Intelligent support for discovery learning

Koen Veermans

2003

Ph.D. thesis
University of Twente

Twente University **Press**

Also available in print:
http://www.tup.utwente.nl/catalogue/book/index.jsp?isbn=9036518377

Intelligent support for discovery learning

ico

This research was carried out in the context of the Interuniversity Centre for Educational Research

Twente University **Press**

Publisher: Twente University Press, P.O. Box 217, 7500 AE Enschede, the Netherlands, www.tup.utwente.nl

Print: Océ Facility Services, Enschede

ISBN 9036518377

INTELLIGENT SUPPORT FOR DISCOVERY LEARNING

USING OPPORTUNISTIC LEARNER MODELING AND HEURISTICS TO SUPPORT SIMULATION BASED DISCOVERY LEARNING

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de rector magnificus,
prof. dr. F.A. van Vught,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
op donderdag 9 januari 2003 te 15.00 uur

door

Koen Huibert Veermans

geboren op 19 april 1968
te Geleen

Dit proefschrift is goedgekeurd door de promotor:
prof. dr. A.J.M. de Jong

en assistent-promotor:
dr. W.R. van Joolingen

Acknowledgements

In the first chapter of this thesis I refer to the feeling of excitement that can accompany the discovery of something 'new'. During the course of working on this thesis I have experienced this feeling of excitement on numerous accounts, and more often than not found out that it was not really ‘new’. The overall feeling is, however, a positive one, and for a large part this is due to the people that have, in one way or the other, contributed to this thesis.

I would like to start thanking Ton and Wouter for giving me the opportunity to do this research, and for sharing their ideas, criticisms, enthusiasm and time. Their support has been important throughout, and indispensable towards the end.

The experiments that are described in this thesis would not have been possible without the participation of the schools and their students. I would like to thank Imme de Bruin for providing the contacts with the schools, and the board, the teachers, and the students of CSG Het Noordik in Almelo, OSG de Bataafse Kamp in Hengelo, Vechtdal College in Hardenberg, and Etty Hillesum Lyceum in Deventer for their willingness to participate in these experiments.

My special thanks go to all the members of ProIST over the years. These meetings have always been a source of inspiration, apart from being a pleasure and joy. Some of the members are still around, some have been gone for years, new members took their places, but none of them will ever be forgotten.

I would also like to express my gratitude to the secretaries, the people of the TO-lab, and the members of the SERVIVE-project for all the work that made my life easier.

I would also like to thank my parents for the support that they have given me over the years. My thanks also go to my friends, for showing interest, and for listening to the ups and downs of my research.

Finally, I would like to thank the ICO for organizing the summerschool of 2000 in Rhodes, where I met Marjaana. The importance of her presence and help during the last part of writing the thesis are something that cannot be expressed in words.

Table of Contents

1

Introduction

Anyone, who has ever experienced the feeling of excitement that can accompany discovering something "new", may recognize the potential of discovery learning. "New" in this sentence relates to the person who made the discovery. This may become apparent after sharing the discovery with someone, who points out that this discovery has been known already to mankind. While this might take away some of the excitement, in general the remaining overall feeling is a positive one. The potential of discovery learning can be realized by the discovery itself: learning domain knowledge, and ideally evoking a certain degree of excitement, and by the discovery process: learning discovery learning skills, thus enhancing the chances of a discovery to occur in another place and another time through transfer of the discovery skills. The research presented in this thesis tries to contribute to realizing the potential of discovery learning.

The main research topic of this thesis is support for learners in simulation-based discovery learning environments; learning environments that contain a simulation of some phenomenon in which the task for the learner is to discover the underlying principles of the simulated phenomenon. A major issue in simulation-based discovery learning is how to support learners in successful learning (de Jong & van Joolingen, 1998). A common approach to resolve this issue is to augment the simulation-based learning environment with *cognitive tools*, that is, components in the learning environment that support the learner by offering information, by externalizing learning processes, or by structuring the task (Lajoie, 1993; van Joolingen, 1999). For example, the learning environment may contain explanations that provide learners with information, dedicated notebooks that allow learners to take structured notes, and forms that prompt learners to make explicit what they are doing (van Joolingen & de Jong, 1997; Shute & Glaser, 1990). These tools can help the learner in learning from the simulation. In general, however, they are not designed to respond to the actual learning process that is going on in the learning environment.

Learners working in a simulation-based discovery learning environment need to employ a number of learning processes, such as, *orientation*, *hypothesis generation*, *experimentation*, and *evaluation* (e.g. Njoo & de Jong, 1993). These learning processes are reflected in *activities* in the learning environment, in interaction with the simulation, or with other elements of the learning environment. These activities include setting values of variables to do experiments, plotting values of variables to interpret results, answering questions, formulating hypotheses, and many more. These activities form the starting point of this research. The idea is to interpret the learners' activities and to use this interpretation to provide support for learners. The assumption is, of course, that learners will benefit from this support, which by nature will be more tailored to the actual behavior of the learner.

The purpose of interpreting learners' activities is to generate a *model* of the learner through abstraction of information from the activities of the learner. Modeling the learner in discovery learning can be a goal in itself (van Rijn, 2003), but in this context the goal is to use the model of the learner to support learners.

When starting this research project the idea was that the model of the learner could be used to adapt the learning environment by setting out a trajectory for the learner. This view was expressed in the design of an experimental study. The idea of that study was to compare the effectiveness and efficiency of a condition in which learners will be guided by assignments triggered by a learner model, and a condition in which the assignments are simply available for selection by the learner.

However, during the research project this view on the potential use of a model of the learner changed. Using the model for instructional planning would result in a learning environment where the learner would lose insight and control over the learning process. The basic idea of a learning environment that dynamically plans the instruction would prevent learners from getting an overview of the trajectory that lies ahead, and the progress that has been made so far. Having such an overview can be considered important for learners in a situation in which they are expected to regulate their own learning, because it represents a framework that can be used as a basis for the construction of knowledge. If this overview is kept from the learner, the learner does not have the framework, and will have to create, extend and/or adjust such a framework during the course of learning, making it harder to maintain a consistent representation.

These considerations yielded a changed view on the role for the learner model. Instead of planning for the learner, the role is to generate support in the form of advice on the discovery learning process. This also changed the view on the contents of the learner model. For the purpose of providing advice, the learner model did not need to keep a full and precise representation of the learner's knowledge. Instead, it can be restricted to a model that contains just enough information as needed to provide advice about the discovery process to the learner.

1.1 Relevance of the research

Apart from the relevance that lies in the general potential of discovery learning (discovery and transfer), the relevance of the research can also be defined in a more practical level in relation to the changed views on science education in schools. The US National Science Education Standards propose a change in emphasis from knowing scientific facts and information towards understanding concepts and developing abilities of inquiry. Along the same lines, a recent reform in the Dutch school system, implemented under the name “Het studiehuis”, changed the situation in the schools in a way that there is less emphasis on frontal expository teaching, and more emphasis on independent learning. The implementation of this reform has not been without problems partly because materials and activities for independent learning where not readily available. The kind of learning environments that were used in this research can fit in this new situation. They can be used along with expository teaching, and provide a different approach towards learning that can be viewed as a valuable extension of the curriculum.

1.2 Purpose of the research

The research in this thesis consists of two main parts. The first part focuses on the design and implementation of a tool to support learners in a simulation-based discovery learning environment, which builds upon the information that can be extracted from learners’ activities. The second part studies the effects of these tools on learners working with these environments. These effects can be subdivided into effects on the way that learners interact with the learning environment, and effects on the learning outcomes of working with the learning environment.

The use of tools in performing a task inherently transforms the nature of the task (Hutchins, 1995; Norman, 1993). If the nature of a task is considered important, as is the case in discovery learning, this notion should be taken into account in the design of tools and in the evaluation of tool use. For a tool to be used in discovery learning, this means that it should be designed in a way that the discovery nature of the activities can be maintained. It should be assessed whether the discovery nature of the activities is indeed maintained, or that the nature of the activities changed. If the latter is the case, both potentials, discovery and transfer might be lost.

This assessment is closely related to the dependency between the effect on the interaction of the learners with the learning environment, and the effect on the learning outcomes. If there is little effect in terms of outcomes, the interaction with the learning environment is most probably also of little value, since appreciation and self-assessment of the interaction are closely related to the learning outcomes.

1.3 Overview of the thesis

The central question for the research is:

> Can we develop a tool that supports learners in the process of discovery learning in a simulation-based learning environment, based on their individual interaction with the learning environment?

The theoretical framework is described in Chapter 2. The chapter starts with a description of discovery learning, defining the term, identifying processes that can be used to describe discovery learning, describing problems that have been identified with these processes, and describing different approaches of supporting learner's experiencing these problems. It then proceeds with a description of Intelligent Tutoring Systems (which will be referred as ITS in the rest of the thesis) a line of research that focuses on providing individualized support to learners. Given the basic assumptions of discovery learning, the ability to provide individualized support can be considered a potentially fruitful extension of these learning environments. However, it turns out that exploiting this potential is not straightforward, because the conditions for providing this individual support seem to be at odds with the conditions for discovery learning. The final part of the chapter focuses on heuristics. It will be argued that heuristics can provide a basis to guide the general design of support for learners, and to exploit the potential of individualized support offered by ITS.

Chapter 3 starts with refining the context, by introducing SIMQUEST, an authoring system for developing simulation-based learning environments, and based on the refined context the research question is redefined. The rest of the chapter describes two versions of the tool that were developed during the course of this research.

The Chapters 4 and 5 describe experimental studies that were conducted with the two versions of the tool. Chapter 4 describes the design and results of a study in which a learning environment with the first version of the tool is compared to a learning environment without this tool. Chapter 5 describes a study in which the second version of the tool was used in combination with a general design of the learning environment guided by heuristics. It compared a learning environment that leaves the heuristics that guided the design implicit to the learner, with a learning environment that explicitly communicates the heuristics to the learner.

The final chapter, Chapter 6, reflects on the results of the two experimental studies and interprets these results by going beyond the conclusions of the individual studies. The last part of this chapter reflects on the thesis as a whole and presents directions for future research.

2

Theory: Supporting discovery learning

2.1 Discovery learning

Zachos, Hick, Doane, and Sargent (2000) define discovery learning as:

> “the self-attained grasp of a phenomenon through building and testing concepts as a result of inquiry of the phenomenon” (p. 942).

“Concepts” in this definition refer to the various forms of representation of natural laws including hypotheses, models, rules, and principles. The definition highlights a number of important aspects of discovery learning. The self-attained grasp of a phenomenon at the start of the definition in combination with building concepts makes it very clear that it is the learner who constructs knowledge in discovery learning. Furthermore, the definition tells us that the concepts a learner builds during discovery learning need to be tested, and that building and testing of concepts are part of the inquiry of the phenomenon. Based on this definition one could say that discovery is related to the outcome, inquiry to the process. It could be argued that the term used for this type of learning, discovery learning or inquiry learning[1], reveals which of the two is considered more important. No such argument is made here. Discovery learning will be the term used throughout this thesis, but both discovery and inquiry are considered equally important.

Discovery learning has a long history in education (Dewey, 1938; Bruner, 1961), but it has seen resurgence in popularity over the last decade for at least two reasons. One reason is the change in the field of education towards more constructivist ideas about knowledge and learning. Discovery learning with its emphasis on the learner’s knowledge construction fits better within this framework than traditional expository teaching. The other reason is related to the fact that computers became known and available to a wider range of people, and also attracted the interest of researchers in research

[1] These are not the only terms that have been used for the type learning that is meant here. Other terms that have been used in the literature are for instance exploratory learning, or inductive learning.

areas related to education. One of the areas of interest was the development and use of simulations. With the computer it became possible to build simulations of processes in the real world and simulate them on the computer, creating environments that were well suited for discovery learning. These environments, also referred to as simulation-based discovery learning environments, will be our main interest throughout this thesis.

Like discovery learning, the idea of simulation-based discovery learning is that it engages the learner actively in the learning process. In an unguided simulation-based discovery environment learners have to set their own learning goals. At the same time they have to find and apply the methods that help to achieve these goals. There are two main goals that can be associated with simulation-based discovery learning; development of knowledge about the domain of discovery, and development of skills that facilitate development of knowledge about the domain (i.e., development of skills related to the process of discovery).

2.1.1 The process of discovery

The definition that was given at the start of this chapter is very dense in its description of the process of discovery. This section will elaborate on this process, taking input from Philosophy of Science, Artificial Intelligence, and Education, to arrive at a description of the process of discovery that will be in the current context.

In philosophy of science there has been an ongoing debate about what exactly constitutes scientific discovery. In the 19th century Peirce developed a model of scientific discovery based on abduction, induction, and deduction. As Peirce (Goudge, 1950) defines it:

> Induction is an argument which sets out from a hypothesis, resulting from a previous Abduction, and from virtual predictions, drawn by Deduction, of the results of possible experiments, and having performed the experiments, concludes that the hypothesis is true in the measure in which those predictions are verified, this conclusion, however being held subject to probable modification to suit future experiments. (p. 198)

As a reaction to the logical positivism in the beginning of the previous century, philosophy of science has been looking for a *logic* of scientific discovery (cf. Popper, 1959) that leaves no room for logically unsound formalisms. According to Popper there is no logic in abduction and induction. Deduction, which he calls the *context of justification*, is the only area in which the scientific procedure can be studied. The reason is that both abduction and induction are not logically sound methods to establish truth. Because abduction was not logically sound it was moved to the margins of science. To a lesser extent the same thing happened to induction, because one can never be sure whether a new observation will not falsify a hypothesis, and because hypotheses are always underdetermined by data, which means that there will always be alternative theories that can be derived from he same data.

Induction, however, remained part of the scientific method because examples of scientific progress that were based on induction proved its value, even without a sound logical basis. De Groot (1969) for instance, left room for induction when he defined five stages of the empirical cycle: orientation, induction, deduction, testing, and evaluation. De Groot also recognized that not all research goes through all stages. Descriptive or explorative research might even restrict the cycle to orientation and induction only, generating hypotheses rather than testing them. Simon (1973a) also challenged Popper's ideas, stating that there is a logic of discovery. In this article Simon claims to banish the problem of induction with the following definitions.

> A *law-discovery process* is a process for recoding, in parsimonious fashion, sets of empirical data.
> A *normative theory of scientific discovery* is a set of criteria for evaluating law discovery processes. (p. 475)

These ideas were inspired by developments in Artificial Intelligence, and resulted in the development of discovery systems that used induction. One of the first systems was the Generalized Rule Inducer (Simon & Lea 1974). Drawing on earlier work on problem solving (Newell & Simon, 1972) human concept formation was described as a search in two spaces: a space of instances, and a space of rules. Concept formation in this view entails searching these spaces and finding rules that can describe the instances. Klahr and Dunbar (1988) elaborated these ideas in SDDS, a theory of Scientific Discovery as Dual Search, replacing instances and rules with hypotheses and experiments. According to SDDS, hypotheses can be generated based on prior knowledge, or based on experiments.

The inductive reasoning systems that were based on the problem space search soon ran into the fundamental problem of underdetermination of theories by data, and researchers had to look for ways to solve this problem. This led to the introduction of heuristics in the inductive reasoning systems. Examples of these rule finding are systems, such as, AM (Lenat, 1979, 1983), Bacon (Langley, 1981; Langley, Simon, & Bradshaw, 1987; Langley, Simon, Bradshaw & Zytkow, 1987), and Kekada (Kulkarni & Simon, 1990). The goal of these systems was to automate the discovery process, and the heuristics were viewed as an internal part of the system, or as Shen (1990) describes it: "It is crucial for a discovery system to have a productive set of operators, as well as an effective set of heuristics to control the search" (p. 271). Effectively this means that the only heuristics that can be used within this view are heuristics that can be formalized, that can be transformed into production rules in a deterministic generate and test system.

Based on the description of these systems it might be concluded that they have a much more narrow view on scientific discovery than the one that is adopted in de Groot's empirical cycle. This is due to the fact that these systems were first designed to replicate discovery of known scientific laws. The systems were given a pre-defined data set, and the outcomes were compared with the known laws. This actually entails that orientation and conclusion are present, but took place outside the system. More recent discovery systems explicitly acknowledge this distributed nature of the discovery systems, activities like preparing data and drawing conclusions by the researcher are no longer seen as cheats, but as an integral part of the system (Langley, 2000).

The process of discovery learning has also received attention in educational research on discovery learning. Descriptions of the discovery process in this research differ in the names used to identify the processes, and the number of processes that are mentioned. Friedler, Nachmias, and Linn (1990) for instance describe the discovery learning processes as: (a) define a problem, (b) state a hypothesis, (c) design an experiment, (d) observe, collect, analyze, and interpret data, (e) apply the results; and (f) make predictions on the basis of results of previous experiment(s). De Jong and Njoo (1992) describe discovery learning as *transformative processes* including analysis, hypothesis generation, testing and evaluation, and *regulative processes* including planning, verifying, and monitoring. Other authors proposed similar descriptions (Kuhn, Black, Keselman, & Kaplan, 2000; Lewis, Bishay, & McArthur, 1993; White, 1993). In this thesis the following processes will be used: orientation, hypothesis generation, hypothesis testing, conclusion, as well as regulative processes. Below, each of these processes will be described in more detail.

