more frequently, and that people have more freedom to learn from this information. Cates continues that "it also offers opportunities for them to experience substantial cognitive overload (...)" (p. 1).

We subscribe Cates' view, and, furthermore, we found that in the evaluation of instructional computer simulations 'unexpected' results are several times ascribed to the cognitive load or overload of the learners involved. Among the researchers, who refer to this cognitive load phenomenon, are Hussy and Granzgow (1987, cited in Leutner, 1993). Their studies indicate among others that an increase in system-transparency was accompanied by an increase in problem-solving achievement. System-transparency was enhanced by giving rich information about system variables inherent in the system. In one study where the information, instead of being eliminated, remained on the screen, the achievement was hindered. The researchers assume, based on the outcome of the (poor) task performance that this effect is the result of information-overload.

Other researchers who implicitly mention excessive cognitive load as a possible reason why, for example, extra support supplied next to a simulation environment does not work are Shute (1991), De Jong et al. (1993) and Njoo (1994). Shute calls it "disruption of the compilation process". She explains that when people are interacting with a simulation and engaged in problem solving, the decision to use on-line tools distracts from, and thus according to Shute, disrupts the compilation process. In the same vein, De Jong et al. (1993) infer from the lower than expected scores on the post-tests, that the extra support given, might have distracted the learners form the main task, the simulation itself. De Jong et al. observed that the support tools formed indeed an extra task for the learners. Njoo (1994) found (study 3) that of two groups working with the same simulation program, the group with the "highest level of support" had the lowest post-test scores. Njoo continues to suggest the possible explanation that "the instructional support measures had placed an additional cognitive load on subjects' working memory" (p. 126).

As can be seen cognitive load is an interesting and relevant concept in the context of learning with computer simulations. Cognitive load is determined by the rather difficult transformative learning processes of discovery learning, the extensive regulative aspects of it, and the complexity of the learning environment that may include extra instructional support. Not much attention, however, is paid to assessing cognitive load in discovery environments. Therefore, we developed a cognitive load scale that was fitted to the learning environments we evaluated, and that was able to tap several aspects of cognitive load. This scale, that we called the S.O.S. scale, is described in Section 5.2.3

4. Hypotheses

The main ideas put forward so far can be summarised as follows: positive effects of discovery learning may be expected if the regulation of the discovery process is adequately supported and if the right assessment techniques are applied to measure the effectiveness of discovery learning. We hypothesise that both the assignments and model progression assist learners in their discovery process. As a consequence, the cognitive load of learners will be reduced thereby leaving sufficient cognitive recourses for learners to actually learn. However, since we also have indications that extra support may increase cognitive load, cognitive load will be measured on-line, in order to have a control. The knowledge gained, will to a large extent, consist of knowledge that is hard to verbalise, and that is best captured by the intuitive WHAT-IF test. The next section describes the empirical study in which our ideas were tested.

5. THE PRESENT STUDY

In this report an evaluation is presented on adding assignments and model progression to a simulation-based environment for discovery learning on the effectiveness of learning. The learning environment evaluated is called SETCOM: System for Exploratory Teaching a Conceptual model of Oscillatory Motion. The subject domain of this environment is oscillatory motion. SETCOM was designed for first year university level students of physics or technical sciences. In its complete form SETCOM employs simple to complex model progression. The learning environment starts with a simple model, of a mass suspended from a spring, and adds two levels of increasing complexity by introducing subsequently a damper and an external force. At each model progression level a series of assignments is available for the students. Apart from assignments and model progression SETCOM also includes a number of explanations as a means of instructional support.

5.1 The learning environment

5.1.1 Domain

SETCOM addresses one-dimensional oscillatory motion. Oscillatory motion is a subject taught to, for example, first-year engineering and physics students. In the practice of engineering, the characteristics of oscillations play an important role in design, since unintended oscillations may severely affect the behaviour of systems. Example of designs in which oscillatory motion is important are shock damping devices in cars, wings of aeroplanes, loud speakers and robots, but also designing earthquake resistant buildings requires deep knowledge of oscillations. Three kinds of oscillatory motion are addressed in SETCOM:

- free oscillatory motion without friction;
- damped motion;
- forced oscillatory motion.

The type of motion that occurs is dependent on the presence of friction and/or an external force. In the case that both are absent, the motion is free, if only friction is present, we have damped oscillation, if both are present, there is forced oscillatory motion.

In *free oscillation* the system shows an undisturbed periodic behaviour, with a frequency dependent on the force constant and the mass of the system. This frequency is called the eigenfrequency of the system.

In the case of *damped oscillation*, different modes of damping exist: subcritical damping, critical damping, and supercritical damping. In the case of a (relatively) small friction coefficient, damping is subcritical. This means that oscillation does occur, but slowly dies out. In the case of supercritical damping, for large friction coefficients, no oscillation occurs at all, the system relaxes to equilibrium without oscillation. The boundary case between these two cases is called critical damping. Here, also no oscillation occurs, but the system quickly relaxes to an equilibrium. This situation only occurs for one specific value of the friction coefficient, for a given mass and force coefficient of the system.

A central place in the analysis of damped oscillation (with or without an external force) is taken by the so called characteristic equation, which is derived from the equation of motion. The two roots are complex numbers of which the real part is associated with the time the motion takes to die out, and the imaginary part is connected to the frequency of oscillation. In the case of free oscillation, the roots are purely imaginary, i.e., their real parts are zero, meaning that the oscillation will not die out. In the case of subcritical damping, both the real and imaginary parts of the roots are nonzero, in the case of critical and supercritical damping, the imaginary parts of the roots are zero. Moreover, in the critical case the two roots are equal to each other.

In *forced motion*, an external periodic force or an external periodic motion is applied to the system. This external force interacts with the autonomous behaviour of the system, meaning that the solution of the differential equation is the sum of the undisturbed (damped motion), and a component stemming from the external force, called the homogeneous and particular solutions respectively. The homogeneous solution dies out in the same manner as for the damped oscillation, including sub- and supercritical behaviour. The particular solution is a periodic motion with the same frequency as the external force, but shifted in phase with respect to the force behaviour. This phase shift as well as the amplitude of the particular solution depend on the eigenfrequency of the unforced oscillator (if existent) and the friction coefficient.

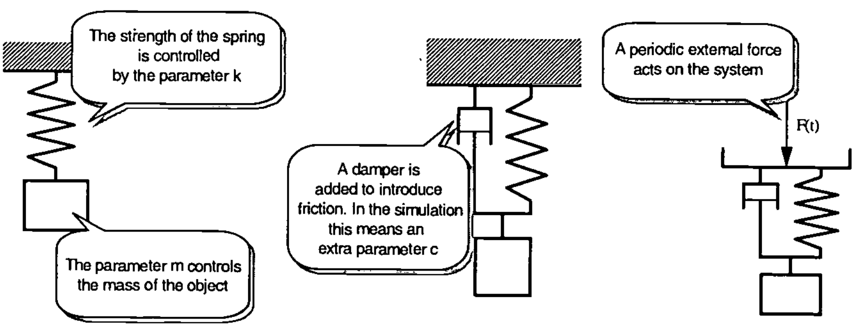

Figure 1. *Three types of oscillatory motion, illustrated by a mass suspended from a spring. From left to right: free oscillation, damped oscillation, and forced motion.*

When the frequency of the force approaches the eigenfrequency of the system, the amplitude of the system grows strongly and the phase shift approaches 90 degrees. This phenomenon is called resonance. When designing systems, it is of crucial importance to prevent situations in which resonance may occur, since this can result in unexpected system behaviour and damage.
The three types of oscillatory motion introduced in SETCOM are depicted in Figure 1

5.1.2 The simulation environment

In this section the full version of SETCOM is described. The other two versions used in the experiment were obtained by omitting features from this version (see Section 5.2). SETCOM contains three simulations of oscillating systems. The simulated systems are the ones displayed in Figure 1. Each simulation model corresponds to a level of model progression going from free oscillation, through damped oscillation to forced motion (see Section 5.1.1). Each of the simulations is a dynamic simulation which allows the learner to control a number of input variables and watch the behaviour of the oscillating system as is expressed in a graph and in numerical output. An example simulation window is displayed in Figure 2. The simulation window in Figure 2 corresponds to the second level of model progression (damped oscillation).