Orientation

During the orientation process learners build their first ideas of the domain and the environment. It might involve reading introductory and/or background information, exploring the domain, identifying the variables in the domain, and relating prior knowledge about the domain to the problem at hand. The activities and the results of the orientation process can be used as input for other processes. Conversely, the activities and results of other processes (especially from the conclusion phase) alter ideas of the domain and might trigger re-orientation of the domain.

Hypothesis generation

In the hypothesis generation process learners start formulating hypotheses about the domain. A hypotheses is a statement about the relation between two or more input and output variables that expresses an idea about the nature of this relation. Hypotheses can be derived from the exploration of the domain or from ideas or other hypotheses about the domain. Deriving a hypothesis from the exploration of the domain can be classified as inductive process, in which an abstraction is made over a number of experiments to arrive at a hypothesis that can account for these experiments.

Deriving a hypothesis from ideas or other hypotheses (for instance through analogy) can be classified as a deductive process, in which existing knowledge is used to create a new hypothesis. This distinction between in inductive and deductive generation of hypotheses coincides with the distinction between experimenters and theorizers that was made in the SDDS framework by Klahr and Dunbar (1988), and the distinction between data-driven and theory-driven discovery systems (Langley, Simon, Bradshaw, & Zytkow, 1987).

Hypotheses can also be generated in relation to both a hypothesis and experiments. For instance, after testing a qualitative hypothesis, a learner could try to generate a more precise hypothesis that also fits the data. Another example would be revising the hypothesis by adding a condition after experiments that show that there are restrictions on the scope of the hypothesis.

Hypothesis testing

The hypotheses that are generated in the hypothesis generation process can not be guaranteed to be correct, and should ideally be tested by the learner. This is the focus of the hypothesis testing process. The learner has to design and execute experiments that put a hypothesis to the test, gather the data from the experiments, and interpret the results. It is essential that the experiment design in this process is set up in a way that evidence that is generated by executing the experiments is suitable for testing the hypothesis.

Conclusion

During the conclusion process the learner should review the hypothesis in the light of the evidence that was generated in the hypothesis testing process. The learner should decide whether the evidence is in line with predictions derived from the hypothesis, or identify discrepancies between evidence and predictions. This may lead to revision of hypotheses and/or the generation of new ones.

Regulation: Planning, Monitoring, and Evaluation

The regulation processes are the processes that manage a learner's movement through the discovery learning processes described above. On a general level the regulation processes keep track on the progress that has been made in the other processes, and the movement between the processes. On a specific level they keep track of the progress that has been made within the processes, and the steps that lie ahead.

Planning involves setting up a goal, and defining a way to achieve that goal. It can be located at the general level defining movement through the processes, or at a specific level, defining steps within a process.

Monitoring is the process that keeps track of the steps and actions carried out within a plan, and their results in a way that the planning and evaluation process can use them.

Evaluation reflects upon both the outcomes of the processes, and the steps taken in processes. Reflection concerns assessing the outcomes of the discovery processes, in relation to their goals. As a result of this assessment the planning might be revised. For instance, if the outcome of the conclusion process is that the evidence is in not line with predictions, the learner has to make a decision about what to do, reformulate the hypothesis or reject the hypothesis. This decision has consequences for planning the next steps. Reformulating the hypothesis implies moving back to the hypothesis generation phase, and generating a new hypothesis based on the old one and the contradicting evidence. Rejection leaves the next step open, and the learner might proceed with any of the other processes. Reflection on the steps taken in the processes concerns assessing the steps in the discovery processes in relation to the goal of the process, and might lead to revision of planning within the processes. It could, for instance mean reflecting on the hypothesis-testing phase, deciding whether the evidence that was generated during this process is “sufficient”, or that additional evidence should be obtained. The latter could lead to a revision of future hypothesis testing plans.

2.1.2 Difficulties with the discovery processes

Learning with simulation-based discovery environments has not always yielded better learning results than expository teaching (for a full review see de Jong & van Joolingen, 1998). Experiments that compared discovery learning with traditional methods do not show a clear picture favoring discovery learning over traditional learning. Positive results were found on some occasions, but on other occasions no differences were found, or even negative results. One reason for these findings is that these studies were usually measuring traditional knowledge only; another reason is that the processes involved in discovery learning appear to be difficult for learners. Learners have been found to experience problems with one or more of the discovery processes.

Problems have been identified in orientation, hypotheses generation, hypothesis testing, conclusion, and regulation of discovery learning.

Orientation can be problematic if learners have relatively little prior knowledge about the domain. This can result in difficulties with regard to identifying variables and potentially interesting relations between these variables.

Problems with hypothesis generation have been identified in a number of experimental studies. Njoo and de Jong (1993) found that learners had problems stating syntactically correct hypotheses. In their study subjects’ hypotheses were syntactically correct in less than half of the cases. Dunbar

(1993) found that learners find it hard to state an alternative hypothesis, and as a result, stick to their current hypothesis. Learners have also been found to be “cautious” towards stating hypotheses, stating general hypotheses that have only a small chance of being rejected (Klayman & Ha, 1987; van Joolingen, 1993).

A problem related to hypothesis testing is that experiments designed by learners often are not suitable to test a hypothesis. This is typically the case when learners change many variables at the same time. A number of researchers (Reimann, 1991; Shute & Glaser, 1990; Tsirgi, 1980) found that learners varied to many variables over experiments making the interpretation of outcomes virtually impossible. Another problem is that learners design experiments looking for evidence that supports their hypothesis, and not for evidence that might disconfirm the hypothesis (Dunbar, 1993; Quinn & Alessi, 1994).

Drawing conclusions about the hypothesis in relation to the experimental evidence can also cause problems. For instance when the experiments are not suitable to test a hypothesis, but are regarded by the learner as such. Even for suitable experiments it has been reported that learners draw conclusions which can not be substantiated by the evidence (Klahr & Dunbar, 1988; Kuhn, Schauble, & Garcia-Mila, 1992; Schauble, Glaser, Duschl, Schulze, & John, 1995). Klahr and Dunbar (1988), for instance, found that learners failed to draw the right conclusions from disconfirming evidence for a hypothesis. In the same study they also found a reversed effect: rejecting a hypothesis without evidence supporting the rejection.

A problem related to regulation is what is sometimes referred to as “floundering” behavior, experimenting with the simulation without having a clear idea of what has been done, and what to do next. This floundering behavior can be seen as problems with respect to planning, monitoring, and evaluation. Another problem with regulation is adoption of what Schauble, Klopfer, and Raghavan (1991) refer to as “the engineering approach”. In this approach learners try to achieve a certain desirable outcome, focus on variables that are expected to have a positive influence on the desired outcome, and as a consequence are only exploring part of the domain.

Some problems with regulation are related to the problems that have been reported for the other processes. For instance, the problem of designing inconclusive experiments can be seen as a regulation problem related to deciding whether the experimental evidence that was generated is sufficient to grant a conclusion about a hypothesis. Not using all evidence, but only confirming evidence in drawing a conclusion can also be seen as a problem related to monitoring.

2.1.3 Support for discovery learning processes

As can be seen from the description of the processes (and from the philosophical discussion on scientific discovery), discovery learning is an illdefined rather than a well-defined process, following Simon's definition (Simon, 1973b). There is no fixed linear sequence between the processes, transitions from one process to another can be in the order as they were described, but are not restricted to this order.

More importantly, there are no fixed procedures that define the steps within a process or rules that define when a process ends. The reason is that the processes can be separated neither from the context in which they occur, nor from the person who carries them out.

As Glaser, Schauble, Raghavan, and Zeitz (1992) pointed out, the structure of a task influences the role of learning processes in the discovery of relations in the domain. They examined the differences between "good" and "poor" learners in three different environments: Smithtown (economics), Voltaville (electric circuits), and Refract (optics). Within the Smithtown domain changes in dependent variables co-vary with changes in independent variables. The learner's task is to assess relationships in the domain in a qualitative manner. The learner needs to find out which variables are involved in a certain relationship, the direction of this relationship and identify which variables are irrelevant. This requires careful experimenting controlling for influences of other variables than the focal variable. Voltaville differs from Smithtown in that the laws that have to be discovered involve all the variables in the domain. Therefore, it is not necessary to design carefully controlled experiments that distinguish between relevant and irrelevant variables. In Voltaville it is sufficient to generate a hypothesis that fits all the experiments that have been conducted. As a consequence, hypothesis testing only plays a minor role in this environment. Important differences between good learners and poor learners were located in different processes for the three learning environments, a finding that Glaser et al. (1992) attributed to differences in the structure of the domain.

The learner involved in the discovery processes also influences these processes. A learner that can be characterized as an experimenter might, for instance, start spending a long time in the orienting process, whereas a theorizer might move to hypothesis generation rather quickly. A learner's knowledge has also been shown to influence the processes that occur in discovery learning. Schunn and Anderson (1999) investigated experiment design skills comparing domain and task experts (researchers from the same field), task experts (researchers from a different field), and task novices with domain knowledge (students form the same field). The results showed clear differences among the groups in the quality and the complexity of their experiments. Domain and task experts designed for instance more complex experiments than task experts with less domain knowledge did. Another finding was that both types of task experts were better with respect to interpretation of the outcomes and drawing conclusions compared to the non-experts. These findings suggest that the task experts were designing experiments to test hypotheses, in a way that they could interpret the results of the experiments in the light of these hypotheses. Their domain knowledge allows domain experts to set up more complex experiments while still being able to interpret the results. Differences have also been found in studies with a more homogeneous population. In most of the studies that identified problems with the discovery processes, differences were found between individual learners. In general, it can be said that successful learners experience fewer problems with the discovery processes.

These notions and the problems that have been identified to appear in discovery learning resulted in design of learning environments that provide support for learners on the discovery processes. The next sections give an overview of support for the different discovery processes.

Support for orientation

Support for orientation process has been provided in simulation-based discovery environments in a numbers of ways. The most obvious way that orientation is supported is through the simulation itself. Defining the model of a simulation restricts the variables of the domain to these variables that are present in the simulation. Learners therefore only need to orient themselves on these variables, and the role they might play in he domain. Model progression (White & Frederiksen, 1989, 1990) can take the support one step further, by defining a series of simulations that progress form simple to more complex; learners are scaffolded in a stepwise orientation on the variables that are part of the simulation model. Another way to support orientation that is frequently used in simulation-based discovery learning is to provide access to domain knowledge. This can be done by providing definitions of the concepts that are used in the simulation (Glaser, Raghavan, & Schauble, 1988; Shute, 1993).

Support for hypothesis generation

Support for hypothesis generation has also been included in many simulation-based learning environments. One form of such support is the use of pre-defined lists of hypotheses (Njoo & de Jong, 1993). Another form of support is a hypothesis menu (Shute & Glaser, 1990) or a hypothesis scratchpad (van Joolingen & de Jong, 1991, 1993) that provides the learner with a list of variables, and a list of possible relations between these variables that can be used to formulate a hypothesis. Yet another form is to force learners to write down one or more hypotheses before they start experimenting (Quinn & Alessi, 1994).

Support for hypothesis testing

Support for hypothesis testing can be subdivided into support for generating predictions, experiment design, and data interpretation. Support for generating predictions usually takes the form of explicitly asking learners to state predictions. Experimentation support can be given through general experimentation hints, like “do not vary too many variables at the same time” and can be given during the process (Lavoie & Good, 1988) or before (Rivers & Vockel, 1987). Experimentation hints can also be specific, describing the precise conditions for the experiment. Data interpretation support can be given in the form of tools that perform curve fitting, or tools that allow learners to draw graphs. Lewis, Stern, and Linn (1993) combine prediction and data interpretation support with a graph in which learners can draw their prediction, upon which the system adds the correct graph as feedback. In Reimann (1991) predictions can be stated numerically, as a graph or less precise as an area in the graph.

Support for conclusion

Support for conclusion usually takes the form of asking learners to draw conclusions and is generally part of an integrated method that addresses all of the processes. The reason is that reviewing a hypothesis in relation to evidence presupposes stating a hypothesis, and generating evidence. Outside an integrative method it is not sure that this situation will occur, and when it will occur, which makes it difficult to provide support.

Support for regulation

Support for the regulation processes can be given on the general level of the discovery learning processes, or at the specific level of these processes.

At the specific level, a notebook facility for storing experiments (Reimann, 1991; Shute & Glaser, 1990) provides support for monitoring these experiments. In a similar vein notebook facilities for hypotheses provide support for monitoring progress in the exploration of the domain on a higher level. Model progression, apart from providing support for orientation, also provides this kind of support. Moving from one level to the other clearly demarcates a form of progress.

At the general level, support for regulation can be provided by presenting a complete method to the learners.

Veenman and Elshout provided learners with a structured working plan that included rephrasing a question regarding the relation between two variables, explicitly stating a hypothesis, work out a detailed action plan, evaluating the experimental outcomes, and drawing a conclusion (Veenman & Elshout, 1995; Veenman, Elshout, & Busato, 1993).

In Thinkertools (White, 1993) learners had to follow the so-called “inquiry cycle” that contained five phases: state a question, make predictions, perform experiments, formulate laws, and investigate the generality of the laws. All phases contained detailed support, but during the course of working with the environment the support gradually disappeared.

Smithtown (Shute & Glaser, 1990), also had a fixed sequence of activities, in which learners had to identify variables of interest, and subsequently were asked to make a prediction before they could do experiments.

Support in discovery learning remains an issue of debate. Not so much whether or not it should be included, since the problems learners have with discovery make a strong case for the need of support, but on how it should be included. Supporting learners in discovery learning always affects the learning process. Njoo and de Jong (1993) identified three dimensions that could be used to classify support: non-directive vs. directive, stimulating vs. restricting, and obligatory vs. non-obligatory.

Directive support stimulates the learner to “do” something, whereas, non-directive support does not. An example could be providing assignments that ask the learner to investigate relations between variables, as opposed to providing definitions of investigation, relation, and variables.

Restrictive support constrains the learners in a certain respect, whereas stimulating support leaves the learner free. An example could be presenting learners with a pre-specified hypothesis list as opposed to allowing learners to express hypotheses in natural language.

Obligatory support is support that is forced upon the learner, whereas non-obligatory support leaves the decision of whether or not to use the support up to the learner. An example of obligatory support could be to force learners to state a hypothesis before an experiment can be done. Non-obligatory support could be providing the learners with the possibility to state hypotheses without requiring them to do so.

In order to maximize learner initiative, support in discovery learning should ideally be non-directive, stimulating, and non-obligatory. A problem with this kind of support is that learners, especially the ones that need support most, sometimes fail to recognize or neglect this kind of support. The practical consequence is that support can usually be located more or less towards the other end for one or more of these dimensions. As an example of an extreme case, consider use of the scientific method as *the* method for discovery learning, requiring learners to go through the phases sequentially and providing precise specifications of what has to be done in these phases. Support in this case can be classified more towards the directive, restrictive, and obligatory end of the dimensions. The learning that occurs in such an environment is most likely related more to memorization than to active construction and mindful abstraction. While this may be sufficient for near transfer, where deep understanding is usually not necessary, it is not for far transfer where deep understanding is required (Salomon & Perkins, 1989). Or, as Schauble et al. (1995) put it: "one can not teach science as a set of declarative facts and concepts and expect students to emerge as skillful reasoners with their new knowledge” (pp. 143).

In conclusion it can be said that discovery learning is difficult for learners, and that there is a clear need to support learners in discovery learning. A wide range of problems in discovery learning is reported in the literature, as are many different ways to support learners. A broad distinction can be made between integrative methods that support all processes of discovery learning that do not always maintain the learner freedom, and more specific support that does not always support all processes.

The difference between learners in their proficiency on the discovery processes also indicates a need for individualized support in these learner environments. The next section will discuss Intelligent Tutoring Systems (ITS), a research area that is entirely devoted to instruction adapted to individual learners.

2.2 Intelligent Tutoring Systems

The primary purpose of tutoring is to provide instruction about a certain domain to a learner. As a result of this instruction the learner's knowledge is assumed to change. This change in the learner's knowledge should be reflected in a change in the instruction provided by the tutor. In order to be able do this, the tutor needs to have an idea about the learner's knowledge, the target knowledge, and an idea of how to change the learner's knowledge. This general description of tutoring also applies to computer mediated tutoring.

The notion of what exactly "an idea of the learners knowledge" means, more or less demarcates the developments in computer-mediated instruction and its transition from CAI (Computer Aided Instruction) to ITS (Intelligent Tutoring Systems).

The first CAI systems were developed at a time that behaviorism was the leading paradigm in psychology and education. Tutoring within this paradigm can be described as presenting stimuli to the learner until the right response is given. When this stimulus response pattern becomes stable the learner will have acquired the knowledge that was the target of the tutoring activity. This is a relatively simple definition of the knowledge of a learner, which can be also be tested relatively easy.