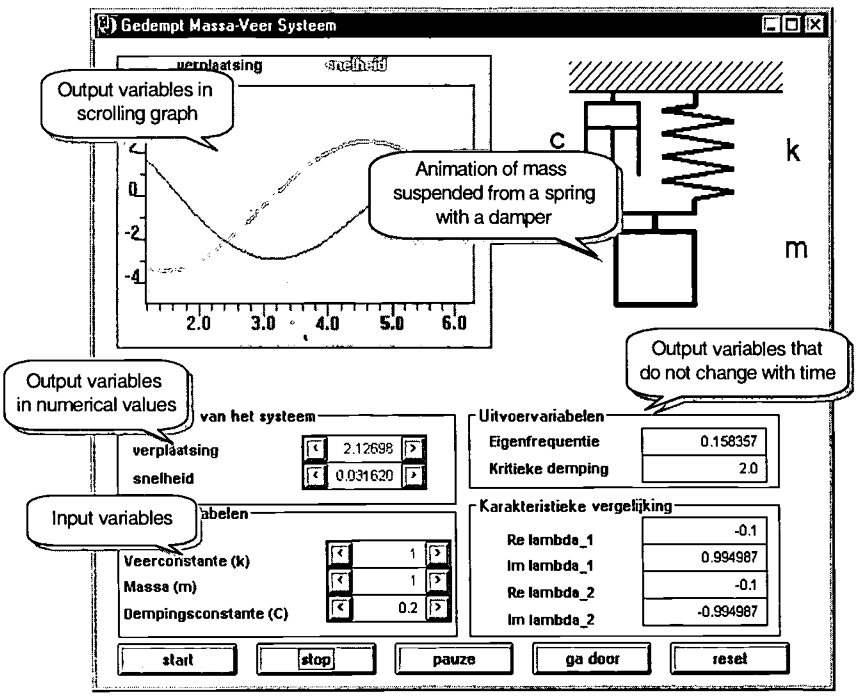

Figure 2. *A prototype simulation window corresponding to the second level of model progression.*

Model progression

The three simulations present in SETCOM define a simple to complex model progression. The number and kind of input variables that can be controlled and output variables that can be observed increases with every level in this progression. At the most simple level the learner can control only two input variables, at the most complex level, the learner can control five variables. In all cases the learner can also control the initial state of the system. In Figure 2 the simulation window corresponds with the second level of model progression. In the simulation window that belongs to the first most simple level the input variable "c" is not available and in the animation no damper is displayed. In the simulation window corresponding to the third and most complex level the input variable "F" is added, and in the picture an external force is shown (see also Figure 1). Table 1 provides an overview of the variables present at each model progression level. Apart from model progression, SETCOM also includes assignments and explanations.

Assignments

On each level of model progression, a number of assignments guide the learner in the exploration of the model behind the progression level. The core of the collection of assignments offered to the learner is formed by the *investigation assignments*. These assignments prompt the learner to start an inquiry of the relations between two variables given. The set of investigation assignments was designed in such a way that for each relevant relation in the simulation model at each model progression level, one investigation assignment was available. In Figure 3 the assignment window, an example of an investigation assignment, and the associated answer window used in SETCOM are displayed.

When learners go through all of these assignments, they have met all the relevant relations in the domain. In Table 1 an overview of the investigation assignments in SETCOM is given. From this table it can be seen that not simply *all* relations between any input variable and any output variable are represented in investigation assignments. Sometimes such relations are just non-existent, sometimes they were too complex to catch them in a single assignment.

Two decisions for the choice of investigation assignments need extra explanation. First, there are no investigation assignments concerning either of the two state variables x and v. The reason for this is that these variables depend on time and there is no direct relationship between any of the inputs and the state variables. In oscillation theory, only the *global* behaviour of oscillating systems, expressed in output variables like the frequency and the amplitude are of interest for understanding. Second, the roots of the characteristic equation appear in the investigation assignments of the second model progression level, whereas they are introduced on the first level. The reason for this is that these output variables only become of real interest at the second level of model progression, when friction is introduced, but that they are already included on the first level, to obtain a consistent look for all model progression levels. The roots of the characteristic equation do appear in a specification assignment on the first level.

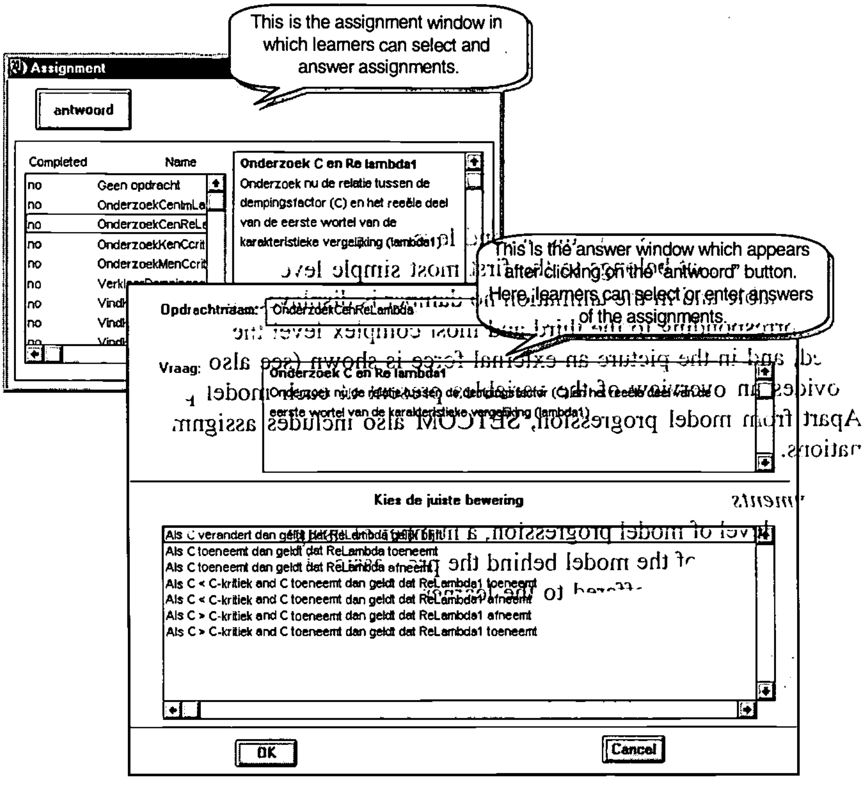

Figure 3. *The assignment window, an example investigation assignment and the answer window of the assignment.*

The set of assignments is completed by three other types of assignments: optimisation assignments, specification assignments, and one explicitation assignment. The optimisation and specification assignments were included to allow learners to test themselves in a game-like situation.

In the optimisation assignments, the learners control one input variable and are given a certain goal, like: try to reach a situation of critical damping (i.e. when $C = C_{crit}$). Usually, in such a situation, the variables involved, like C_{crit} can be manipulated indirectly, using one of the relations that can be found in one of the investigation assignments. During the activity of an optimisation assignment, some *constraints* are active. Once such a constraint is violated, the simulation is stopped and the learner is informed that the constraint has been violated.

Specification assignments ask the learner to predict the value of a variable in a given situation. The situation is presented in the simulation interface, and the learner can type in the prediction in the answer window.

Table 1. *Overview of the model progression levels and investigation assignments in SETCOM. Input variables are printed in bold type.*

Model progression level	Variables introduced	Investigation assignments
simple harmonic	velocity (v) position (x) **force constant** (k) **mass** (m) frequency (f) roots of the characteristic equation ($Im\ \lambda_1$, $Im\ \lambda_2$, $Re\ \lambda_1$, $Re\ \lambda_2$)	investigate the relation between ... k, f m, f
damped harmonic	**damping constant** (C) critical damping (C_{crit})	k, C_{crit} m, C_{crit} $C, Im\ \lambda_{1,2}$ $C, Re\ \lambda_{1,2}$
forced oscillation	**force amplitude** (F_0) **force frequency** (ω_1) phase shift (θ) equilibrium amplitude (a_1)	C, a_1 k, a_1 k, θ ω_1, a_1 ω_l, θ

Explicitation assignments ask the learner to explain a phenomenon simulated by the system. The single explicitation assignment present in SETCOM presents three states of the system and the learner is asked to give the underlying principle for the observed phenomenon, which is super- and subcritical damping. This assignment is present at the *damped harmonic* model progression level. An overview of the optimisation and specification assignments in SETCOM is given in Table 2.

Table 2. *Overview of the specification and optimisation assignments in SETCOM.*

Model progression level	Optimisation assignments	Specification assignments
simple harmonic	<none>	predict *frequency* predict $\lambda_{1,2}$
damped harmonic	control C, find C_{crit} control k, find C_{crit} control m, find C_{crit}	predict $Im\ \lambda_{1,2}$ predict $Re\ \lambda_{1,2}$
forced oscillation	control k, find maximum a_1 control ω_1, find maximum a_1	<none>

Table 2 shows that the optimisation assignments and specification assignments do not appear on all model progression levels. The reason for this is that at the highest model progression level specification assignments would require complex calculations, which we did not intend to train learners in. On the lowest level of model progression, any optimisation assignment would be trivial.

Explanations

SETCOM includes an explanation for each variable present in any of the model progression levels. These explanations consist of simple text and graphics. For most variables the formula(s) describing the variable are given together with clarifications of the other parameters involved in the formula. Figure 4 displays the explanation window and an example of such an explanation.

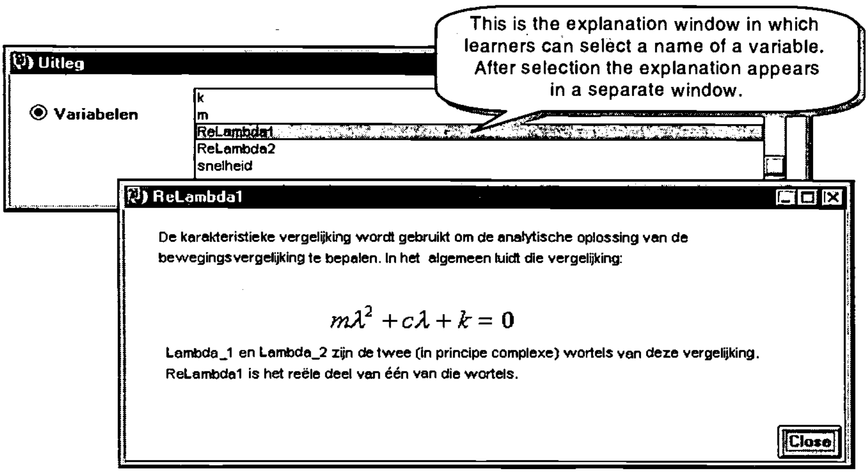

Figure 4. *The explanation window and an example explanation.*

A second set of explanations is that of the *feedback* explanations. Feedback explanations appear as feedback on assignments, e.g. “this is not the right answer try to set the value of the damping constant to a value greater than the critical damping”. For all alternatives in investigation assignments, and for all constraints in optimisation assignments a feedback explanation is defined.

Learner control

In SETCOM, at any point, the learner may choose from a set of assignments, explanations, manipulate the simulation, or choose for a new level of model progression. In SETCOM we introduced a number of controls that limited learners somewhat in their freedom. These controls were:

- When a model progression level is active, all assignments for that model progression level are enabled;
- When a model progression level is active, all explanations for variables appearing in that model progression level are enabled;
- Feedback explanations are never available for free selection by the learner, they only appear as feedback;

- Once all investigation assignments for a model progression level have been completed, or when 20 minutes are spent on a model progression level, the next model progression level is enabled;
- Once enabled, a model progression level stays enabled. This means that learners can always return to a model progression level that was previously visited.

These settings assure that the learner has a great freedom in exploring SETCOM. Mainly, just inconsistent choices are prevented, such as trying to select an assignment of a different model progression level than the one currently active. The constraints on proceeding to a next model progression level are implemented to enforce the idea of model progression by ensuring that learners spend a relevant amount of time on a level before proceeding to the next one.

5.2 Method

The study described here aimed to measure the effects of offering model progression, in combination or without assignments, in a simulation learning environment on oscillations. Three versions of this learning environment were developed, one with model progression and assignments, one with model progression without assignments, and one with neither model progression nor assignments. All three learning environments included –the same– explanations. Subjects participated in a session with one of these three environments. Before and after this session they received some tests, measuring different kinds of knowledge. During the session, the subjects' actions were recorded in a logfile, and they were queried on several aspects of their cognitive load.

5.2.1 Experimental conditions

Three versions of SETCOM were created. One was the full version as described above. In the second version no assignments were available, and in the third version, also the model progression was omitted. In this last version, subjects only saw the *forced oscillator* model progression level, so from the start of their session they could access *all* variables.

5.2.2 Subjects

Sixty-three subjects participated in the study. They were first year physics students who had just followed an introductory course on dynamics. The students were randomly assigned to one of the three conditions such that N = 21 for each experimental condition. Subjects participated in the study voluntarily and received a small fee for their participation.

5.2.3 Tests

For assessing the learners' knowledge a series of three tests was used. The *definitional knowledge test* aimed at measuring students' knowledge of concepts, the *intuitive knowledge test* intended to measure the student's difficult-to-verbalise-insightful knowledge of the topic, and the *propositional knowledge test* aimed at directly measuring (i.e., knowledge articulated by student) the student's knowledge of relations in the domain. The definitional and intuitive knowledge test were presented as pre- and post-test; the propositional test was presented as post-test only. For the definitional knowledge test the same test was used for pre- and post-test, for the intuitive knowledge test parallel versions were used. The definitional and intuitive tests were computer administered, the propositional test was a paper-and-pencil test.

Definitional knowledge

The tests for definitional knowledge concerned the knowledge of individual elements from the domain. Multiple choice items (presenting three answer alternatives) assessing the definitional knowledge about the facts and concepts of the domain were used to measure this kind of knowledge. An example of two definitional test items is depicted in Figure 5. The definitional knowledge test consisted of 25 items.

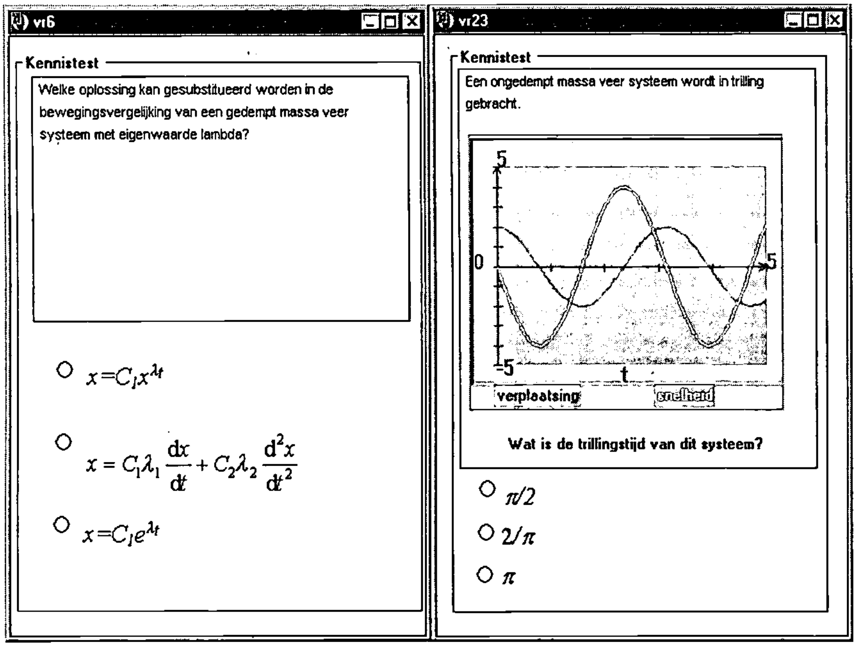

Figure 5. *Two example definitional items used in the experiment. The left hand item asks "Which solution can be substituted in the equation of a damped-mass-spring-system with eigenvalue lambda?". The right hand item states "A mass-spring system without damping is brought into oscillation". What will be the oscillation time of the system?"*

Intuitive knowledge

For measuring intuitive knowledge about the relations between the variables of the domain, we created a test that we called the *speed* WHAT-IF *test.* In the speed WHAT-IF test each test item contains three parts: conditions, actions, and predictions. The conditions and predictions are states in which the system can be. The conditions are displayed in a drawing of the system and some text. The action, or the change of a variable within the system, is presented in text. Finally, the predicted states are also presented in text. The speed WHAT-IF task requires the learner to decide as accurately and quickly as possible which of the predicted states follows from a given condition as a result of the action that is displayed. The items of the task are kept as simple as the domain permits, and the items have a three-answer format. Two parallel versions of the intuitive knowledge test were developed, each consisting of 25 questions. The versions differed on details of the changes given. One of these versions was given as pre-test, the other as post-test. These versions were developed to prevent memorisation effects. For determining the level of intuitive knowledge both correctness and answer time required were used. Students were instructed to answer as accurately and quickly as possible. Two example WHAT-IF items are depicted in Figure 6.