Within this frame it is not very difficult to define a computerized version of a tutor. The tutor only needs to present the stimuli that the learner has to respond to, and if the learner responds incorrectly, present it again, and if the learner responds correctly present the next stimulus. In practice this is probably not what happened in even the most blunt first CAI systems, but it sketches an extreme implementation of the behaviouristic view that influenced the first designs of CAI systems, and might help to illustrate that the line from these systems to the later ITS's is a continuum rather than a discrete transition (Shute & Psotka, 1996; Sleeman & Brown, 1982; Wenger, 1987). A step away from this extreme position might be realizing that presenting the same stimulus over and over until the correct response is given, might not be the best strategy in tutoring, and that presenting different, but similar stimuli might be an alternative. The next step, was to leave the behaviouristic framework with its stimulus-response terminology, and switch to the information processing framework, and start using terms like problems and solutions.

Although this might sound like a radical change, changing from one theoretical frame to another, the implications for a computerized tutor do not need to be that big. The answer or end solution still provides clearly demarcated points that can be assessed by the system. The system can compare this answer or end solution with the correct answer, and provide the learner with the correct solution path, if the answer is wrong. Including so-called bug libraries (Brown & Burton, 1978; Burton, 1982) allows the system to provide more specific feedback and differentiate more between learners. These bug libraries describe possible deviations from the correct solution, in relation to misconceptions or “malrules” that cause these deviations. The system attempts to remediate these misconceptions or malrules with dedicated feedback.

With the introduction of the bug-library, the analysis of the system became more fine grained, no longer viewing the solution path as a single entity, but as a collection of steps that have to be taken to reach the solution. What remained is the idea that there is only one correct path.

What did change was that a model of the learner was introduced in the systems, and the term for the systems changed from CAI to ITS. In 1973 Hartley and Sleeman, presented the requirements for an ITS, and these requirements (a domain model, a learner model, and a tutoring strategy) are still used today (Shute & Psotka, 1996). The domain model, or expert model, contains a description of the knowledge that the learner is supposed to learn or use during the course of working with the ITS, and is used as a benchmark to assess the learner’s interaction with the system. This assessment of the learner is then used to construct a learner model.

A learner model is a collection of all the information about the learner that can be used to adapt the teaching system to the individual user. It serves as a base to tailor the instruction plans and generate explanations and feedback to the learner. The design of the learner model is influenced by the future users, the type of model to be produced, the kind of information that can be obtained from the learner, and the way in which the system should interact with the user.

Holt, Dubs, Jones and Greer (1994) suggested that in order to choose an appropriate learner model four questions need to be answered: who is being modeled, what is being modeled, how is the model to be acquired and maintained, and why is the model there? In the following section these questions will be elaborated upon.

Who is modeled?

It is important to have information about the learners that are going to use the system. It is not hard to realize that without this kind of information tutoring becomes very difficult. Using information about potential users can help to restrict the size and complexity of the system.

What should be modeled?

Vassileva (1990) argues that the learner model should contain a model of the knowledge that is subject to changes and a model of characteristics that will remain the same over time. The dynamic model might include knowledge about the domain, and / or skills that will be tutored. The static model might include learner characteristics like a learner's learning preference, general intelligence, age, and so forth.

Theoretically these static characteristics could be of interest for determining the interaction between the system and the learner, in practice however, it is hard to translate this information into different tutoring strategies for different learner characteristics. The dynamic model of the learner's knowledge might differ in terms of granularity. A fine grained model (Anderson, Farrell, & Sauers, 1984; Anderson, Boyle, & Jost, 1985) keeps a detailed model of the learner, that is believed to reflect the actual knowledge of the learner at a particular point in time. In principle the contents of such a model could even be used to predict a learner's behavior. A more coarse model represents the learner's knowledge on a much more general level, for instance, by rating certain skills as good, moderate, or poor.

The kind of information that the learner model should contain depends on how this information is going to be used by the system to adapt its behavior to the learner. It only makes sense to strive for detail up to the level that is needed by the system.

How should it be modeled?

Two types of information that can be obtained from the user can be distinguished: explicit and implicit information. Explicit information concerns all the information that can be directly obtained from the learners' actions. This includes the information about the learner provided at design time, the learners' actions in the application, answers to question asked by the system etc. This explicit information is easy to obtain but one should be careful in asking the learner too much. The system's need for extra information might not be clear to a learner and its questions could unnecessarily disturb the learning process. Implicit information is all the information that is not directly observed from the learners' actions, and is therefore more difficult to obtain. It requires careful observation of the learners' actions and careful analysis of the domain in order to infer what knowledge the learner will have used to be able to perform the actions.

Whether explicit information will suffice to obtain a learner model depends on the way the learner model is related to the domain knowledge and the tutoring goals. A system focused only on whether certain outcomes are obtained will not need information other than these outcomes. If, however, the system wants to give proper feedback on why a certain outcome is not obtained, it needs to infer what caused the failure. Different models of relating the learner model to the domain model have been described in the literature (eg. Holt et al., 1994; Ragnemalm, 1996; Vassileva, 1990; Wenger, 1987). The models differ with respect to the knowledge that is taken into account and the completeness of the domain knowledge that is assumed.

- *Overlay model*: In these models the only knowledge included in the domain model is correct knowledge. Moreover, the knowledge contained in the model is assumed to be complete. This implies that errors according to this model can only be the result of missing knowledge.
- *Extended overlay*: These models differ from overlay models by incorporating incorrect knowledge in the model. The way this is usually done is by extending the domain model with a library of buggy knowledge. The assumption of completeness still holds. Errors can result from missing correct knowledge as well as buggy knowledge.
- (*Advanced) perturbation*: Overlay and extended overlay models do not take into account that learners can possess correct knowledge that cannot be recognized by the system. Because it can not be guaranteed that the domain model that the system uses is complete the learner can always posses correct and incorrect knowledge that is not recognized by the system as such. Perturbation models do not automatically label this unrecognized knowledge as incorrect.

Why should it be modeled?

The rationale behind the learner model is that it can be used to adapt the tutoring system to the individual learner. Any learner model in a learning environment is an abstract representation of the learner based on the learner's interaction with the learning environment. This abstract representation of the learner is used to change the way the environment interacts with the learner under the assumption that this change has a positive effect on the learning outcomes. VanLehn (1988) describes several common uses for a learner model:

- *Advancement*: The system uses the learner model to decide whether the learner has mastered the current topic well enough to be advanced to a next topic.
- *Offering (unsolicited) advice*: Based on the information in the learner model the system prompts the learner with advice when it thinks this is necessary or on a learner's request.
- *Problem generation*: The system uses the learner model to choose problems that are just beyond the learners current capabilities.

The demands on precision and accuracy of the learner model depend on the purpose that it needs to serve. For advancement, the accuracy of the learner model's description of the learner's level of mastery should be very high. Only then the learner model can make the correct decision about when to advance the learner to the next level. If this high level accuracy can not be achieved in the learner model the decisions about advancement will be out of sync with the learner's actual level of mastery and this could easily lead to frustration in the learner.

The purpose of offering unsolicited advice to a learner is to provide the learner with information that might be relevant to the learner in order to achieve better learning results within the learning environment. As advice does not directly alter the instructional planning it leaves more bandwidth for the learner model with respect to the level of accuracy of the learner model. In case the learner model is not accurate and gives advice that is not appropriate the learner can decide to ignore the advice and continue working in the same way.

If the learner model has to serve as a basis for problem generation just beyond the level of the learner's capabilities, the learner model should, again, have a very accurate description of the learner's knowledge. Apart from that, the system also needs to have an accurate description of the problems to be able to select the one that is just above the current level of the learner. If either of these two descriptions is incorrect there will be a mismatch between the selected problem and the learner's capabilities.

The basic ideas and principles behind ITS were outlined in this section. In the next section it will be analyzed to see if and how these ideas can be used to provide support to learners in discovery learning.

2.3 Discovery learning and ITS

Discovery learning environments and ITS are not natural companions. The traditional ITS systems containing learner models, usually cannot be designated as discovery environments in the sense that they do not offer the amount of learner freedom necessary for discovery learning. Although the degree of learner control varies, the stereotypical ITS has traditionally been system controlled. Conversely, simulation-based discovery learning environments usually do not offer learner modeling in the sense of creating a cognitive model of the learner's knowledge. There are two reasons for the latter observation:

- the amount of learner freedom offered in simulation-based discovery environments is so large that a full learner model is beyond the scope of practical application, the number of parameters is simply too large.
- often learner modeling is seen as contradictory to discovery learning and the related concept of constructivism, for which 'measuring' the learners' knowledge conflicts with the idea that each learner builds his or her own representation of the external world (for a discussion on this point see, e.g., Jonassen, 1991).

This may explain why traditional ITS techniques, such as, learner modeling and instructional planning, have not been used very often for discovery learning. Imperative teaching from within the tutoring system is not an appropriate teaching strategy for discovery learning, as it will decrease the learner's freedom. The arguments however plausible they might sound do not provide a fair picture. Firstly, the overview of the support that is provided to learners in discovery learning environments showed that learner freedom is not always fully preserved in these environments. Dismissing the use of ITS techniques in discovery environments based on this argument should consequently also dismiss these means of providing support to learners. Secondly, the argument that learner modeling conflicts with construction of knowledge is less problematic than it appears at first sight.

Influenced by problems with the traditional approach towards ITS, like the difficulty of handling uncertainty, and new directions in the field of education, the view on ITS has changed (Holt et al., 1994). This move, away from systems with detailed learner and expert models that aim at remediating "incorrect" knowledge creates opportunities to use methods from ITS in a less directive way. Instead of modeling the domain knowledge that the learner is supposed to acquire while working with the ITS, modeling could focus on the processes that are expected to lead to the acquisition of that knowledge. The role for learner modeling in discovery learning could be to support the learner in the *process* of discovery. To do this, the system would not need to maintain a full cognitive model of the learner's domain, knowledge, only to infer just enough from steps taken by the learner to support the learner in the process of discovery.
Following Self (1990) a pragmatic approach is advocated in which:

- interactions are designed in a way that the information needed by the system is provided by the learner
- the contents of the system's model of the learner are linked to specific instructional actions
- a collaborative, advisory role for the system is adopted

The reason to adopt an advisory role is that the other roles, advancement and problem generation, are not in line the ideas behind discovery learning. Discovery learning can be seen as an open domain in which there are no explicit criteria for success (Barnard & Sandberg, 1994). This makes the expert model limited in its scope; it can not contain a full description of the path learners should follow from beginning to end. In a way a formal expert model of discovery is constrained to the Popperian view on science, since it is only within this view that logically sound conclusions can be drawn, which can be compared to the learner's conclusions.

As a consequence ITS should give up control in favor of the learner. Where traditional application of ITS can be characterized as directive, restrictive, and obligatory, a different approach is needed for application in discovery learning. This implication has an important advantage for the learner model. Since the role of the system is collaborative, the validity of the learner model is less critical. The aim will, therefore, not be to establish a complete model of the learner's domain and discovery knowledge but a model that contains just enough information as is needed to assist learners in their discovery processes. The system's role will be restricted to providing the learner with advice that addresses certain parts of the processes that are difficult for learners, and the learner should feel free to discard the advice. The support may be directive, should if possible be simulative, and by no means obligatory.

In the next section heuristics will be introduced, and it will be argued that heuristics can fulfill a role in providing this kind of support, and at the same time extend the expert model of discovery beyond the Popperian view.

2.4 Heuristics

Heuristics have a long history dating them back to their Greek origin. In modern times Polya (1945) brought them back to the attention in his books on mathematical problem solving. Polya (1971) argued that heuristics are hard to define, in part because heuristics are interdisciplinary, and are of interest to other areas of science like psychology, computers, and education. In the years that followed these areas indeed took interest in heuristics.

2.4.1 Defining Heuristics

Influenced by the view that scientific reasoning should be rational, researchers on human reasoning in cognitive psychology started a line of research that investigated the human ability to reason according to logical rules or bayesian statistics. This led to a series of publications in which humans performed rather poorly when measured according to these formal standards (Kahneman & Tversky, 1973; Mynatt, Doherty, & Tweney, 1977; Tversky & Kahneman, 1974; Wason, 1960, 1983). Heuristics became equated with a fallacy that leads to incorrect decisions that should be replaced with formal methods from logic or probability theory. Based on this research the notion of heuristics became negatively connotated within psychology.

Recently, in psychology there is also research going on that explicitly focuses on the virtues of heuristics. This research deals with similar topics as the early research in psychology, but inspired by the work of Simon (1956, 1990), their approach is different. The researchers from psychology approached the problem from a logical end, concluding that humans are using inferior heuristics. The research by Gigerenzer and the ABC research group (Gigerenzer, Todd & the ABC group, 1999; Gigerenzer, 2000) starts from the idea that humans are very profound at making decisions in situations with limited time, knowledge and computational power. The key to that success according to them can be found in the use of heuristics for guiding searches, stopping searches, and decision making. These heuristics are not necessarily based on logic or probability theory, their function is not to be coherent, but to enable making quick reasonable decisions in situations where there is no optimal solution, or where pursuing this optimal solution requires a lot of effort and / or time.

The differences in the areas that dealt with heuristics and their appreciation of the heuristics enable that numerous definitions of heuristics exist. Some definitions do not specify heuristic at all ("heuristic method or procedure"), or are too general to be of any use ("serving to discover or find out"), others only address the use of heuristics in one specific area of science. For the present purpose the following definition will be used:

> A rule of thumb, simplification, or educated guess that reduces or limits the search for solutions in domains that are difficult and poorly understood. Unlike algorithms, heuristics do not guarantee optimal, or even feasible, solutions and are often used with no theoretical guarantee. (The Free On-line Dictionary of Computing, © 1993-2001 Denis Howe)

This definition covers some important characteristics of heuristics. First and foremost, they can serve as a means to make a decision about a problem without the need for a complete and exhaustive analysis of the problem and the context. This characteristic is especially important in situations were there are no established methods of analysis to arrive at an indisputable decision. Even if such methods do exist, the application of heuristics might still be a viable option, since their application usually saves a lot of time. An additional advantage is that heuristics can even be used without a complete understanding of the origins of the heuristic. This can be illustrated with a heuristic, known as Occam's Razor, that says: "if you have two theories which both explain the observed facts then you should use the simplest until more evidence comes along". The problem in this heuristic is the definition of simplest. One commonly used criterion to decide which of the theories is the simplest is Kolmogorov complexity. Use of the heuristic is of course not restricted to people who know how to compute the Kolmogorow complexity for two theories. Instead of Kolmogorow complexity, people use other heuristics (heuristics about simplicity), to decide which of the theories is simpler.

A disadvantage is that using heuristics might lead to a wrong decision. This brings us directly to the next characteristic, that heuristics are not guaranteed to lead to success. This characteristic can be illustrated with a heuristic related to experimental design: "vary one thing at a time" (Tsirgi, 1980). The idea behind this heuristic is that changing only one thing at a time provides a sound basis for interpreting the outcomes of experiments. Differences in the outcomes of the experiments can be attributed to the one thing that was changed. This heuristic is generally considered to be important in experimentation (Schauble et al., 1995). However, as Zohar (1995) argues, it does not work in situations with interacting variables. In such a situation it is necessary to look at situations in which more than one variable is changed as well. This does not mean that the heuristic is "wrong" or worthless. On the contrary, in order to notice an interaction effect it is necessary to have a good idea about the effects of the individual variables, and "varying one thing at a time" is a good way to get a good idea about these effects. What it does mean is that "vary one thing at a time" is not a "golden bullet", but a heuristic. It can be used to make decisions in experiment design but it can not be guaranteed to lead to the best choice in every situation.

Another important characteristic of heuristics is the scope of the heuristic (i.e. the range of situations in which the heuristic can be applied). Two dimensions can be identified that influence the scope of the heuristic.

The first dimension that can influence the scope is the domain dependency. *Domain independent* heuristics have a broader scope than *domain dependent* heuristics. A heuristic like "keep records of what you are doing" is for instance more domain independent than "if choosing the third value of a variable, then choose an equal increment as between the first and the second value" which only applies in domains with continuous variables.

The second dimension that influences the scope of a heuristic is the generality. *General* heuristics have a wider range of situations in which can be applied than *specific* heuristics. A heuristic like "simplify the problem" is fairly general, whereas "construct an analog problem with less variables" is more specific and, therefore, more restricted in use.

2.4.2 Heuristics and discovery learning

There are two views on scientific discovery that are important in the discussion of support for discovery learning and the role that heuristics might play in this support. The formal view deals with the *logic* of scientific discovery: when should a hypothesis be rejected? Which predictions can be derived from a hypothesis? Which logical steps can be taken in a process that preserve truth? Alternatively, the *productive* view deals with questions like: what hypothesis to state next? Which experiment to design to test the hypothesis? Heuristics can have a role in both views.

In the first view they can be used to introduce the formal view on scientific discovery before presenting the formal *logic* behind this view. This use of heuristics stretches the definition of "domains that are difficult and poorly understood" from the universal context, to the personal context of the learner. The logic of discovery is both difficult and poorly understood by learners, and learners can not be expected to learn and understand the logic of discovery through expository teaching. In discovery learning learners can experience the logic of discovery in a context. Heuristics can provide guidance to learners during these experiences, and provide a basic informal structure that can later be transformed into a formal structure. The heuristics form a scaffold that the learner can gradually replace with a more formal understanding.