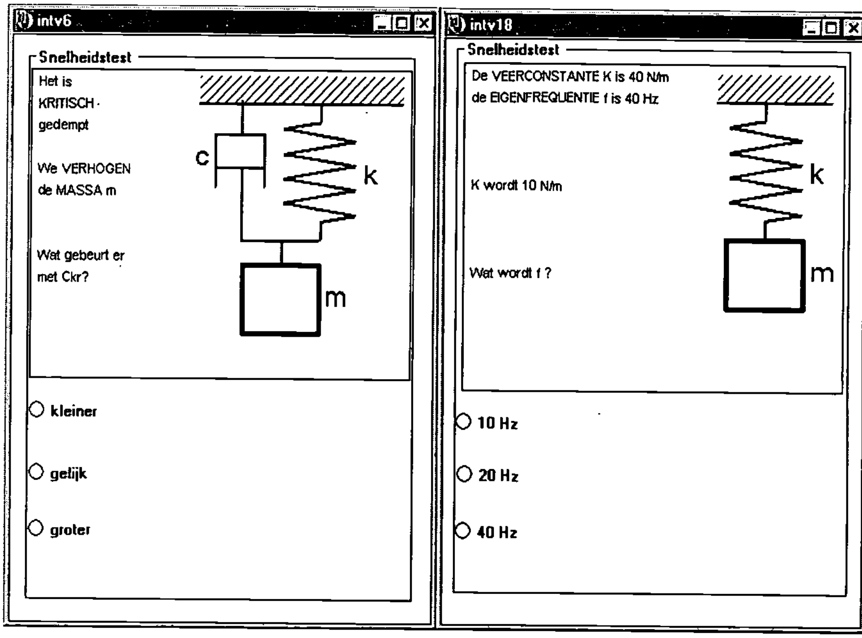

Figure 6. *Two example WHAT-IF items used in the experiment. On the left, a qualitative item, on the right, a numerical item for which no calculation is needed to solve it. The left hand item tells "The damping is critical, the mass m is increased, what happens to Ckr? decreases, same, increases". The right hand item states "The force constant K is 40 N/m, the eigenfrequency f is 40 Hz. Now K becomes 10 N/m. f becomes?"*

Propositional knowledge

Propositional knowledge, on relations between variables in the domain, was measured using a propositional knowledge test. On this test, learners were confronted with pairs of variables, present in the simulation. For each of those pairs, they had to state a relation they thought valid between the variables given. Also they had to indicate whether the relation always holds, or only in a limited number of cases. Students were told that they could use both their own words and/or formula. Furthermore, it was explained to them that of all the descriptions they gave, only the correct ones were counted, and that no attention would be paid to the incorrect ones. The propositional knowledge test aims, like the WHAT-IF tests, on relations between specified variables. The range in level of detail of the two test formats can be considered identical. However, the two formats contrast on the demand they place on the verbal skills of the learners.

5.2.4 Interaction behaviour and cognitive load

We registered all the actions learners made while interacting with the simulation. This provided us with data on the use of the simulation and the supportive measures that were present. These data were used to make a comparison between groups, but also to relate specific interaction patterns with outcomes of the post-tests.

Another type of measurement we introduced is a measurement of subjectively experienced *cognitive load*. Subjects' cognitive load during the learning session was measured by means of a pop-up electronic questionnaire, the S.O.S. scale. Subjects' opinion on three aspects of the environment were gathered: *subject matter* difficulty (is the subject matter experienced as easy or difficult), *operating* the system (is working with the system easy or difficult), and usability of *support* tools (do support measures make the understanding of the subject matter better or worse).

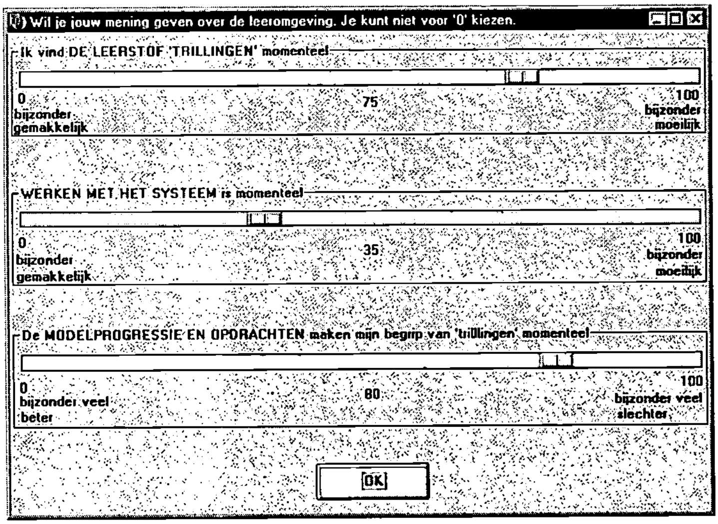

Figure 7. *The S.O.S. scale measuring three aspects of the cognitive load of interacting with the learning environment: subject matter ("leerstof") difficulty, operating ("werken met") the system, and support (i.e., model progression and assignments "modelprogressie en opdrachten") added.*

At regular moments the S.O.S. scale appeared and subjects had to complete it before they continued working with the environment. By pulling sliders subjects could indicate their ratings. Subjects scores could range from 0 to 100, where 100 was the 'negative' side, meaning that the subject matter was extremely difficult, the environment was extremely difficult to work with, and support made the task much more difficult. This scale[2] is depicted in Figure 7.

The questionnaire was set to pop up every 10 minutes, but display was always postponed until an event occurred that marked the end of a coherent subject's action, such as closing an explanation or completing an assignment. This was done in order not to let this measurement interfere with the discovery behaviour.

5.2.5 Procedure

Each experimental session had a duration of approximately three and a half hours. It consisted of the following parts in chronological order:

- *Introduction (5 minutes)*
  Subjects were welcomed and given an overview of the activities that they would be engaged in during the session. They were also explained the target of the learning session and the subject domain (oscillatory motion).

- *Pre-tests (30 minutes)*
  After the general introduction the definitional and intuitive pre-tests were administered. This took about 30 minutes all together.

- *Introduction to the learning environment (10 minutes).*
  After having completed the pre-tests subjects read an introduction on the SETCOM 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 students that both their performance on the tests and their interaction with the learning environment would be recorded. Furthermore, it was clarified how their performance would be evaluated.

- *Interaction with SETCOM (set at 2 hours and 15 minutes)*
  After the introduction subjects learned with the SETCOM environment on their own. The experiment leader was present and could give assistance on questions concerning the operating of the environment, but not on the subject matter contents. Subjects were encouraged to use the full two hours and a quarter available for the interaction. If they wanted to stop earlier they were stimulated to explore more of the environment, however, they were not forced to do so. During the interaction, coffee, tea, and sweet snacks were served.

- *Post-tests (30 minutes)*
  After the interaction with the simulation environment the post-tests were presented. The sequence of presentation was first, the definitional test, then the

[2] The S.O.S. scales were adapted to the learning environments of each of the three experimental conditions. The example given in Figure 7 displays the s.o.s scale for the condition with assignments and model progression.

intuitive knowledge test, and finally the propositional knowledge test. The first two of these tests were presented electronically (as were the pre-tests), the propositional test was administered using paper and pencil.

5.3 Predictions

With respect to the knowledge tests, we expect across all three experimental conditions a considerable gain at the intuitive WHAT-IF test, and a small or no gain at the definitional knowledge test. If the definitional test, the WHAT-IF test, and the propositional test measure different aspects of knowledge, we could possibly expect low correlations between the scores on these tests.

We predict that learners of the experimental conditions that have assignments and/or model progression available will perform better on the intuitive WHAT-IF test than the learners who are not supported. Furthermore, it is speculated that he learners of the experimental condition with assignments and model progression will perform better than on the intuitive WHAT-IF test the learners with just model progression.

We do not expect that the experimental conditions that have assignments and/or model progression available will necessarily score better on the definitional knowledge test than the control condition, as the effects of discovery learning in general should especially improve the intuitive character of knowledge.

In addition, we do not expect differences between the experimental conditions on the propositional test. However, we are especially interested in possible relations between scores on this test and the WHAT-IF test as both tap knowledge on relation between variables of the domain.

With regard to cognitive load, we expect that learners of the experimental conditions that have assignments and/or model progression available are supported in their learning process and for this reason will experience a lower cognitive load than the non-support group. Nevertheless, we do not foresee that the reported cognitive load of learners who have assignments and/or model progression will necessarily be lower than the reported cognitive load of the learners of the non-support condition Obviously, as the support tools make the environment, per definition, more complex, adding support tools may also raise cognitive load. These 'contradictory' effects, however, should be reflected in the different aspects of the cognitive load measure (see Section 5.2.4).

6. Results

In this section we will first report the results on different knowledge tests, then we will give an account of the interaction behaviour and the cognitive load measure, and, finally, we relate a number of the interaction behaviours to performance measures.

6.1 The definitional knowledge test

The definitional knowledge test was given in the same form as pre- and as post-test. It consisted of 25 multiple choice items with 3 alternative answers each. A reliability analysis on the definitional pre-test ($N = 63$; $n = 25$ items) resulted in the removal of one item that lowered the total test reliability to a considerable extent. The resulting test reliability was .49 (Cronbach's α). The reliability analysis on the same test used as post-test also resulted in the removal of one item and then yielded a reliability of .62 (N = 63; n = 24 items). The average number correctly answered items on the definitional pre-test was 15.9 with an SD of 2.8 and a range going from 9 to 21 correct items out of 24. On the definitional post-test the number correct scores had a mean of 18.0 with a SD of 3.0 and a range from 11 to 24. Table 3 and Figure 8 give the average numbers of correct items for the definitional pre- and post-tests for the three experimental conditions averaged over subjects.