In the second view they can be used as a set of “rules” that constitutes “good practice”. When used in this sense heuristics can provide support for the parts of the discovery process that are not well defined, and that should therefore not be taught as if they were. They also provide a “hook” for the combination of ideas from ITS and discovery learning as they can extend “the expert model” of discovery learning beyond the logic of discovery. Productive heuristics can be included in the expert model, not as a prescriptive model of correct behavior (that is used to correct a learner) but as a descriptive model of good practice. They can be used to provide the learner with advice, triggering reflection on the learner’s own practice.

This more productive view on heuristics should also be reflected in education. In education for too long emphasis has been on teaching procedures that lead to solutions, while neglecting uncertainty in the search for these solutions. Heuristics provide the possibility to support problem solving, or discovery learning, while at the same time highlighting this uncertainty.

The characteristics of heuristics make them well suited to support learners in discovery learning. The uncertain nature of heuristics implies that they should not be used in an obligatory way. Their scope makes that they can provide more or less direction for learners, and allows learners to relate the heuristics to their own existing knowledge. Thinking of heuristics in this way also makes them stimulating.

3

From Theory to Practice: Design and implementation of the support

In Chapter 1 the central question for this thesis was defined as: Can we develop a tool that supports learners in the process of discovery learning in a simulation-based learning environment, based on their individual interaction with the learning environment?

Two constraints have to be taken into account in the design of the support. The first constraint is related to the nature of the support as it was discussed in the previous chapter. The idea behind simulation-based discovery learning is that learners construct knowledge through the exploration of the domain and during the process also develop discovery learning skills. To maintain the exploratory nature of the environment, the support may be *directive*, should try to be *stimulating* and must be *non-obligatory*. In other words, the support should leave room for the learner to explore. The second constraint is related to the context in which the support should be operating, SIMQUEST, an authoring environment for simulation-based learning environments. In an authoring environment the domain will not be known in advance, therefore, the support cannot rely extensively on domain knowledge. In the next section this context will be described in more detail. Then the question will be further specified towards the specific context of the research. Furthermore, the design an implementation of two versions of a support tool will be presented as an attempt to answer the redefined question.

3.1 The SIMQUEST authoring environment

The SIMQUEST authoring environment (de Jong, van Joolingen, Swaak, Veermans, Limbach, King, & Gureghian, 1998; van Joolingen, King, & de Jong, 1997, van Joolingen & de Jong, in press) is an authoring environment to design and develop simulation-based learning environments. The main focus and intent of the authoring environment is on conceptual domains and the acquisition of conceptual knowledge (e.g., laws in physics), but in principle it can also support development of simulation models for procedural (operational) domains.

Simulation models form the core of the learning environments created with SIMQUEST. Authors can present different views on the domain to the learners by using more than one simulation model in the learning environment. They can use models that gradually increase from simple to complex, zoom in on parts of the model, or present different representations of the model. These ideas, known as *model progression* (White & Frederiksen, 1990) support learners on the regulative aspects of the learning process, by demarcating different models, structuring the environment, and presenting an overview. Technically model progression is facilitated through the use of different model contexts and/or the use of different *simulation interfaces*. Each model context has one model associated with the context, and will appear as a separate model progression level in the learning environment. Each simulation interface can present the learner with a different view of the model, for example, by providing a more restricted view on the model or by providing a different representation of the model (numerical, graphical, or animated).

Authors can also use so-called *instructional measures* to provide support for the learners on the discovery processes. The SIMQUEST authoring environment includes several types of *instructional measures* that authors can readily use in designing their learning environments. One way to provide support is by using *assignments*, small tasks or exercises that provide the learners with subgoals that are within reach. These assignments can support regulation and monitoring processes, but can, for instance, also take over the hypothesis generation from the learners by presenting the learners with a research question and possible hypotheses related to this research question. *Explanations* can be used to provide the learner with just-in-time information in relation to the assignments. Finally, there is a *monitoring tool* that allows learners to store and organize experiments, thus, providing support on monitoring the experiments. The next sections will shortly describe the instructional measures, and the control structure that allows the author to define the level of system control.

3.1.1 Assignments

Within the SIMQUEST authoring environment different types of assignments are available to the author. A common feature of all assignments is that they have a goal, and through that goal the assignment supports the regulative learning processes. The difference between the assignments is in the focus of the goal, the amount of implicit support that is provided, and the way the learner can achieve the goal. Assignments can also bring the simulation in a certain state that serves as a starting point for the pursuit of the assignment’s goal.

Do-it assignments present the learners with a goal and can set up a specific situation in the environment. The responsibility over the process of achieving this goal is left to the learners.

Open answer assignments are similar to do-it assignments; the only difference is that in these assignments the learners are asked to write down their answer, ideas or conclusions about the goal. This addition might trigger self-explanation / reflection in the learners, when they have to write down an answer.

Do-them assignments are also similar to do-it assignments, but here the author can specify additional situations. These additional situations provide support for experiment design, where different situations can be seen as experiments.

Investigation assignments (Figure 3-1) can be viewed as do-it assignments that are extended with hypothesis generation support. The goal is to investigate a relation between variables, and the answers can be seen as possible hypotheses about that relation. The learner can choose between hypotheses and after having made a choice they will receive feedback that was specified in advance by the author.

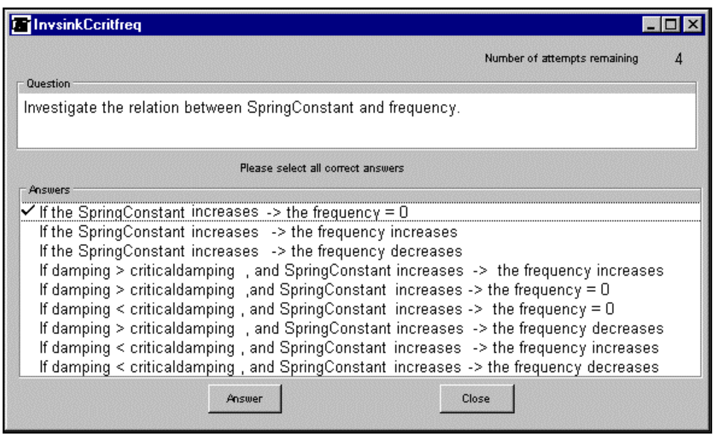


Figure 3-1. An investigation assignment in the SIMQUEST environment.

Explicitation assignments relate to investigation assignments just like do-them assignments relate to do-it assignments. They extend the investigation assignment with a set of experiments. The learner can run these experiments and see the impact of these experiments, choose a hypothesis, and receive the feedback that was specified by the author.

Specification assignments have a somewhat different goal focus. The focus of the other assignments is *inductive* (finding out something about the domain), but in this assignment it is *deductive* (predicting based on knowledge about the domain). The relation is supposed to be known, and the learners are asked to use this knowledge to predict an outcome for one or more variables in a specific situation. The learners will receive pre-defined feedback on the prediction(s).

Optimisation assignments are similar to specification assignments, but the task of the learner is reversed. In the specification assignment the conditions are specified, and the results have to be predicted. Here, some end result is specified, and the learner has to change the conditions in a way that this end result will be obtained. The learner will receive pre-defined feedback when violating conditions or when the end result is not according to the target.

3.1.2 (Feedback) Explanations

Explanations can contain audio, video, text, html, images, or a combination of text and images. They can be used to provide feedback, but also to provide background information (Figure 3-2) about the domain or the learning environment.

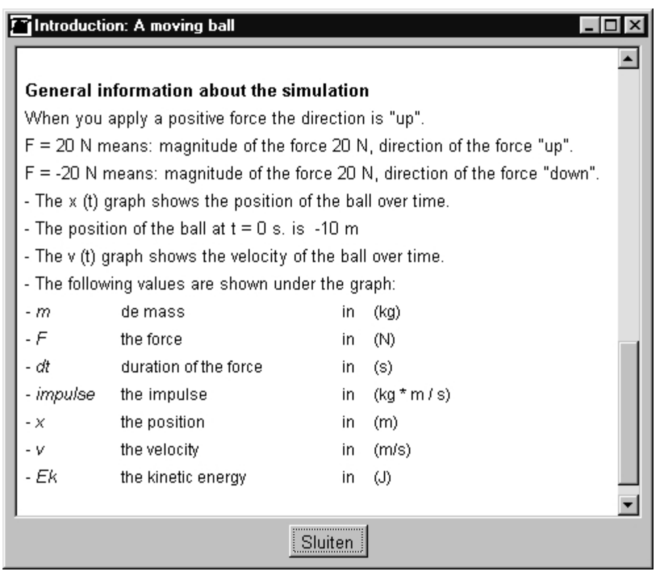


Figure 3-2. An explanation with background information about the simulation.

3.1.3 Monitoring tool

The monitoring tool (Figure 3-3) is comparable to the notebook facilities that were described in Section 2.1.3. It supports the learners in monitoring their experiments with the simulation. They can store experiments in the tool, which then presents the values of the variables in a table format. They can later replay experiments if they want to see them once more, or sort variables to compare different experiments.

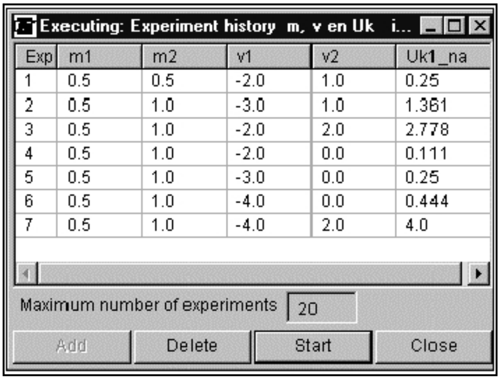


Figure 3-3. A monitoring tool with experiments.

3.1.4 System vs. learner control in SIMQUEST learning environments

The balance between system control and learner control in the interaction between the SIMQUEST learning environment and the learner is specified by the author. The author does this by making use of the control mechanism. The control mechanism determines when instructional measures present themselves to the learner. Figure 3-4 gives an overview of the basic control and data flow within the system.

The control is specified in the instructional measures. Instructional measures can be invisible or visible for the learner, and visible instructional measures can be disabled, enabled, or active. An enabled instructional measure can be activated by the learner or the system, at which point it will open its associated window. For some of the assignments the state can change into succeeded (correct answer) or failed (no correct answer after a specified amount of attempts), and for all of the instructional measures into exited (when the window is closed). The author can specify actions for each of the states (activated, succeeded, failed, and exited) that alter the state of other instructional measures. With the exception of succeeded and failed, all states can be set. This means that one instructional measure can hide, show, enable, disable, activate or abort any other instructional measure. Figure 3-5 shows an example of the control structure. The example shows that when assignment “D 2” is closed, the assignment itself will be enabled again, meaning that the learner can open it again, and the monitoring tool will be closed.

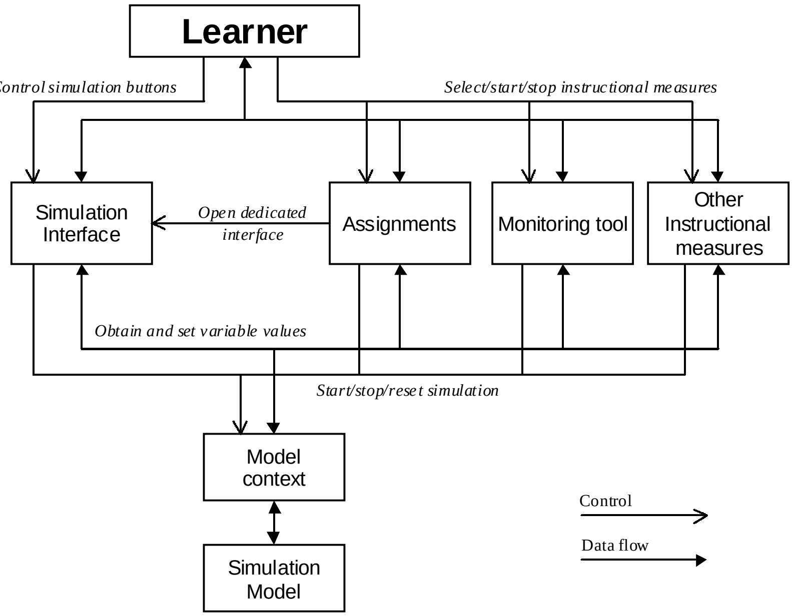


Figure 3-4. The basic structure of control and data flow in a SIMQUEST learning environment.

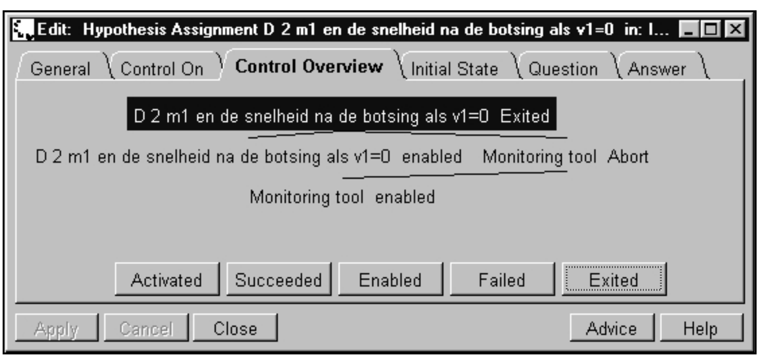


Figure 3-5. Control overview of an assignment.

These possibilities allow an author to design an interactive learner environment that can set out different paths that learners can follow while working with the learning environment. Authors can design environments that range from a very structured system controlled environment to less structured, and more learner controlled environments, and everything in between. In a very structured, system controlled environment, the author would present a simulation to the learners with an assignment, and specify the next assignment based on the state of the assignment (succeeded, failed or exited) when the assignment is closed, thus, controlling the trajectory for the learner. A more free, learner controlled environment would present simulations, assignments, explanations, monitoring tools to the learner, and hand the regulation over to the learner.

3.2 Redefining the research question

The framework that was outlined in the previous paragraph allows authors to design and develop simulation-based learning environments and to provide support for learners that are working with these learning environments. They can design assignments that support the learner with orientation, by pointing out relations between variables that might be worth investigating. Presenting possible hypotheses about this relation as answer alternatives in these assignments can support hypothesis generation. Adding a list of experiments that can be used to test the hypotheses can support hypothesis testing. Connecting feedback to answers in assignments gives the author the possibility to include feedback that reflects the author's idea about the learners' reasons for choosing a particular answer in an assignment.

However, it does not provide a way of assessing the learners' experiments, while they are working on an assignment, in combination with the answer that is given. Such an assessment could be a valuable source of information that could be used to provide support for learners related to hypothesis testing and drawing conclusions. It should leave the learners freedom to design their own experiments for testing the hypotheses in the assignment, and try not to disrupt this process more than is really needed. Therefore, it should not try to collect a lot of extra information from the learners to get a better picture of the learners plans, or change the process in a stepwise procedure that learners have to follow.

This leads to a more focussed version of the research question:

> Can we develop a tool that supports learners in the process of testing a hypothesis and drawing conclusions in a simulation-based learning environment, based on their individual interaction with the learning environment?

The next sections will describe two versions of a tool that was developed for this purpose.

3.3 Design of the first version of the tool: Using induction and deduction to support discovery learning[1]

This section will describe the design of a tool that generates a learner model based on a learners' experimenting behavior in a discovery learning environment, and uses this learner model to provide feedback to the learner. The tool operates as a self standing module within applications created with SIMQUEST, and strengthens the relation between one type of assignment (investigation assignments see Figure 3-1) and the learner's experimentation while working on the assignment. Figure 3-6 illustrates how the tool can be positioned within the existing control structure. The principles of learner modeling and generating feedback will be outlined first, followed by a scenario of how it could operate in a real situation. The scenario will show that restricting the analysis to assessing the correctness of the learner's hypothesis, can yield information that can be utilized to give advice about experimentation without relying on domain knowledge.

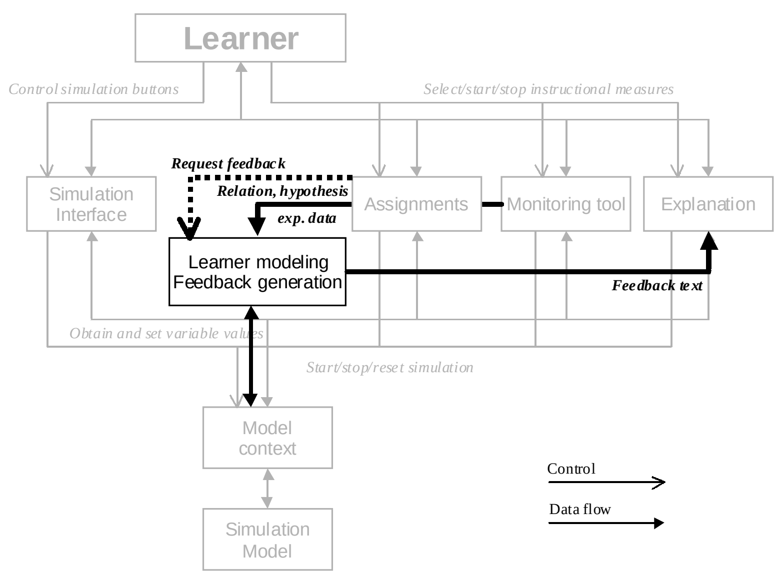


Figure 3-6. The structure of control and information exchange in a SIMQUEST learning environment with learning modelling added.