Table 3. *Average number of correctly answered items on the definitional pre-test and definitional post-test (n = 24 items)*

Condition	definitional pre-test	definitional post-test
I (model progression and assignments)	16.0 (*sd* = 2.5)	17.9 (*sd* = 3.2)
II (model progression, no assignments)	16.3 (*sd* = 2.5)	19.1 (*sd* = 2.3)
III (no model progression, no assignments)	15.5 (*sd* = 3.4)	17.1 (*sd* = 3.2)
Overall average	15.9 (*sd* = 2.8)	18.0 (*sd* = 3.0)

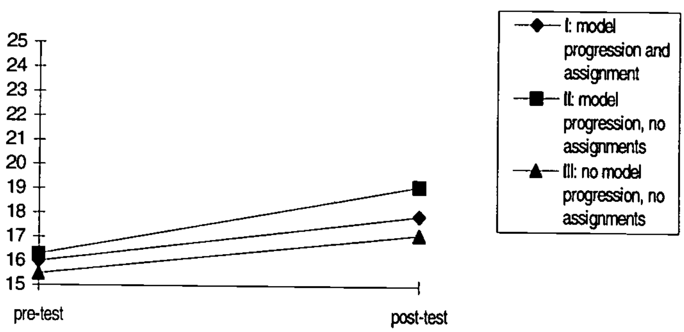

Figure 8. *Average number correctly answered items on the definitional pre-test and definitional post-test*

In every condition I, II and III, some students had a lower post-test than pre-test score. In conditions I five students, and in condition III three students scored 1 item less. In condition II one student had two items less correct, and one student had one item less correct on the post-test in comparison with the pre-test.

A repeated measurement analysis on the definitional test-scores showed a significant within-subject effect of number of correct items ($F_{1, 60} = 45.5$, $p < .001$). No interaction between experimental condition and test scores was revealed in this analysis. ANCOVA's –on post-test scores with pre-test scores as covariate– over pairs of conditions showed that the difference between condition II and III was significant ($F_{1, 39} = 4.4$, $p < .05$). The other comparisons yielded no significant differences in definitional post-test scores.

6.2 The intuitive knowledge test

For the intuitive test, items are scored on both the correctness of the answer and on the time used for giving the answer. On the basis of a reliability analysis and an analysis of outliers in response time, a number of items were excluded from further analysis.

Reliability analyses on the WHAT-IF pre-test across 63 students resulted in the removal of two items (of the total number of 25 items) that lowered the total test reliability to a considerable extent. The resulting test reliability was .43 (Cronbach's α). Reliability analyses on the WHAT-IF post-test resulted in the removal of one item and then yielded a Cronbach's α of .70 ($N = 63$, $n = 24$ items).

In order to identify outliers in the response times to the WHAT-IF items, for every student ($N = 63$) average response times and SD's across WHAT-IF pre-test and post-test items were computed. A response time was defined an outlier if it was more than three standard deviations from the individual average response time. We have chosen this method to identify outliers because the method takes into account individual differences. Using this procedure, overall no more than 1.7 % of the data was excluded from further analyses.

The number of items over which analyses were done differed between students because the removal of outliers was performed on the basis of individual data. For Condition II an average total over all students of 23.7 items remained, for Conditions I and III this was 23.8 items on the average.

The average number of correctly answered items – after exclusion of the outliers – on the WHAT-IF pre-test was 7.1 with an SD of 2.6 and a range going from 2 to 15 correct items out of 23. On the WHAT-IF post-test the number correct scores had a mean of 12.9 with a SD of 3.7 and a range from 6 to 20. The average time to answer WHAT-IF pre-test items was 16.9 seconds with a SD of 4.6 and a range from 8.8 up to 33.5 seconds to respond to the WHAT-IF pre-test items. For the WHAT-IF post-test the average item response time was 17.1 seconds, the SD was 3.8 and the range of the latencies went from 9.5 to 27.1 seconds (see Figure 9, Table 4, and Table 5).

Table 4. *Average number correctly answered items on the WHAT-IF pre-test and WHAT-IF post-test*

Condition		WHAT-IF pre-test	WHAT-IF post-test
I	(model progression and assignments)	7.2 ($sd = 2.6$)	14.8 ($sd = 3.3$)
II	(model progression, no assignments)	6.9 ($sd = 2.8$)	12.8 ($sd = 3.7$)
III	(no model progression, no assignments)	7.0 ($sd = 2.5$)	10.3 ($sd = 2.7$)
Overall average		7.1 ($sd = 2.6$)	12.9 ($sd = 3.7$)

Table 5. *Average item response times (in seconds) of the WHAT-IF pre-test and WHAT-IF post-test*

Condition		WHAT-IF pre-test	WHAT-IF post-test
I	(model progression and assignments)	17.0 (sd = 4.4)	17.5 (sd = 3.6)
II	(model progression, no assignments)	16.1 (*sd* = 2.9)	16.9 (*sd* = 3.1)
III	(no model progression, no assignments)	17.7 (sd = 5.9)	16.9 (sd = 4.6)
Overall average		16.9 (*sd* = 4.6)	17.1 (*sd* = 3.8)

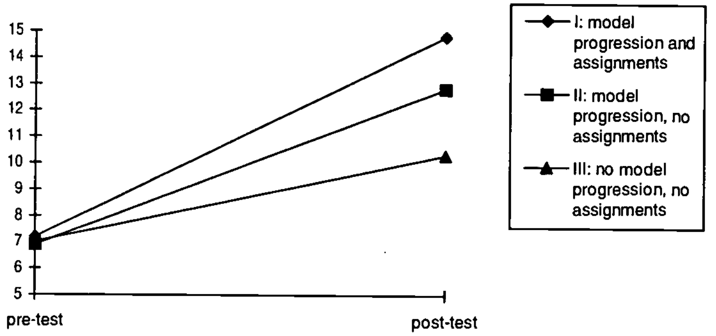

Figure 9. *Average number correctly answered items on the WHAT-IF pre-test and WHAT-IF post-test*

We did not find a trade-off between correctness and speed. The correlations found between answer time and correctness had a value of $r = .14$, $p > .10$, when computed within students across the WHAT-IF pre-test items, a value of $r = .16$, $p > .10$, when computed within students across the WHAT-IF post-test items, a value of $r = -.29$, $p > .10$ when computed within WHAT-IF pre-test items across students, and finally a value of $r = -.46$, $p < .05$ when computed within WHAT-IF post-test items across students.

Across all the 63 students only one student had a lower post-test correctness score which was in condition I where one student had one item less answered correctly on the WHAT-IF post-test in comparison with the pre-test. All other 62 students showed a knowledge gain on the WHAT-IF test. Furthermore, as can be read from Table 5, no gain or loss in average item response time was found in this experiment.

A repeated measurement analyses on the WHAT-IF test scores showed a significant within-subject effect of number of correct items ($F_{1,\,60} = 237.3$, $p < .001$). Moreover, an interaction between experimental condition and test scores was found ($F_{2,\,60} = 11.83$, $p < .001$). Subsequent ANOVA's on the gain of the WHAT-IF test over pairs of conditions yielded both a significant difference between the experimental conditions I and III ($F_{1,40} = 23.9$, $p < .001$), and between conditions II and III ($F_{1,40} = 11.5$, $p < .05$). No significant differences in WHAT-IF test improvement was found between conditions I and II[3].

6.3 The propositional knowledge test

For assessing the subjects performance on the propositional knowledge test we scored the completed hypotheses lists of the students. This resulted in two different measures: the number of correct hypotheses and the average precision of the hypotheses. The precision score of the hypotheses could range from 1 to 4, with "1" indicating that learners successfully stated that a relation between two variables existed or not, a "2" was given if learners indicated the right qualitative relation (e.g., "if a increases b also increases"), "3" was scored if the correct quantitative relational specification (e.g., "if a multiplies by 2 then b multiplies by 2") was stated, and a relation was scored "4" if the right numerical formulation (i.e., the exact formula) was given by the learners (Van Joolingen, 1995). Table 6 shows the average number correct hypotheses and the average precision of the relations specified by the learners.

Table 6. *Average propositional knowledge measures*

Condition		number of correct hypotheses out of 7	average precision of the hypotheses ranging from 1 to 4
I	(model progression and assignments)	4.1 (sd = 1.0)	2.5 (sd = .55)
II	(model progression, no assignments)	4.0 (sd = 1.8)	2.3 (sd = .44)
III	(no model progression, no assignments)	4.4 (sd = 1.6)	2.5 (sd = .39)
Overall average		4.2 (*sd* = 1.5)	2.4 (*sd* = .46)

ANOVA's showed no differences between conditions on neither number of hypotheses nor average precision of hypotheses ($F_{2,60} < 1$ for both analyses).

6.4 Relations between the different tests

Table 7 displays the correlations between the three knowledge tests over all three conditions. For the WHAT-IF speed test results are given for correctness of the items and for time separately. In Table 8 the correlation between the gain in definitional test score and gain in WHAT-IF correctness test score is given.

The pattern, resulting from the use of the three measures, which emerges from this analysis is that we find three clear clusters. The first one consists of the definitional test and WHAT-IF correctness, the second one is the WHAT-IF time aspect. The test for propositional knowledge, as measured through the hypotheses lists, correlates neither with the first cluster nor with the second one, and could be regarded as a third cluster.

[3] However, if we remove from our analysis the only student who scored lower on WHAT-IF the post-test compared to the pre-test, the difference in WHAT-IF test gain between conditions I and II becomes significant ($F_{1,40} = 5.1$, $p < .05$).

Table 7. *Correlations between the different aspects of knowledge over the three conditions on the post-test scores (levels of significance between parentheses)*

	WHAT-IF correct	WHAT-IF speed	Number of correct hypotheses
definitional	.49 ($p < .05$)	-.02 ($p > .10$)	.21 ($p > .10$)
WHAT-IF correct		.16 ($p > .10$)	.12 ($p > .10$)
WHAT-IF speed			.11 ($p > .10$)

Table 8. *Correlations between the different aspects of knowledge over the three conditions on the post-test scores (levels of significance between parentheses)*

	WHAT-IF correct gain
definitional gain	.06 ($p > .10$)

However, if we look at the gain in correctness on the definitional and WHAT-IF tests we can clearly see that both scores are not related, indicating that a gain in intuitive knowledge does *not* automatically yield a gain in definitional knowledge.

6.5 Interaction behaviour

We registered all the actions learners made while interacting with the simulation. This provided us with data on the use of the simulation and the supportive measures that were present. Due to technical problems two log-files of students in Condition III were lost. In the subsequent analyses the complete interaction data of 61 subjects were used.

Number of runs

Students were rather active in the simulation. Table 9 shows the average number of runs[4] over the three conditions. As can be seen from the standard deviations the individual differences are enormous.

Table 9. *Average number of runs in the three conditions*

Condition	Number of runs
I (model progression and assignments)	61.9 (*sd* = 36.9)
II (model progression, no assignments)	86.2 (*sd* = 30.6)
III (no model progression, no assignments)	81.5 (*sd* = 59.0)
Overall average	76.4 (*sd* = 43.9)

An ANOVA on number of runs across the three conditions showed no significant differences: $F_{2,58} = 1.86$, $p > .10$. Subsequent ANOVA's including pairs of conditions yielded a significant difference between the experimental conditions I and II: $F_{1,40} = 5.4$, $p < .05$, but not between the other experimental groups.

[4] This is the number of times students click on the "run" button. An other way of running the simulation and perceiving changes of manipulated variables is to dynamically change values of variables, while the simulation is running. These "runs" are *not* included in our count.

Number of assignments and explanations used

Most subjects made moderate to extensive use of assignments and explanations, for some subjects however the explanations were less popular. One subject in Condition I consulted no explanations at all, and one subjects in Condition I, and one in Condition III just opened one explanation. Table 10 displays the average number of different assignments and explanations used.

Table 10. *Average number of different assignments and explanations used for the three conditions*

Condition	number of assignments (total 27)	number of explanations (total 16)
I (model progression and assignments)	13.6 (sd = 2.3)	6.4 (sd = 2.9)
II (model progression, no assignments)	-	8.6 (sd = 1.7)
III (no model progression, no assignments)	-	8.2 (sd = 2.4)
Overall average	-	7.7 (*sd* = 2.5)

An ANOVA on the number of explanations indicated significant differences between the three experimental conditions ($F_{2,58} = 5.22$, $p < .05$). Subsequent ANOVA's including pairs of conditions showed significant differences between the experimental conditions I and II ($F_{1,40} = 9.3$, $p < .05$), and conditions I and III ($F_{1,38} = 4.6$, $p < .05$), but not between the experimental groups II and III ($F_{1,38} < 1$).

6.6 Cognitive load

Subjects' cognitive load during the learning session was measured by means of a pop-up electronic questionnaire, the S.O.S. scale. Subjects' opinion on three aspects of the environment were gathered: *subject matter* difficulty (is the subject matter seen as easy or difficult), *operating* the system (is working with the system easy or difficult), and usability of *support* tools (do support measures make the understanding of the subject matter better or worse). At regular moments the S.O.S. scale appeared and subjects had to complete it before they continued working with the environment. By pulling sliders subjects could indicate their ratings. Subjects scores could range from 0 to 100, where 100 was the 'negative' side, meaning that the subject matter was extremely difficult, the environment was extremely difficult to work with, and support made the task much more difficult.

In condition I the learners indicated their perceived difficulty of the topic, their appreciation of the system and their opinion on the helpfulness of the support on the average 10.5 times with a range of 9 to 12 times. In condition II the average was 10.3 times and the range was 7 to 12 times. In condition III the average was 9.6 times and the range was 5 to 12 times.

Table *11* displays the correlations between the three rated cognitive load aspects.

Table 11. *Correlations between the cognitive load aspects across the three conditions*

	subject matter difficulty	operating the system	support provided
subject matter difficulty		.34 ($p < .05$)	.27 ($p < .10$)
operating the system			.36 ($p < .05$)

Correlations, though on two out of three occasions significant, are of a moderate level, indicating that the three measures assess different aspects of cognitive load. The scores at the cognitive load measures in the three conditions are given in Table 12.

Table 12. *Average scores on the three measures of 'cognitive load'*

Condition	subject matter difficulty	operating the system	support provided
I (model progression and assignments)	60.7 (*sd* = 18.0)	30.6 (*sd* = 22.4)	38.4 (*sd* = 7.9)
II (model progression, no assignments)	46.5 (*sd* = 13.9)	25.5 (*sd* = 22.4)	39.3 (*sd* = 10.5)
III (no model progression, no assignments)	54.3 (*sd* = 19.6)	22.9 (*sd* =13.8)	-
Overall average	53.8 (*sd* = 18.0)	26.5 (*sd* = 17.8)	38.8 (*sd* = 9.2)

ANOVA's showed no differences between the two experimental groups on helpfulness of the support ($F_{1,40} < 1$), nor on operating the system across the three conditions ($F_{1,58} < 1$). Likewise, subsequent ANOVA's including pairs of conditions showed no significant differences between the experimental conditions on the appreciation of operating the system ($F_{1,40} < 1$ and $F_{1,38} < 1$ for comparisons between I and II, and for II and III, $F_{1,38}= 1.64$, $p > .10$ for comparison between I and III). However, the experimental conditions differed with respect to the subject matter difficulty rating: $F_{1,58} = 3.6$, $p < .05$). ANOVA's including pairs of conditions showed significant differences between the experimental conditions I and II ($F_{1,40} = 8.2$, $p < .05$), but not between the other conditions ($F_{1,38} = 1.15$, $p > .10$ for I and III, $F_{1,38} = 2.2$, $p > .10$ for comparisons of II and III).

6.7 Interaction of behaviour and learning results

We already found that the experimental manipulations, i.e., the extent to which support is provided to the students, had their effects on the post-test scores. Here, we take a closer look at the relations between use of instructional measures (assignments and explanations, model progression could not be *used* but was present) and scores on the knowledge tests. Table 13 displays the correlations within condition I between number of assignments used and the scores on the post-tests. Table 14 shows the correlations across the three experimental conditions between number of explanations used and the scores on the post-tests.

Table 13. *Correlations in condition I between knowledge scores (post-tests) and number of assignments used*

post-test score	number of assignments
Definitional post-test scores	-.05 ($p > .10$)
WHAT-IF post-test correctness scores	.23 ($p > .10$)
WHAT-IF post-test item response times	-.23 ($p > .10$)
number of correct hypotheses	.12 ($p > .10$)

Table 14. *Correlations across all three experimental conditions between knowledge scores (post-tests) and number of explanations used*

post-test score	number of explanations
Definitional post-test scores	-.17 ($p > .10$)
WHAT-IF post-test correctness scores	-.09 ($p > .10$)
WHAT-IF post-test item response times	-.02 ($p > .10$)
number of correct hypotheses	-.13 ($p > .10$)

No correlation reached a level of significance below .05, indicating, among others, that within condition I we can not identify a relation between the number of assignments used and the post-test scores. Neither can we say anything about the relation between the explanations consulted, across the three experimental conditions, and the post-test scores.

Table 15. *Correlations between knowledge scores (post-tests) and number of runs used*

post-test score	number of runs
Definitional post-test scores	-.38 ($p < .05$)
WHAT-IF post-test correctness scores	-.30 ($p < .05$)
WHAT-IF post-test item response times	.11 ($p > .10$)
number of correct hypotheses	-.06 ($p > .10$)

The figures in Table 15 show that two of the correlations reach a level of significance below .05. Considering the correlations taken over the three conditions we may therefore conclude that it appears that a higher number of runs is associated with lower post-test scores. When this correlation is computed within experimental conditions the picture in Table 16 emerges.