[1] The Section 3.3 is based on Veermans, K., & Joolingen, W.R. van (1998). Using induction to generate feedback in simulation-based discovery learning environments. In B.P. Goetl, H. M., Halff, C.L. Redfield, & V.J. Shute (Eds.), *Intelligent Tutoring Systems, 4th International Conference, San Antonio, TX USA* (pp. 196-205). Berlin: Springer-Verlag.

3.3.1 Generating a learner model

The mechanism that is used in the tool to generate a learner model, that serves as a base for generating individualized advice on discovery learning, is based on the principles of induction and deduction. An induction process tries to infer a hypothesis from a given set of data, while a deduction process tries to predict experimental outcomes from a given hypothesis. In the mechanism we invert both processes: instead of reasoning forward from given data to a candidate hypothesis, or from a given hypothesis to predicted data, we reason back from a candidate hypothesis to supportive data or from experimental outcomes to a given hypothesis. In this way can be assessed the steps the learner takes and used as a basis for generating tailored advice.

A set of experiments performed by the learner, described as a set of values assigned to input and output variables, is taken as a starting point. A second input for the learner model is the hypothesis that the learner is investigating. The source is an assignment in the environment. In the assignment (see Figure 3-1, p. 29) the learner has to investigate the relation between the variables spring constant and frequency. The hypothesis the learner is working on is: “If the spring constant increases then the frequency = 0”.

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

- First, a set of *informative experiments about the relation* is filtered from the complete set of performed experiments. An experiment (or pair of experiments) is considered to be informative when the input variables that have been manipulated take part in the relation. If this set is empty, the process stops here.
- Then, a set of *informative experiments about the hypothesis* is filtered. This process uses the form of the hypothesis to divide the set of hypotheses resulting from the previous filter into sets which each can generate a prediction using the hypothesis. For instance, for a hypothesis with the form: “If a doubles, b is divided by two”, experiments will be selected where a is doubled, quadrupled etc. where all other input variables are kept constant.
- For each of the informative sets for the hypothesis, *predictions* are generated for the output variables. This can be done based on the hypothesis. For instance, if the hypothesis is a quantitative relation, such as, $y = 2x$. Then the output variable y can be computed directly from the input variable x. If the hypothesis is qualitative, such as: “When x increases, y increases”, it can be inferred from the (x,y) pairs: (1,5), (2,?) that the value on the question mark must be greater than 5. The more information available, the more accurate the prediction can be.
- The predictions generated are compared to the values actually found in the experiments. On mismatches, advice can be generated.

The analysis of experiments as described here yields two kinds of information. Firstly, the presence or absence of informative sets of experiments for a certain hypothesis contains information about what a learner *could* have inferred about a certain relation in the domain. Secondly, the information collected can be used as a source to assess the learner's experimenting behavior. In this sense the learner model contains information about the quality of the learner's conclusion process, and the quality of the learner's hypothesis testing process. For both processes, this information can be used to generate advice, directed at improving the efficiency and effectiveness of the discovery processes, without disrupting the self-directed nature of these processes.

3.3.2 Using the learner model to generate advice

In the advice the relation between the experiments and the hypothesis, or the conclusions can be questioned, followed by a suggestion on how to improve this aspect in the future. Being able to distinguish "good" and "poor" experimenters gives an opportunity to present poor experimenters with information that concerns their experimenting behavior. This advice is presented in the form of questioning the relation between the experiments and the hypothesis, and a suggestion on how to conduct experiments that contain information about the hypothesis. When the experiments contained no information about the hypothesis the content of this suggestion depends on the kind of experiments that the learner has done. If the learner has changed more than one variable at a time, it is questioned whether this kind of experiments can serve as a basis to draw conclusions, and at the same time it will be suggested that it might be a better idea to change just one variable at a time to be able to keep track of the changes for the independent variable. For not manipulating the dependent, the value of experiment outcomes will be questioned and it is suggested to focus on the dependent variable.

A discrepancy between learners' beliefs and the evidence that is generated for a specific hypothesis can be translated into advice as well. If learners do not reject a hypothesis when they are confronted with disconfirming evidence, they might either oversee this information or misinterpret the evidence. In either case the attention is drawn to the evidence that should have led to the rejection of the hypothesis. This is done by presenting the predicted outcomes and the observed outcomes for the experiment(s).

It is important to note that no assumptions are made on the correctness of a confirmed hypothesis. No knowledge of correct and incorrect hypotheses is involved and any hypothesis can, in principle, be disconfirmed by future evidence, because strictly speaking, evidence within the induction paradigm can never be conclusive (Klayman & Ha, 1987). Based on the current evidence, the conclusion can only be, that the hypothesis cannot be rejected, and that we can hold on to the hypothesis until further evidence comes along.

3.3.3 A learning scenario

The current section will be used to illustrate how a tool using the ideas that were described in the previous sections could work within a learning environment. The domain used in the scenario is oscillatory motion (Figure 3-7). The simulation consists of a mass, a spring and a damping force. The learner is asked to find out the relations between the mass, spring constant, and the frequency of the oscillation in the situation where there is no friction (damping = 0), and in the situation where friction is present, represented by a damping force. The learner can change the input variables mass, spring constant and damping. There are four output variables: position, velocity, frequency of the oscillation, and critical damping indicating the point beyond which the system will no longer oscillate. For this simulation let us assume that a learner performs the set of experiments presented in Table 3-1, states a hypothesis, and, see what kind of information can be extracted.

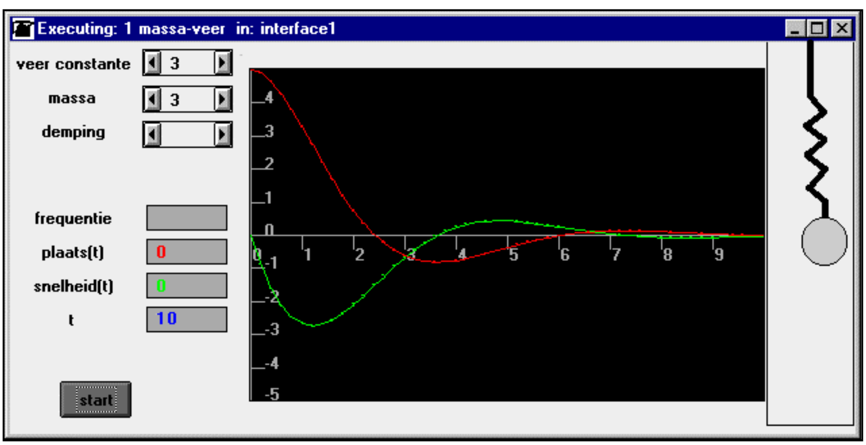


Figure 3-7. The simulation interface of the oscillatory motion simulation.

The hypothesis the learner will start to investigate is:
If the spring constant increases then the frequency = 0
In case a learner only conducted the first three experiments in Table 3-1, the result of the analysis is that the set of experiments that are informative about the relation is empty because there are no two experiments for which the spring constant is changed and the mass is not changed at the same time. At this moment the analysis can stop, knowing that the learner had no support

for the hypothesis because the experiments were not relevant for the relation to what was under investigation.

Table 3-1. Set of experiments used in the first example.

Experiment	Mass	Spring constant	Damping	Critical damping	Frequency
1	1.00	1.00	10.00	2.00	0.00
2	8.00	3.00	10.00	9.80	0.00
3	3.00	5.00	10.00	7.75	0.00
4	1.00	4.00	10.00	4.00	0.00
5	1.00	2.00	10.00	2.83	0.00
6	8.00	8.00	10.00	16.00	0.78
7	8.00	4.00	10.00	11.31	0.33

If the learner would proceed with the next two experiments and again claimed this hypothesis to be correct, the tool would re-evaluate the evidence starting with selecting the experiments that are informative about the relation. This time, the experiments 1, 4 and 5 form a set that is informative about the relation. As the hypothesis is qualitative ("spring constant increases") the same set is informative about the hypothesis. From the hypothesis and experiment 1 predictions can be generated for experiments 4 and 5 and it can be concluded that for this set of experiments, the hypothesis holds. If later on the student conducts experiment 6 and 7 there are two sets of experiments that are informative about both relation and hypothesis. In this case both experiment 1, 4, and 5 and experiment 2, 6 and 7 are informative. The first set supported the hypothesis, the latter, however, predicts a frequency of zero for the experiments 6 and 7, whereas the observed frequencies are not equal to 0. This means that the hypothesis can no longer explain the observations and, thus, has to be rejected. Figure 3-8 shows the kind of feedback that the learner receives in this situation.

The feedback has a format that consists of three components: general information, an overview of informative experiments together with predictions that can be derived from the hypothesis for these experiments, and a conclusion. This distinction provides an opportunity to separate different types of information in the feedback to the learner. The general information gives information concerning testing a hypothesis, and outlines the basic principles behind hypothesis testing. The overview of informative experiments shows the learner, which of the experiments were set up according to these basic principles, and also which predictions can be derived from the hypothesis for these experiments. The conclusion shows how the experiments can be interpreted in relation to the hypothesis to arrive at a (tentative) conclusion about the hypothesis. The system uses templates to generate the components of the feedback. The choice of the specific template depends on the assessment of the learner's experiments in relation to the hypothesis. The content of the templates is related to the present context by using parts of the hypothesis and/or the names of the variables in the hypothesis.

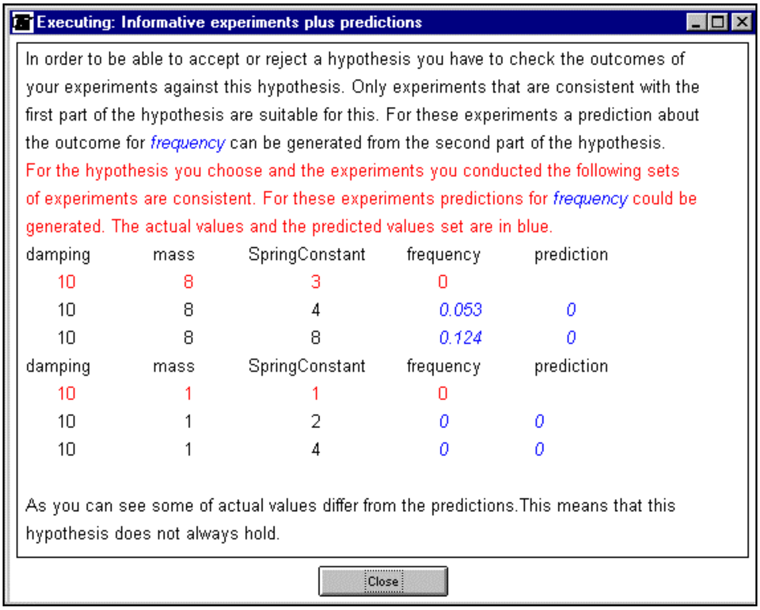


Figure 3-8. Feedback after the experiments 1-7.

A second example is the investigation of the semi-quantitative hypothesis stating: "*If the mass doubles then the frequency will become twice as large*". This example will use the experiments from Table 3-2. After the first four experiments, the learner concludes that the hypothesis is correct. This conclusion is incorrect. The information the analysis offers is as follows: there is a set of experiments for which the mass doubles (1 and 3) but for these two experiments the spring constant is not left constant; this means that the values for the frequencies cannot be interpreted to confirm or disconfirm the hypothesis. If the learner had performed the complete set of experiments in Table 3-2, the conclusion would still have been false but now, because there is evidence (viz., experiments 6 and 3), which yielded a different result than would have been predicted by the hypothesis. In the first situation the learner receives feedback pointing out that the experiments are not suitable for testing this hypothesis and how to conduct experiments to test this hypothesis. In the second situation, the tool confronts the learner with the conflict between the prediction generated by the hypothesis and the actual outcome of the experiment in a similar way as shown in Figure 3-8.

Table 3-2. Set of experiments used in the second example.

Experiment	Mass	Spring constant	Damping	Critical damping	Frequency
1	3.00	3.00	0.00	6.00	1.00
2	3.00	5.00	0.00	7.75	1.29
3	6.00	4.00	0.00	9.80	0.82
4	2.00	3.00	0.00	4.90	1.22
5	2.00	12.00	0.00	9.80	2.45
6	3.00	4.00	0.00	6.93	1.15

3.3.4 Concluding remarks about the design of the first version of the tool

The previous sections outlined a mechanism to model experimentation behavior in a discovery environment. The tool that uses this mechanism is capable of analyzing the experiments that learners perform in the process of testing a hypothesis, and, based on the result of the analysis, it makes inferences about the quality of the learner's hypothesis testing and conclusion process. In a scenario it was demonstrated how the tool generates feedback that supports these processes.

The tool is not a learner-modeling tool that keeps and updates a model of the domain *knowledge* of the learner, but is a learner-modeling tool in the sense that it interprets the behavior of the learner, and that it uses this interpretation to provide individualized feedback to the learner.

In relation to the question and the constraints it can be concluded that:

- *The tool can support testing hypotheses and drawing conclusions.* Through generating feedback on experimentation, and supporting both interpretation of experiments and drawing a conclusion about a hypothesis based on these experiments, the tool can be said to support the hypothesis testing and conclusion processes.
- *It leaves room for the learners to explore.* The assignments present the learners with a list of predefined hypotheses, but within these assignments learners are free to set up their own experiments leaving room for the learners to explore the relation between variables in the simulation.
- *It is able to operate within the context of an authoring environment.* There is no reference in the tool to the actual model of the domain. What the tool needs is a formalized description of the learner's hypothesis, in combination with experiments that are meant to test this hypothesis. The hypothesis can be retrieved from the context offered by the learning environment (as was done here), or from a hypothesis generation tool like a hypothesis scratchpad (van Joolingen & de Jong, 1993). This means that it can be used in any domain for which a more or less formal description of the domain is possible. This applies to most domains for which simulation models can be constructed.

3.3.5 Revisions to the tool based on an explorative study

The tool as described in the previous section was used in a pilot study with 5 university students from the faculty of Educational Technology. It was used with a simulation on harmonic and damped oscillations.

At the beginning of the session, the students could only get access to the simulation. Assignments were hidden from the students during these first fifteen minutes. This was done to get an idea about the students' capabilities of exploring an environment without support. The results of this phase were only recorded for two of the five subjects, due to technical problems. Analysis of the experimentation behavior of the two subjects for whom the interaction was recorded revealed that one was working in a relatively structured way changing more than one variable only once in nine experiments. The other worked in a much more unstructured way. This student was changing more than one variable on five occasions in fourteen experiments. This student even explicitly stated having no idea of what she was doing, or what she was supposed to be doing. After the exploratory phase neither of the two students was able to answer assignment questions about the qualitative relations in the domain. They both had to switch to the simulation and start experimenting again before they could answer these questions. Only one of the other three students derived a qualitative understanding of some of the relations in the domain during the exploration phase.

With regard to the dynamic feedback a number of issues came up during the sessions or the interview afterwards. One was that the students really liked the fact that the feedback was based on their own experiments. Even though they liked it they also made some remarks about it. In this exploratory study only the informative experiments were used in the feedback, the non-informative experiments were left out. Some of the students asked why only these experiments were shown, failing to notice why the experiments that were left out were not informative in relation to the hypothesis that they were investigating. Another remark was related to the general style of the feedback, being too long and too abstract. One of the reasons was that the general and specific information in the feedback was not clearly separated, making it more difficult for the students to separate the information about their experimenting, and their conclusions from the general information about experimenting and drawing conclusions. Based on these remarks in the pilot study a few of changes were made to the tool before taking it into the schools.

The feedback was changed in a way that the general information on hypothesis testing was no longer presented to the learners, making the feedback shorter. Instead of the general information, the feedback starts with the hypothesis that the learner is evaluating, and proceeds with a short statement about the learner's experiments in relation to this hypothesis. If the learner's experiments are suitable for testing the hypothesis, the statement simply says that the experiments in the table are the experiments that are relevant for testing the hypothesis. If the learner's experiments are not suitable for testing the hypothesis, it is explained why these experiments are not suitable for testing the hypothesis. Both the learner's experiments and the hypothesis are referenced in this explanation, making the feedback less abstract. The overview of relevant experiments and predictions that can be derived from the hypothesis for these experiments, are presented after this statement in a table format. The feedback ends with a conclusion. If the learner's experiments are suitable for testing the hypothesis, the conclusion states that for a hypothesis to be correct the observations in the experiments should match with the predictions that can be derived from the hypothesis. It then proceeds with a statement that says whether this is case in these experiments and this hypothesis, that is, whether the hypothesis should be rejected, or not. If the learners experiments are not suitable for testing the hypothesis the conclusion starts saying that based on these experiments no conclusion can be drawn about the hypothesis, and proceeds with advice that explains what kind of experiments could be done that would put the hypothesis to the test. The advice uses the hypothesis, to present the learner with specific a specific example of experiments that could test this hypothesis, making the feedback also less abstract in this respect. Figure 3-9 shows an example of the feedback in the new format.