Table 16. *Correlations between knowledge correctness scores (post-tests) and number of runs for each experimental condition*

post-test score x number of runs	condition I	condition II	condition III
Definitional post-test scores	-.41 ($p < .10$)	.17 ($p > .10$)	-.74 ($p < .05$)
WHAT-IF post-test correctness scores	-.24 ($p > .10$)	-.05 ($p > .10$)	-.47 ($p < .05$)

Finally, we computed the correlations between aspects of cognitive load, as measured with the S.O.S. scale, and the post-test scores. They were determined across the experimental conditions and are displayed in Table 17.

Table 17. *Correlations between knowledge scores and measures of cognitive load*

post-test score	subject matter difficulty	operating the system	support provided
definitional post-test scores	-.14 ($p > .10$)	-.08 ($p > .10$)	-.33 ($p < .05$)
WHAT-IF post-test correctness scores	.03 ($p > .10$)	.01 ($p > .10$)	-.26 ($p < .10$)
WHAT-IF post-test item response times	.11 ($p > .10$)	-.33 ($p < .05$)	.11 ($p > .10$)
number of correct hypotheses	-.03 ($p > .10$)	.05 ($p > .10$)	.25 ($p > .10$)

From Table 17 we can read that one of the significant correlations can be found for definitional post-test scores, indicating that subjects who appreciate the support provided (low score), have higher correctness scores on the definitional post-test. The negative correlation between the operating system and WHAT-IF post-test item response times indicates that students who estimate operating the system as easier have longer item response times. This correlation disappeared in conditions I and II, when computed within the experimental conditions. In condition III the correlation was -.48 ($p < .05$).

7. Discussion

The first main finding of this study is that, as a whole, subjects improved on the knowledge tests, in all three experimental conditions. For definitional knowledge, there was a small gain between the pre- and post-test, meaning that on average students acquired some definitional knowledge during the session of a little more than two hours. We believe that the availability of explanations for all students is the main contributor to the definitional knowledge gain. On intuitive knowledge, the average gain in correctness was substantial in all three conditions. This is in line with our expectations that simulations do have most effect on intuitive knowledge, and not on learning facts and definitions. On the basis of the results of the study we can therefore conclude that, in the context of simulation based discovery learning, it makes sense to introduce new ways of measuring knowledge in addition to traditional 'definitional' type of knowledge tests.

A second important conclusion from the study is that adding support to the simulation helped. On the correctness scores of the intuitive post-test the two conditions with support added outperformed the control group, and the condition with model progression and assignments was very close to outperforming the group with only model progression. We can therefore conclude that, in this situation, adding model progression to a simulation helped the learners in gaining intuitive knowledge, and adding assignments to model progression was close to being successful.

The interaction data as measured with the log-files revealed differences between experimental groups. The activity of subjects, measured in the number of simulation runs and the number explanations used, showed that learners in condition I looked up less explanations and performed less runs than in the other conditions. A likely explanation is that learners in condition I simply devoted considerable part of their time to assignments, which were not available in conditions I and II.

In our data we found a relation between the interaction patterns and the post-test results of the definitional knowledge and the WHAT-IF tests. An overall negative correlation was found between the number of runs and the correctness scores of the post-tests. These negative correlations were significant when computed across the three experimental groups. So, on the whole, the more runs the learners used, the less items they responded correctly. When these correlations were calculated within experimental conditions they only remained significant (at the .05 level) in condition III. In condition I the correlations between number of runs and post-test scores stayed negative and of a moderate size, but were not significant. In condition II this pattern disappeared. These results are in contrast with De Jong et al. (1993) where a correlation between interaction level (measured as number of iterations) and performance was found. A possible explanation for the negative correlations is that learners may have devoted much of their time to running the simulation at the expense of thinking about the domain of harmonic oscillations, and the relations between variables within the domain. It is not clear why the negative correlations are found in just these experimental conditions, and not in the other. Finally, it should be noted that the way we counted the runs (see footnote 4) may be responsible for this inconclusive picture.

Like in other evaluations of SMISLE learning environments (de Jong et al., 1995; van Joolingen, van der Hulst, Swaak, & De Jong, 1995), subjects showed that they like the idea of assignments. The virtue of this is that assignments seem to have their expected guiding role, in the sense that they get learners going with the simulation. A drawback that we found in previous studies was that subjects showed a tendency to identify the discovery task with completing all assignments. We tried to overcome this issue in the present experiment by telling the learners explicitly that this was *not* the purpose of assignments; assignments should only be used if learners thought them helpful in their discovery process. The same was told about the use of the explanations, the model progression tool, and the simulation window. We directly explained the students to use the environment, the way they liked most, the way they conceived most fruitful to their learning. We think we were successful in explaining the students the freedom was in their hands: the range of used assignments was 8 to 16 assignments out of 27, for the explanations this was 0 to 10 out of 16, and the number of runs ranged from 18 to 252.

For complex (learning) environments, cognitive load might be an important factor in the learning process. In this study we measured cognitive load by means of an electronic questionnaire that popped up once in a while. Three different aspects of cognitive load were measured: *subject matter* difficulty, *operation* of the environment, and *usability* of the support measures in understanding the subject matter. These different aspects indeed appeared to measure to a certain extent different aspects as became clear from their correlations. The environments from the three experimental conditions differed with respect to subject matter difficulty, but not on the ratings of system operation or helpfulness of support measures. A possible explanation for the higher subject matter difficulty rating in the experimental condition in which assignments are available, is that in trying to answer the assignments, students learned that they not always were right at the first trial. They received negative feedback (i.e., "this is not the right answer"), some hints (e.g. "try to set the value of the damping constant to a value greater than the critical damping"), and were encouraged to try again. As a consequence learners in this conditions became more conscious of their understanding of the domain, and rated it accordingly as more difficult.

The fact that the subjects on the whole, rated the *subject matter* difficulty higher than the operating aspect of cognitive load, indicates that tackling the domain of harmonic oscillations took more cognitive resources than *operating* the instructional environment. Moreover, the fact that the average rating of helpfulness of the *support* is below 50, tells us again that the support is indeed helping learners to understand the domain at hand, instead of interfering their learning process. These results may not seem very surprising. However, when disappointing results are found in simulation-based discovery studies, extreme cognitive load as resulted from operation difficulties or extra support, is several times mentioned as a possible reason why so less is learned (e.g., Shute, 1991; De Jong, et al. 1993; Njoo, 1994).

The results of this study can also be used for a further validation of the WHAT-IF test format, which is a relatively new format.

In contrast with previous studies (De Jong et al., 1995a,b; Van Joolingen et al., 1995), in this experiment, no decrease in item answer time for the WHAT-IF test is found. However, at the same time, the gain in correctness is much larger in the present study and the average item answer time of the items in this study was far lower than those of WHAT-IF items of earlier work. In one particular study (Van Joolingen et al., 1995), we used a comparable set of items and found average answer times of 25 seconds for the pre-test and 20 seconds for the post-test items. In the present experiment learners needed on the average 17 seconds for both pre- and post-test items. Further experimentation should tell us more about the maximum speed to be expected in answering these type of items within complex physics domains.

In agreement with prior studies, no trade-off between correctness and completion times is detected. This entails that there is no evidence that the incorrect items are answered quicker than the correct ones. There even seems an indication of the reverse, i.e., that the quicker an item is answered, the higher the chance it is correct which fits well with our preconceptions on intuitive knowledge.

Like in former work (Van Joolingen et al., 1995), comparisons between the post-test scores showed no correlation between the WHAT-IF test and the number of correct hypotheses. This is interesting as both tests require the same knowledge about relations between variables of the domain. However, the two formats contrast on the demand they place on the ability of the learners to formulate the relations: while in the WHAT-IF test there is no need at all for verbalisation, in the propositional test this is of uttermost importance. If it is argued that for knowledge to become intuitive, knowledge first has to go through a verbal phase then considerable correlations should have been expected between the WHAT-IF test scores and the propositional knowledge test scores. If, it is, on the other hand, believed that intuitive knowledge is acquired by a more implicit experiential learning mode, without need for explicitation, then no relations are foreseen between the two test scores. Our data are in line with the latter hypothesis.

We started this work with highlighting two general issues in research on effectiveness of discovery learning. We stated as a first cause for finding disappointing results in this research that learners might experience problems in the discovery learning process. Therefore, in this study we supported the learners in regulating their discovery process with several types of instructional measures. A second mentioned option for the studies lacking positive effects of discovery, entailed that discovery learning might lead to a more 'intuitive', deeply rooted form of knowledge that might not measured adequately by the tests used in these studies. For that reason, in this experiment, we applied a new type of test, the WHAT-IF test, intended to tap intuitive knowledge. The results of this study were in line with our expectations: the support measures worked and students mainly acquired intuitive knowledge. We will continue this line of research in future work. In upcoming experiments we will fine-tune the support measures and the assessment –both the knowledge acquired and the activities of the learners–. These modifications will then hopefully, on the one hand, improve learning even further, and on the other hand, provide still a more clear picture on what is going on during discovery learning and tell us more about what *is* learned, and what is *not* learned.