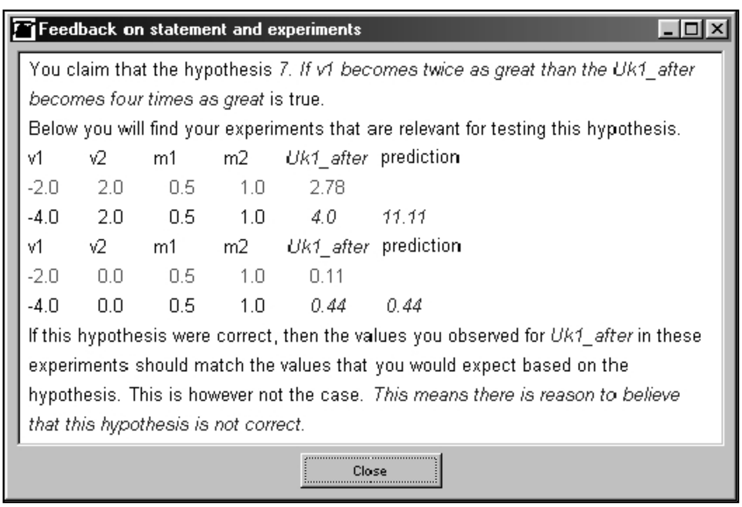


Figure 3-9. Revised format of the feedback.

In Chapter 4, a study is presented in which this version of the tool is evaluated by comparing an environment that contains the tool, with a similar environment without the tool.

3.4 Design of the second version of the tool: Using heuristics to support discovery learning

The main focus of the design of the first tool was to extend the learning environments in a way that it promotes and supports experimenting by learners based on a formal analysis of their experimentation in relation to a hypothesis. The first experiment (see Chapter 4) showed that it was able to fulfill that role, but it also highlighted two important problems with the tool.

The first problem with the tool is that one of the strengths of the tool is also one of the weaknesses. The tool that was described in Section 3.3 does not rely on domain knowledge in the analysis of the learners' experimentation. The strength of this approach is that it is domain independent, and therefore can be used in different environments without the need of knowledge about the domain. The weakness is that it can not use knowledge about the domain to correct learners when this might be needed. This weakness can be illustrated with the example given in Table 3-1 (p. 38). Consider the situation that a student only conducted the first five experiments from this table, and based on these five experiments concluded that the hypothesis "There is no relation between the spring constant and the frequency" is correct. Given the experimental evidence that this learner gathered it could only be concluded that this hypothesis is still valid, even though this hypothesis is not correct in this domain. This might lead to incorrect domain knowledge in the learners, which of course is not among the goals of a learning environment, but on top of that it can also lead to an incorrect self-assessment of the exploration process. This is because the outcome of the exploration process also serves as feedback that is used by learners in assessing the exploration process (Butler & Winne, 1995). In the absence of external feedback, learners have to rely on their own assessment of the outcome of the process. If this assessment is incorrect, the resulting assessment of the exploration might also be incorrect. Providing feedback on the correctness of hypotheses can therefore serve two purposes. It can prevent construction of incorrect domain knowledge, and it can serve as input for self-assessment of the exploration process.

The second problem has to do with the basic approach that was taken in designing the tool. The tool was designed based primarily on formal principles related to induction and deduction, which allowed the tool to analyze the learner's experiments in a formal way resulting in a sound verdict. The problem with this formal approach is that it can not be used to give detailed feedback on experiments unless there are conditions in the hypothesis that allow for a detailed analysis of the experiments. This led to the use of semi-quantitative hypotheses, such as: *"If the velocity becomes twice as large then kinetic energy becomes four times as large"*, because these had conditions that could be used to assess the validity of the experiments related to that hypothesis. This type of hypothesis is slightly artificial, in the sense that in more common language it might be expressed as follows:

"*There is a quadratic relation between velocity and kinetic energy*". The problem with the latter formulation is that it has no condition part that can be used to make a formal assessment of experiments as to whether they comply with the condition or not.

A solution for this second problem is to extend the tool in such a way that it does not only use formal methods to assess the experimentation, but also less formal, i.e. heuristic assessment of the experimentation. Heuristic assessment of the experimentation would allow the tool to provide feedback on experimentation without needing the hypotheses in the assignments as input for the process of evaluating the learners' experiments. This entails that there is no longer a need to formulate the hypotheses in the assignments in a way that allows the tool to distinguish between "good" experimenting and "poor" experimenting. Consequently, the hypotheses in the assignments can now be stated in "normal" language, which makes it easier for the learners not only to investigate, but also to conceptualize them. If the hypothesis in the assignment is no longer used as input for the analysis of the learners' experimentation, it is also no longer needed to connect the feedback to the moment that the learner evaluates a hypothesis as true. This means that feedback on the hypothesis the correctness of the hypothesis can be given in the assignment, thus, solving the first problem. It also means that the feedback on experimentation can be moved to the monitoring tool, the tool in which the experiments are stored; therefore, it is a more logical place to provide feedback on experimentation. Moving the feedback to the monitoring tool requires this tool to be redesigned in a way that it will provide feedback to the learners, and this was the starting point for the design of the second version of the tool.

The heuristics that were included in the tool originate from an inventory by Sanders, Bouwmeester, and van Blanken (2000), who reviewed literature about problem solving, discovery learning, simulation-based learning, and machine discovery, to identify heuristics that can be used in simulation-based discovery learning. From their inventory, a not too large set of heuristics covering the hypothesis testing process was selected. This selection is presented in Table 3-3.

Table 3-3. Selected heuristics related to hypothesis testing.

Keep track	Keep records of what you are doing. (Klahr & Dunbar, 1988; Kulkarni & Simon, 1988; Schauble et al., 1991)
Simple values	Design experiments giving characteristic results. (Klahr, Fay, & Dunbar, 1993) Choose special cases, set any parameter to 1,2,3 (Schoenfeld, 1979)
Votat	If a variable is not relevant for the hypothesis under, test then hold that variable constant, or vary one thing at a time (VOTAT), or If not varying a variable, then pick the same value as used in the previous experiment (Glaser et al., 1992; Klahr & Dunbar, 1988; Schunn & Anderson, 1999; Tsirgi, 1980)
Identify Hypothesis	Generate a small amount of data and examine for a candidate rule or relation. (Glaser et al., 1992)
Equal increments	If choosing a third value for a variable, then choose an equal increment as between first and second values. Or if manipulating a variable, then choose simple, canonical manipulations (Schunn & Anderson, 1999)
Confirm Hypothesis	Generate several additional cases in an attempt to either confirm or disconfirm the hypothesized relation (Glaser et al., 1992)
Extreme values	Try some extreme values to see if there are limits on the proposed relationship(Schunn & Anderson, 1999)
Inductive discovery heuristics	-If you have recorded a set of values for X and a set of values for Y, and the values of X and Y are have a constant ratio of increments, then infer that a linear relation exists between X and Y -If you have recorded a set of values for X and a set of values for Y, and the absolute value of X increases and the absolute value of Y increases, and these values are not linearly related, Then consider the ratio of X and Y -If you have recorded a set of values for X and a set of values for Y, and the absolute value of X increases and the absolute value of Y decreases, and these values are not linearly related, then consider the product of X and Y (Langley, 1981; Qin and Simon, 1990)

3.4.1 Extending the monitoring tool

The idea for the monitoring tool in its original form was to support the students in keeping track of what they are doing while experimenting by providing a storage place for their experiments. The learners could decide to store experiments in the monitoring tool, which would then show the values of the variables in a table format. Extending the tool in a way that it provides feedback on experimenting entails that there should be a point at which this feedback is communicated to the learner. One option is to give feedback after each experiment. This has the disadvantage that it disrupts the learner's experimentation needlessly. Another option is to extend the tool in a way that it combines the feedback with some other action that is initiated by the learner. This is the approach that was taken here, and the learner-initiated action that was combined with the support was drawing a graph.

Drawing a graph was actually one of the heuristics that was frequently mentioned in the literature. It was not yet possible with any of the existing tools in the environment. Extending the tool with the possibility to draw graphs could therefore serve two purposes. Firstly, to support the learners in the process of interpretation of the experiments, since drawing a graph provides a way to visualize the relationship between variables in a graphical way. A linear relationship between two variables is, for instance, easily identified in a graph, since the points in the graph will be located on a straight line. This will usually show more clearly from a graph than from the numbers in a table. Secondly, it would provide a learner initiated action that could be used to present feedback to the learners.

Drawing a graph is not a trivial task and has been the object of instruction in itself (Karasavvidis, 1999). If learners had to draw the graphs themselves this would take a relatively large amount of time. It was therefore decided not to have learners draw graphs themselves. The tool will take care of drawing the graph, and in addition will provide the learner with feedback related to drawing and interpreting graphs, as well as, feedback related to experimenting. All the learner has to do is to select a variable for the x-axis, and a variable for the y-axis. This selection provides the tool with important information that can be used for generating feedback. Through the choice of variables the learner expresses the relation that he/she is interested in. This is similar to the information about the relation that was used in the first tool.

Adding the possibility to draw a graph was not the only extension of the tool. Along with the graph learners can now also ask the tool to fit a function on the experiments. A number of basic functions are provided to the learners. These include qualitative functions (monotonic increase and monotonic decrease), and quantitative functions (constant, linear, quadratic, and reciprocal). More functions could of course be provided, but it was decided to restrict the set of functions to the functions that are actually part of the domain. The reason for restricting the number of functions is that learners might be overwhelmed by the possibilities, some of which they might not even be familiar with. Fitting a function is optional, but when a learner selects this option it provides the tool with valuable extra information for the analysis of the experimentation.

The last extra functionality that was added to the tool was inspired by the inductive discovery heuristics at the bottom of Table 3-3.. It allows learners to construct new variables based the variables that are already present in the monitoring tool. New variables can be constructed using basic simple arithmetic functions add, subtract, divide, and multiply. Whenever the learner creates a new variable, a new column will be added to the monitoring tool, and this column will be updated with the values that the new variable would have in each of the experiments. The learner can then compare these values to the other values in the table to see how this newly constructed variable relates to the variables that were already listed in the monitoring tool. The learner can exploit this functionality in an exploratory way, by constructing variables and then checking for patterns between the newly constructed variable and the output variable, or to validate a hypothesis about a relation, by constructing a new variable that according to the hypothesis, and compare the results to see if the hypothesis is correct. The extended version of the monitoring tool with its new functionality is shown in Figure 3-10.

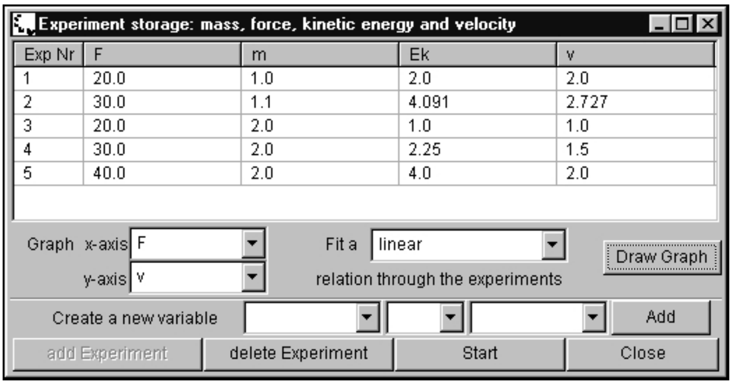


Figure 3-10. The extended monitoring tool.

3.4.2 Providing heuristic support

The previous section described the basic design of second version of the tool. In this section, it will be described in more detail how the tool will provide support for the learner. Basically the support that is provided to the learner upon drawing a graph can be subdivided into three different parts: drawing the data points, calculating and drawing a fit, and providing feedback based on the heuristics from Table 3-3. The first two parts are rather straightforward, and will therefore not be described in detail.

The heuristics from Table 3-3. were divided into general heuristics and specific heuristics (for the division of heuristics see Figure 3-11). The former includes heuristics about experimentation in general, that is, heuristics that are valuable regardless of the context of application. A heuristic like “keep track of your experiments” is, for instance, important to keep in mind in any situation. The latter includes heuristics that are more dependent on the context of application. “Choosing equal increments” between experiments, for instance, depends on the kind of hypothesis that the learner is looking for. It is a valuable heuristic when you are looking for a quantitative relation between variables, but when you are looking for a qualitative relation between variables it is not really necessary to use this heuristic. In this case it might be more useful to look at a range of values, also including some extreme values, than to concentrate on using “equal increments”.

The division between general and specific heuristics is reflected in the feedback that is given to the learners when they draw a graph. The general heuristics will always be used to assess the learner's experiments, and can always generate feedback. The specific heuristics will only be used to assess the learner's experiments *if* the learner fits a function on the experiments. Which of the specific heuristics will be used, depends on the kind of function that the learner wants to fit. For instance, the 'equal increments' heuristic will not be used if the learner fits a qualitative function on the experiments.

The structure that is shown in Figure 3-11 was created for the purpose of analyzing the learner's experiments in the light of the heuristics. The "heuristic tool" class stands at the center in the process of generating feedback, and is created when the monitoring tool is opened. It keeps a reference to the feedback tracker, (the place where the actual feedback for the learner is stored), the heuristics, (a list of heuristics that are compared to the learner's behavior, and that might want to give feedback), and the experiment set (a list of the learner's experiments).

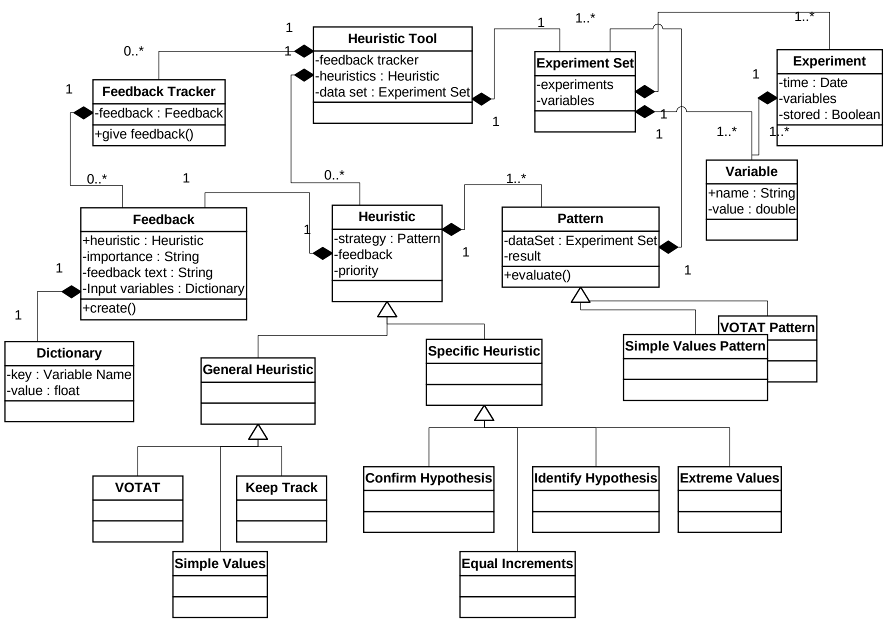


Figure 3-11. Class diagram of feedback generation part of the tool.

The experiment set holds references to all experiments that were done since the heuristic tool was activated. It also holds a reference to all variables that can be used to distinguish between input and output variables whenever this is needed. This is for instance the case when experiments have to be sorted into sets in which the independent variable might change, but other input variables remain unchanged. Associated with each of the experiments in the experiment set, is a boolean that indicates whether the learner stored this experiment in the monitoring tool or not.

The feedback tracker is the place where feedback from the individual heuristics is collected and transformed into the actual feedback text that will be presented to the learner.

The list of heuristics is a list with references to all the heuristics that might generate feedback to the learner. As mentioned before, there are two types of heuristics, general heuristics and specific heuristics. The general heuristics are heuristics that take the “raw” experiment set as input. The specific heuristics use a preprocessed version of the experiments in which the experiments are sorted into sets in which the input variable that will be put on the x-axis of the graph changes, but none of the other input variables. Each of the heuristics has its own associated pattern (see Table 3-4 for an example). This pattern describes behavior that can be compared to the learner’s behavior to identify use of a heuristic.

Keep track	If a learner did not store all experiments in the monitoring tool And at least one of the experiments that were not stored is not a duplicate of one of the stored experiments Then remind the learner of the keep track heuristic
Equal increments	If in a set of experiments in which the value for input variable on the x-axis changes, and the other input variables are kept the same There is no set of experiments in which the increment between the first and the second experiment is equal to the increment between the second and the third experiment Then remind the learner of the equal increment heuristic

Table 3-4. Example patterns for a general and a specific heuristic.

The specific heuristics “identify hypothesis” and “confirm hypothesis” can be said to represent the formal analysis that of the experiments that was used in the first version of the tool. Actually, the first version of the tool used only the “identify hypothesis” heuristic. The analysis that is performed in the first version of the tool only checks whether the hypothesis could be identified based on the experimental evidence that was generated by the learner. It also checks whether this identification was proper. It does, however, not check if the experimental evidence could also confirm the hypothesis. For instance, if the hypothesis was that two variables are linearly related, and only two experiments were done, the analysis of the experiments would be that this hypothesis could be inferred[2]. It then checked if this linear relationship could describe the data (i.e. whether the identification of the hypothesis was properly done). For confirming this hypothesis at least one other experiment is needed. This could show that the hypothesis that was identified is able to account for this additional experiment, but it could also show that the additional experiment is not on the line with the hypothesis that was identified based on the first two experiments.

3.4.3 A learning scenario

A learner working with the simulation can do experiments, and for each of these experiments decide whether to store the experiment in the monitoring tool or not. The heuristic tool keeps track of all these experiments and keeps a tag that that indicates whether an experiment was stored by the learner or not. At a certain moment, the learner can decide to draw a graph. The prerequisites for drawing a graph are that the learner selects a variable for the x-axis and one for the y-axis, and ask the tool to draw a graph for these variables. At this point, the tool checks what kind of variables the learner is plotting, and based on this check the tool decides whether to proceed with drawing the graph, or not to draw a graph and present feedback to the learner. Figure 3-12 shows the sequence diagram for generating feedback.

The latter will happen if a learner tries to draw a graph with two input variables, since this does not really make sense. Input variables are independent, and any relation that might show in a graph will therefore be the result of the changes that were made by the learner, and not of a relation between the variables. The tool will not draw a graph either when a student tries to draw a graph with an input variable on the y-axis, and an output variable on the x-axis. Unlike with the two input variables this could make sense, but it is common practice to plot the variables the other way around. In both cases the learner will receive feedback that explains why no graph was drawn, and what they could change in order to draw a graph that will provide more insight on relations in the domain.

If the learner selects an input variable on the x-axis, and an output variable on the y-axis, or two output variables the tool will generate feedback following the sequence that is shown in Figure 3-12. It starts with asking the general experimenting heuristics to evaluate the experiments that the learner has performed. Each of the heuristics will ask its associated pattern to compare the learner’s experiments with the pattern that was defined for the heuristic. If necessary the heuristic can ask the dataset to filter the experiments (for instance only stored experiments). The feedback is

[2] For two experiments in which the independent variable is changed and all other variables are kept unchanged, one can in principle always hypothesize a linear relation between the two variables.

then asked to generate the feedback text based on the result of this comparison, and the heuristic returns this feedback to the tool. The tool temporarily stores the feedback in the feedback tracker until it will be presented to the learner.

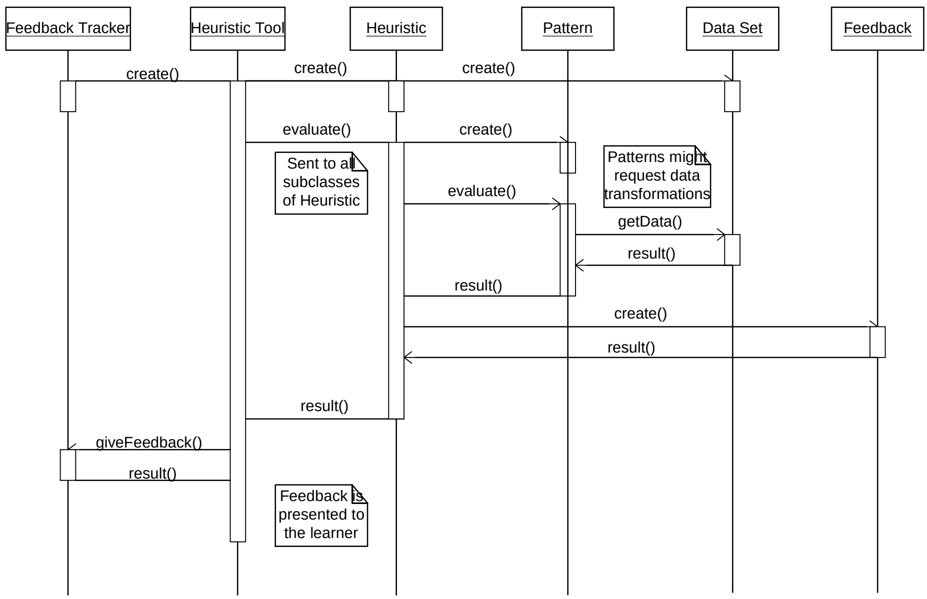


Figure 3-12. Sequence diagram for generating heuristic feedback.

After the general heuristics have assessed the learner’s experiments the next step will be that the tool analyses the experiments using the same principles that were used in the tool that was described in Section 3.3. Based on these principles the tool identifies sets of experiments that are informative for the relation between the input variable and the output variable. This is done by grouping the experiments into sets in which all of the other input variables are kept constant. The result will be one or more sets of experiments. Each of the sets is now sent to the specific experiment heuristics, which will compare them with the heuristic pattern, and, if necessary, generate feedback.

At this point the tool will draw the graph and present the feedback to the learner. Each of the sets of experiments is assigned a different color in the graph to make it easy for the learner to distinguish between them. Learners can view the plots for the different sets, and use them to guide their decisions on hypotheses about the relation.

Together with the plots the tool will now present the feedback that was generated by the general experimenting heuristics. This feedback consists of the name of the heuristic, the outcome of the comparison with the heuristic pattern, an informal text that expresses why it could be useful to set up experiments according to this heuristic, and the explicit statement that it what the learner did is not “wrong”. The tool will also provide information on each of the experiment sets. This information consists of the values of the input variables in this set and the feedback on the specific experiment heuristics.

In the situation in which the learner decides to plot two output variables against each other, it is not possible to divide the experiments formally into separate sets of informative experiments. Both output variables are dependent on one or more input variables, and it is not possible to say what kind of values for the input variables make up a set that can be used to see how the output variables are co-varying given the different values for the input variables. Some input variables might influence both output variables, and some only one of them. Also the way the input variables influence the output variables might also be different. This makes it impossible to assess the experiments and the relation between the outputs formally.

What can and will be done in this situation is that this uncertainty is communicated to the learners, warning them that drawing a graph for two output variables will not always result in a clear picture, and that they should be careful with drawing conclusions based on such a graph. It is also suggested that they could remove some of the experiments to get a set of experiments in which only one of the input variables is varied, as to make sure that this change is the one the causes variation in the output variables. This feedback is combined with the feedback that was generated by the general experiment heuristics.

Learners can also decide to fit a function through the experiments. In general this works the same as in the cases that were just described with the exception that, if possible, for each of the experiment sets a fit would be calculated. If a fit can be calculated, it will be drawn in the graph, and additional feedback will be generated that will be presented to the learner (see for example Figure 3-13). This additional feedback consists of a calculated estimation of the fit and more elaborate feedback from the specific experiment heuristics. The estimation of the fit is expressed with a value on a scale ranging from 0% to 100%, with 0% denoting no fit at all, and 100% a perfect fit. The feedback that is generated by the specific experiment heuristics is more elaborate when the learner fits a function, than it was without fitting. The reason is that the function that the learner tries to fit can be seen as a hypothesis. This hypothesis allows a more detailed specification of the specific experimentation heuristics. The minimum number of experiments that is needed to be able to identify a function through the experiments can be compared with the actual number of experiments in each of the experiment sets. If the actual number is smaller than the required number this is used to generate feedback. The minimum number to confirm a hypothesis is the minimum number that can identify the hypothesis, plus one extra experiment that can be used for confirmation. Learners are also suggested to look at both the graph and the estimation of the fit to guide their decision on the correctness of the fit. At the same time one of the inductive discovery heuristic is used to suggest the learner to create a new variable that could help to establish a firm conclusion on the nature of the relationship.

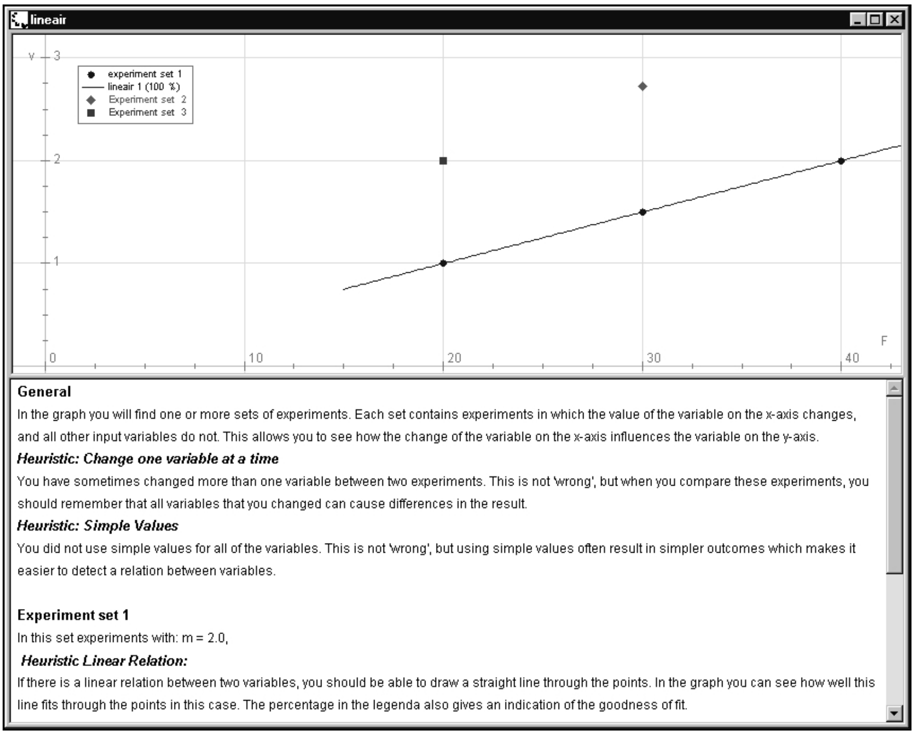


Figure 3-13. Example of a graph with heuristic feedback.

3.4.4 Concluding remarks about the design of the second tool

In the previous sections the design of the second tool for supporting hypothesis testing was described. The tool uses both formal and heuristic methods to analyze the experiments that learners perform in the process of testing a hypothesis, and, based on the result of the analysis, draws conclusions about the quality of the learners' hypothesis testing process. In a scenario it was shown how the tool could generate support for the learners. As with the first version of the tool, it is not a learner-modeling tool, in the sense that keeps and updates a persistent model of the learner's *knowledge*, but is in the sense that it interprets the behavior of the learner, and uses this interpretation to provide feedback to the learner. A difference between the tools is that the first tool only uses formal methods to assess the learner, while the second tool uses both formal and heuristic methods. This makes that the second version of the tool is broader in its scope, since it covers a wider range of behaviors that can be used to provide feedback to the learner. Another difference is that the first version of the tool had a strong connection with assignments, and the hypotheses that were presented to the learner in these assignments. The second version of the tool no longer has this strong connection with the assignments. It can be used with assignments just like the first version of the tool, but now also “stand alone”.

In relation to the research question and the constraints it can be concluded that:

- *The tool can support testing hypotheses and drawing conclusions.* By sorting the experiments into sets that are informative for the relation in the graph, and drawing these sets as separate plots, generating feedback on experimentation, and generating feedback that can help the learner in the design and analysis of the experiments, the tool can be said to support hypothesis testing. Drawing separate plots, and presenting an estimated fit for a fitted function supports drawing conclusions.
- *It leaves room for the learners to explore.* The tool leaves learners free to set up their own experiments, to draw graphs, and to fit relations through these graphs, thus leaving room for the learners to explore the relation between variables in the simulation.
- *It is able to operate within the context of an authoring environment.* The tool is designed as a self-standing tool, and can be used as such. It does not have dependencies other than the dependency on the model context, the central manager of the simulations within the authoring system.

Chapter 5 presents a study in which this version of the tool is used within a broader setting. The study compares two learning environments that differ in their use of heuristics to support the learner.

4

Study 1: The effect of intelligent feedback on discovery learning[1]

Due to the difference in locus of control and learning paradigm, ITSs and discovery learning seem to be incompatible. However, as was argued in Chapter 2, problems with the traditional ITS approach lead to systems that shift the control more to the learner (Holt, Dubs, Jones, & Greer, 1994; Shute & Psotka, 1996) while at the same time, problems with discovery learning (de Jong & van Joolingen, 1998) created a need for advanced support for discovery learning processes. The previous chapter described two versions of a tool intended to provide this kind of advanced support to learners. Section 3.3 described the design of the first version of the tool. The tool uses ideas from ITS to support the hypothesis testing and conclusion processes. This chapter will describe a study that was designed to test this tool in a learning situation.

4.1 Simulation-based discovery learning environments

In simulation-based discovery learning environments, learners change values of input variables and observe values of output variables, inducing characteristics of the underlying model. An example of such an environment is given in Figure 4-1. This is the interface of a simulation about the physics topic of collisions as it is used in the present study. Learners can do experiments by manipulating the variables mass and velocity for two balls and then running the simulation by pressing the start button. The variables can be manipulated by clicking on the arrows next to the values or by just changing the values. Visual information on the screen includes graphs

[1] This text is based on : VEERMANS, K., JONG, T. DE, & JOOLINGEN, W.R. VAN (2000), Promoting self directed learning in simulation based discovery learning environments through intelligent support. *Interactive Learning Environments 8,* 229-255.

(displaying position, velocity, and kinetic energy of the balls), numerical output, and on the left side, an animation of the movement of the balls.

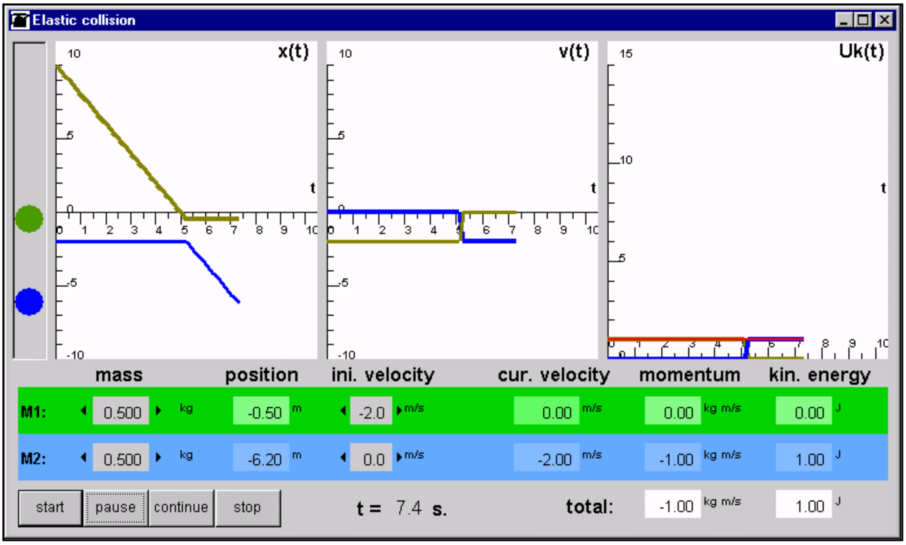


Figure 4-1. Elastic collision simulation window.

Support for discovery learning aims at providing context and tools for performing learning processes essential for discovery learning. The paradigm is that the environment provides the learner with cognitive tools (Lajoie, 1993; van Joolingen, 1999). These tools help the learner perform learning processes by offering information, externalizing learning processes, or structuring the task. Examples of learning environments that follow this paradigm can be found in the environments based on the SIMQUEST authoring system for simulation-based discovery learning that was described in Section 3.1. Simulations are the core of the environments and learners are supported in various ways to help them in the learning process. SIMQUEST-based learning environments can structure the task using *model progression* (White & Frederiksen, 1990); provide support for learners by using *assignments*, small tasks that provide the learner with sub-goals that are within reach; providing the learner with just-in-time information in the form of *explanations* and allowing learners to organize the experiments they have done with the simulation in a *monitoring tool*.

Typically, a learner will utilize the support in a SIMQUEST learning environment by opening an assignment and trying to reach the goal presented in the assignment. Figure 4-2 is an example of an assignment that goes together with the simulation presented in Figure 4-1. This particular type of assignment, an *investigation assignment*, asks the learner to investigate a certain relation in the simulation. The learner would conduct experiments with the simulation and the results would be displayed in the monitoring tool window (Figure 4-3). Next, the learner would analyze the results and choose one of the hypotheses in the bottom part of the assignment window. The learner would then receive feedback on this choice in the form of a pre-defined explanation containing information on the correctness of the hypothesis and optional extra information to support the learning process.

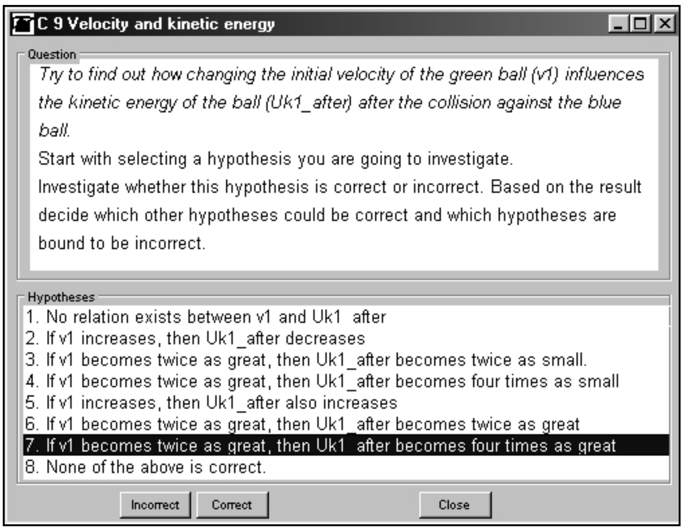
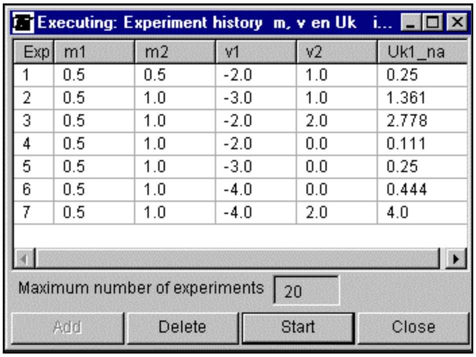


Figure 4-2. An investigation assignment.