8. References

Anderson, J.R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review, 94(2),* 192-210.

Berkum, J.J. van, & Jong, T. de (1991). Instructional environments for simulations. *Education & Computing*, *6*, 305-359.

Berry, D.C., & Broadbent, D.E. (1984). On the relationship between task performance and associated verbalizable knowledge. *The Quarterly Journal of Experimental Psychology*, *36A*, 209-231.

Berry, D.C., & Broadbent, D.E. (1988). Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology*, *79*, 251-272.

Berry, D.C., & Broadbent, D.E. (1990). The role of instruction and verbalization in improving performance on complex search tasks. *Behaviour and Information Technology, 9(3),* 175-190.

Broadbent, D.E., Fitzgerald, P., & Broadbent, M.H.P. (1986). Implicit and explicit knowledge in the control of complex systems. *British Journal of Psychology, 77,* 33-50.

Brown, D.E. (1993). Refocusing core intuitions: a concretizing role for analogy in conceptual change. *Journal of Research in Science Teaching, 30 (10),* 1273-1290.

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

Carlsen, D.D., & Andre, T. (1992). Use of a microcomputer simulation and conceptual change text to overcome students preconceptions about electric circuits. *Journal of Computer-Based Instruction, 19,* 105-109.

Cates, W. M. (1992). *Considerations in evaluating metacognition in interactive hypermedia/ multimedia instruction.* Paper presented at the annual meeting of the American Educational Research Association, April 20, 1992, San Fransisco, CA.

Chambers, S.K., Haselhuhn, C., Andre, T., Mayberry, C., Wellington, S., Krafka, A., Volmer, J., & Berger, J. (1994, April). *The acquisition of a scientific understanding of electricity: Hands-on versus computer simulation experience; conceptual change versus didactic text.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

Chase, W.G., & Simon, H.A. (1973). Perception in chess. *Cognitive Psychology, 4,* 55-81.

Choi, B., & Gennaro, E. (1987). The effectiveness of using computer simulated experiments on junior high school students understanding of the volume displacement concept. *Journal of Research in Science Teaching, 24,* 539-552.

diSessa, A. (1993). Towards an epistomology of physics. *Cognition and Instruction, 10 (2 & 3),* 105-25.

Fischbein, E. (1987). *Intuition in science and mathematics.* Dordrecht, The Netherlands: Reidel.

Friedler, Y., Nachmias, R., & Linn, M.C. (1990). Learning scientific reasoning skills in microcomputer-based laboratories. *Journal of Research in Science Teaching, 27,* 173-191.

Glaser, R., Schauble, L., Raghavan, K., & Zeitz, C. (1992). 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. 345-373). Berlin: Springer.

Goodyear, P., Njoo, M., Hijne, H., & van Berkum, J.J.A. (1991). Learning processes, learner attributes and simulations. *Education & Computing, 6,* 263-304.

Grimes, P.W., & Willey, T.E. (1990). The effectiveness of microcomputer simulations in the principles of economics course. *Computers & Education, 14,* 81-86.

Hayes, N.A., & Broadbent, D.E. (1988). Two modes of learning for interactive tasks. *Cognition, 28,* 249-276.

Hussy, W. & Granzow, S. (1987). Komplexes Problemlösen, Gedächtnis und Verarbeitungsstill. *Zeitschrift für Experimentelle und Angewandte Psychologie, 34,* 212-227.

Jong, T., de (1991). Learning and instruction with computer simulations. *Education & Computing, 6,* 217-229.

Jong, T., de & Joolingen, W.R. van (1995). The SMISLE environment: Learning with and design of integrated simulation learning environments. In P. Held & W.F. Kugemann (Eds.), *Telematics for education and training* (pp. 173-187). Amsterdam: IOS Press.

Jong, T. de, & Joolingen, W.R. van (in preparation). *Discovery learning with computer simulations.*

Jong, T. de, Härtel, H., Swaak, J., & Joolingen, W.R. van (1995a). *Support for simulation-based learning; the effects of assignments in learning about transmission lines.* SMISLE project, deliverable D51. Enschede: University of Twente.

Jong, T., de, Hoog, R., de & Vries, F. de (1993). Coping with complex environments: The effects of navigation support and a transparent interface on learning with a computer simulation. *International Journal of Man-Machine Studies, 39*, 621-639.

Jong, T. de, Joolingen, W.R. van, Scott, D., de Hoog, R. , Lapied, L., Valent, R. (1994). SMISLE: System for Multimedia Integrated Simulation Learning Environments. In T. de Jong & L. Sarti (Eds.), *Design and production of multimedia and simulation based learning material* (pp. 133-167). Dordrecht: Kluwer Academic Publishers.

Jong, T. de, Martin, E., Zamarro J-M., Esquembre, F., Swaak, J. , & Van Joolingen, W.R. (1995b). *Support for simulation-based learning; the effects of assignments and model progression in learning about collisions.* SMISLE project, deliverable D50. Enschede: University of Twente.

Joolingen, W.R. van (1995). QMaPS: Qualitative reasoning for simulation learning environments. *Journal of Artificial Intelligence in Education, 6(1)*, 67-89.

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: Springer.

Joolingen, W.R. van, Hulst, A. van der, Swaak, J, & Jong, T. de (1995). Support for simulation-*based learning; the effects of model progression in learning about oscillations.* SMISLE project, deliverable D24a. Enschede: University of Twente.

Klahr, D., Fay, A.L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology, 25*, 111-146.

Lajoie, S.P. (1993). Cognitive tools for enhancing learning. In S. P. Lajoie & S.J. Derry (Eds.), *Computers as cognitive tools* (pp. 261-289). Hillsdale (NJ): Lawrence Erlbaum.

Lavoie, D.R., & Good, R. (1988). The nature and use of predictions skills in a biological computer simulation. *Journal of Research in Science Teaching, 25*(5), 335-360.

Leutner, D. (1993). Guided discovery learning with computer-based simulation games: effects of adaptive and non-adaptive instructional support. *Learning and Instruction, 3*, 113-132.

Lewis, E.L., Stern, J.L., & Linn, M.C. (1993). The effect of computer simulations on introductory thermodynamics understanding. *Educational Technology*, 45-58.

Njoo, M. (1994). *Exploratory learning with a computer simulation: Learning processes and instructional support.* PhD thesis. Eindhoven: Eindhoven University of Technology.

Njoo, M., & de Jong, T. (1993). Exploratory learning with a computer simulation for control theory: Learning processes and instructional support. *Journal of Research in Science Teaching, 30*, 821-844.

Rieber, L.P. (1990). Using computer animated graphics in science instruction with children. *Journal of Educational Psychology, 82*, 135-140.

Rieber, L.P., & Parmley, M.W. (in press). To teach or not to teach? Comparing the use of computer-based simulations in deductive versus inductive approaches to learning with adults in science. *Journal of Educational Computing Research.*

Rieber, L.P., Boyce, M., & Assad, C. (1990). The effects of computer animation on adult learning and retrieval tasks. *Journal of Computer-Based Instruction, 17*, 46-52.

Rivers, R.H., & Vockell, E. (1987). Computer simulations to stimulate scientific problem solving. *Journal of Research in Science Teaching, 24*, 403-415.

Showalter, V.M. (1970). Conducting science investigations using computer simulated experiments. *The Science Teacher, 37*, 46-50.

Shulman, L.S., & Keisler, E.R. (1966). *Learning by discovery.* Chicago: Rand McNally.

Shute, V. J. (1991, April). *A comparison of learning environments: All that glitters* .... Paper presented at the American Educational Research Association (AERA) Annual Meeting, Chicago, USA.

Shute, V.J., & Glaser, R. (1990). A large-scale evaluation of an intelligent discovery world: Smithtown. *Interactive Learning Environments, 1*, 51-77.

Swaak, J. (1995). Development of tests for intuitive knowledge in formal domains: the WHAT-IF test. Internal IST-MEMO-95-03, Enschede: University of Twente.

Veenman, M.V.J., & Elshout, J.J. (1995). Differential effects of instructional support on learning in simulation environments. *Instructional Science, 22*, 363-383.

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

White, B.Y. (1984). Designing computer games to help physics students understand Newton's laws of motion. *Cognition and Instruction, 1*, 69-108.

White, B.Y. (1993). ThinkerTools: causal models, conceptual change, and science education. *Cognition and Instruction, 10*, 1-100.

Universiteit Twente
universiteit voor technische en maatschappijwetenschappen