Using only standard techniques, the feedback that can be given on the learner's actions is limited. Given a list of hypotheses, and basic true/false feedback, it will be up to the learner to set up experiments to test the hypotheses. This is not a trivial task for all learners, and a learner that has problems with this task might get entrapped in a trial and error-game with the assignments and answers. The first version of the tool (Section 3.3), which intents to support the learners in the hypothesis testing and conclusion processes, was designed to prevent this from happening. Instead of the predefined feedback, learners receive feedback from the tool that is based on the hypothesis and their own experiments.



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

Figure 4-3. The monitoring tool, with experiments.

4.2 Intelligent support for discovery learning

This section will shortly summarize the principles of the first version of the tool (for a more elaborate description see Section 3.3).

The starting point for the tool is a set of experiments performed by the learner and the hypothesis that the learner is investigating. The hypothesis is obtained from an investigation assignment but could in principle also be obtained from another source like a hypothesis scratchpad (van Joolingen & de Jong, 1993). In the assignment in Figure 4-2, the learner has to investigate the relation between the variables *v1* (initial velocity of ball 1) and *Uk1_after* (kinetic energy of ball 1 after the collision). The hypothesis the learner is working on is "*If v1 becomes twice as great than the Uk1_after becomes four times as great*". In order to assess whether a hypothesis can predict a given set of data the stepwise process that was described in Section 3.3.1 is applied to the experiments. Based on the outcome of this stepwise process, it can be assessed whether the experiments suffice to draw conclusions about the hypothesis, and, if necessary the tool can give advice on the kind of experiments that can test hypothesis. The tool can also infer whether a conclusion drawn by the learner is correct or not, and draw a learner's attention to conflicting data in case the learner draws an incorrect conclusion.

In the example shown in Figure 4-4, the learner evaluated the hypothesis "*If v1 becomes twice as great than the Uk1_after becomes four times as great*" to be true. The analysis of the experiments resulted in two sets of two experiments that contain information that can be used for evaluation of the hypothesis. The other experiments that the learner did were not informative because variable *v1* was not changed, or did not fit the condition part of the hypothesis. The first experiment of each of the two sets is used to generate a prediction for the value of variable *Uk1_after* and this value is then compared to the value that was actually observed in the simulation. It is then explained that a correct hypothesis should generate correct predictions in all cases. In this example, the predictions did not match the actual values. Thus, the learner is informed about this discrepancy and is advised to reject the hypothesis.

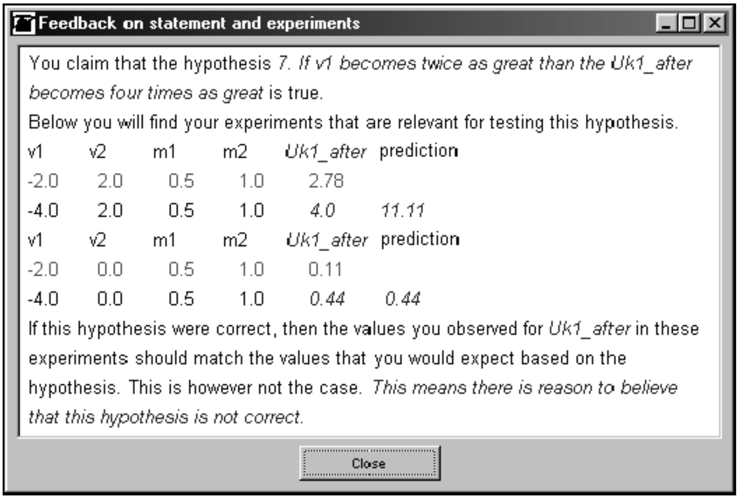


Figure 4-4. Feedback explanation with analysis of the experiments in Figure 4-3 and the hypothesis from Figure 4-2.

4.3 Design of the study

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

4.3.1 Conditions

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

1 A Control Condition: students interacted with a SIMQUEST simulation environment on collisions. This simulation included assignments stimulating students to detect the principles behind moving and colliding

particles by means of manipulating input variables and interpreting the outcomes of their experiments. The assignments covered relations in the domain and contained hypotheses about these relations. The learners were asked to select the correct hypothesis or hypotheses from this list. If they did so, they received feedback that was pre-defined and contained a statement about the truth-value of the hypothesis and additional information. For instance, if the hypothesis was correct, but there was another, more precise, correct hypothesis the additional information prompted the learner to look for this other hypothesis as well.

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

4.3.2 Participants

The participants were 46 Dutch students from two schools. The students took part in the study on their fourth year of pre-scientific education (15-16 year-olds). Students attended physics classes and had some computer experience. The students were transferred from their schools to the university to participate in the experiment. One school participated with 23 students from two classes in one experimental session. The other school participated with 23 students from one class. The students of these classes were distributed equally over the two experimental conditions. The students received no compensation or credit for their participation.

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

4.3.3 The learning environment

The discovery learning environment used in this study was called *Collision* and covered the physics domain of central collisions between two balls. *Collision* was developed in the SERVIVE project (van Joolingen et al., 1997; de Jong et al., 1998; de Jong, Martin, Zamarro, Esquembre, Swaak, & van Joolingen, 1999). The *Collision* learning environment was designed for learners in the fourth year of secondary school. It included four levels of complexity: non-accelerated movement, collisions against a wall, completely elastic collisions between two balls in one dimension and, completely inelastic collisions between two balls in one dimension. Figure 4-1 displayed a simulation interface for the third level. Apart from the four levels, the environment contained support in the form of assignments, a monitoring tool that registered the experiments with the simulation, background information about the simulations, and feedback explanations.

Assignments

A total of 41 investigation assignments guided the students in exploring the domain. Learners in both conditions were free to choose any assignment. The assignments offered four to eight hypotheses to the students. The students were advised to select one of the hypotheses from the list, to experiment with the simulation, to evaluate the evidence for the hypothesis and, if they felt this was necessary, to investigate other hypotheses as well. The content of the assignments was identical for the two conditions. In both conditions, assignments guided the students in investigating the relation between (a) mass, velocity, and momentum; (b) mass, velocity, and kinetic energy; (c) mass, velocity, and resulting velocities after a collision. In addition, conservation of momentum was treated in assignments on elastic and inelastic collisions and contrasted with the loss of kinetic energy in inelastic collisions.

Monitoring tool

Whenever a student opened an assignment concerning a relation between variables, a monitoring tool (Figure 4-3) was activated which automatically registered the values of the important variables in that relation. This monitoring tool served as a kind of external memory for the students. After each experiment the values for the variables were listed in the monitoring tool. The students had the opportunity to re-run any of the experiments that were in this list and could re-arrange the experiments in the list to get a better overview of the experiments. The general idea behind an instructional measure like the monitoring tool is to allow the students to focus on discovering the relations in the domain. Without the presence of the monitoring tool, students would have to remember the results of their experiment and think of an appropriate next experiment at the same time and then interpret the results afterwards.

4.3.4 Tests

Three different tests, developed by Swaak (1998), were administered to assess the students' knowledge: the *definitional knowledge test*, the *what-if knowledge test*, and the *what-if-why test*.

Definitional knowledge test

The definitional knowledge test consisted of three-answer items and aimed to measure conceptual knowledge of a declarative quality like definitions and equations. An example of the test is presented in Figure 4-5. The same definitional test was given both as a pre- and as a post-test. Whenever learners selected an answer, the item disappeared from the screen and the next item popped-up. Learners were allowed to return to previously answered items. The definitional knowledge test consisted of 20 items.

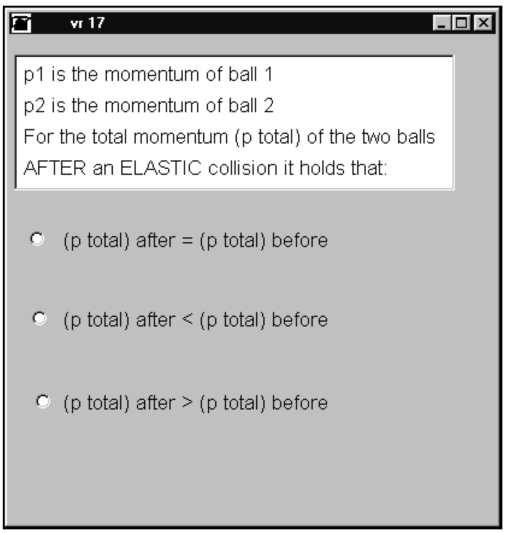


Figure 4-5. Example question from the definitional knowledge test.

Intuitive knowledge test (what-if test)

To measure intuitive knowledge about the relationships between the variables of the domain, a test called the what-if test was created (Swaak & de Jong, 1996; Swaak & de Jong, 2001). In the what-if test (see Figure 4-6), each test item contained three parts: conditions, actions, and predictions. The conditions and predictions were possible states of the system. The conditions were displayed in graphs. The action was presented in text. The predicted states were, like the conditions, presented in graphs. In the instructions of the what-if test the learners were asked to decide which state would follow from a given condition as a result of the action. The items of the task were kept as uncomplicated as the domain permitted. The items had a three-answer format. In order to prevent memorization effects, two parallel versions of the intuitive knowledge test were developed (however, 9 of the 24 items were identical in both versions because no parallel item could be constructed). Whenever learners selected an answer, the item disappeared from the screen and the next item popped-up. Learners could not go back to previously answered items.

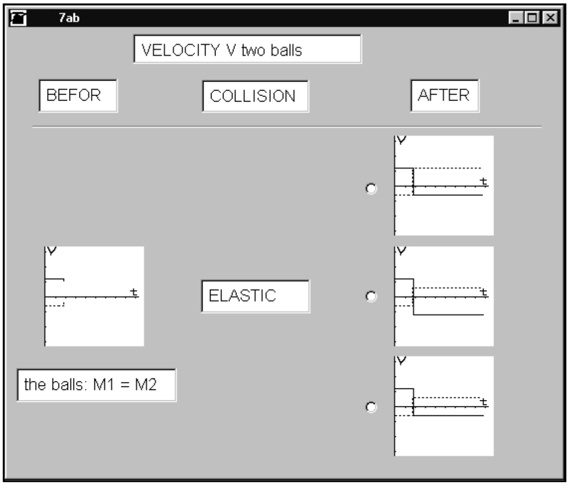


Figure 4-6. Example question from the intuitive knowledge test.

What-if-why test

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

4.3.5 Process measures

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

4.4 Procedure

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

1 *Introduction and pre-tests (40 minutes).* Students were welcomed and given an overview of the activities in the session. After the introduction, the definitional knowledge and what-if pre-test were administered electronically.
2 *Introduction to the learning environment (10 minutes).* Upon completion of the pre-tests, students read an introduction on the Collision environment. This was followed by a demonstration in which the experiment leader showed the function of the various elements of the learning environment and explained how they could be operated. It was explained to the students that both their performances on the tests and their interaction with the learning environment would be recorded.
3 *Interaction with Collision (set at 1 hour and 30 minutes*). After the introduction, students worked with the Collision environment on their own. The experiment leader was present and could give assistance on questions concerning operating the environment, but not on questions concerning subject matter. Students were encouraged to use the full one and a half-hour available for the interaction. If they wanted finish earlier they were asked to explore more of the environment, but were not forced to do so.
4 *Post-tests (45 minutes).* After the interaction with the learning environment the post-tests were administered. The definitional knowledge test was administered first, then the what-if test, and finally the what-if-why test followed. The what-if-why (with the prediction and the explanation part) test was administered using paper and pencil, the other two tests were presented electronically.

4.5 Predictions

The learning environments in the two conditions were identical except for the possibility to reject a hypothesis in the experimental condition and the feedback that the student received upon evaluating a hypothesis. Both environments required the students to investigate relations between the variables in the domain through experimenting with the simulation.

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

Our second prediction was that both conditions would gain equally in definitional knowledge as measured by the definitional knowledge test, but that the experimental condition would also gain more knowledge as measured by the what-if and what-if-why tests. The definitional knowledge test assesses the formal principles of the domain, which are not explicitly dealt with in either of the two environments. It was anticipated that students in the experimental condition perform better on the what-if test because, as a result of the more intense interaction, the students in this condition would have constructed more intuitive knowledge about the domain.

The third prediction was that the students in the experimental condition would perform better on the why part of the what-if-why because they would reflect more on their interaction with the environment, therefore being able to explain more adequately in their own words why a certain situation had occurred.

4.6 Results

The result section presents analyses of the learning outcomes and processes, and consists of five parts. The first part of this section presents the overall results of the knowledge tests and a comparison between the two conditions. In the second part, as an indication for the integration of knowledge, correlations between post-tests will be shown. The third part will investigate the relationship between the pre-tests and the post-tests to see what role the definitional and intuitive knowledge play in the two conditions. The fourth part presents data on the interaction of learners with the learning environment, and in the fifth and final part the relation between the process and the knowledge measures is presented.

4.6.1 The knowledge measures

The *definitional knowledge* test was administered before and after the session. Reliability analysis (N = 43; n = 20 items) resulted in a reliability of .46 (Cronbach's α) for the pre-test and .38 for the post-test. The reliability of the pre-test was moderate, but the reliability of the post-test was relatively low. The *what-if* test was administered in two parallel versions as pre- and post-test. Reliability analysis of the pre-test (N = 44; n =2 4 items) resulted in a reliability of .70 (Cronbach's α) and .56 for the post-test. The reliability of the pre-test was good and the reliability of the post-test moderate. The *definitional knowledge* test and the *what-if* test are assumed to measure different types of knowledge in learners. The low correlation between the pre-tests (.29) supported this assumption. The *what-if-why* test was scored on correctness of the prediction given and on the correctness of the explanation given for the prediction. The score for the explanation could be either 0 (incorrect), 0.5 (partly correct), or 1 (correct) and was rated by two independent domain experts. The inter-rater reliability yielded a Kappa of .65, which can be considered substantial. The reliability of the *what-if-why correct* results (Cronbach's α) was somewhat low (.38) and the reliability of the *what-if-why explanation* results was good (.69).

The results of the knowledge tests are given in Table 4-1. As can be seen, learners gained on both the definitional and what-if tests in both conditions. Paired-samples T-Tests showed a significant within-subject effect for the definitional test across conditions ($t = 5.96$, $df = 42$, $p = .001$) and in each of the conditions (for the experimental condition: $t = 3.35$, $df = 19$, $p = .003$; for the control condition: $t = 5.14$, $df = 22$, $p = .000$). The overall effect size was $d = .88$. Similarly, there was a significant within-subject effect for the what-if test (overall: $t = 7.81$, $df = 42$, $p = .000$; experimental: $t = 5.11$, $df = 20$, $p = .000$; control: $t = 6.04$, $df = 21$, $p = .000$). Overall effect size again was $d = .88$.

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

Table 4-1. Mean scores and standard deviations for the different knowledge tests in the different conditions. Standard deviations are given within parentheses.

	Condition					
	Experimental		Control		Total	
	Pre	Post	Pre	Post	Pre	Post
Definitional (max=20)	11.6 *(2.6)*	13.7 *(2.4)*	10.9 *(2.9)*	13.4 *(2.3)*	11.3 *(2.8)*	13.5 *(2.4)*
What-If (max=24)	14.7 *(4.0)*	18.3 *(3.2)*	14.1 *(3.6)*	17.6 (2.7)	14.4 *(3.8)*	17.9 *(3.0)*
What-If-Why correctness (max=13)		11.2 *(1.4)*		11.1 *(1.6)*		11.1 *(1.5)*
What-If-Why explanation (max=13)		7.0 *(2.1)*		7.2 *(2.5)*		7.1 *(2.3)*

4.6.2 Relations between the different knowledge measures

In order to find out to what extent the different knowledge measures assess similar or different constructs, their correlations were computed. Table 4-2 displays the correlations between the post-tests. In the experimental group, correlations between the definitional post-test and the what-if and what-if-why explanation post-test were not significant. All other correlations were significant. In the control group all correlations were significant.

Table 4-2. Correlations between the scores on the different post-tests, for each of the conditions.

		Definitional post-test	What-if post-test	What-if-why correctness
What-if post-test	Experimental	.38		
	Control	.51*		
What-if-why correctness	Experimental	.44*	.74**	
	Control	.75**	.54**	
What-if-why explanation	Experimental	.13	.60**	.58**
	Control	.61**	.56**	.72**

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

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

Relations between pre- and post-test knowledge measures

Table 4-3 displays the correlations between the pre- and post-tests scores, thus showing to which extent the two types of prior knowledge (definitional and intuitive knowledge) predicted the students' performance on the post-tests. A basic expectation is that a high score on a pre-test leads to a high score on the post-test of the same type. This expectation was met for the what-if tests in both conditions, for the definitional tests in the control condition, but not for the definitional tests in the experimental condition. Remarkable differences can be seen in the pattern of correlations with post-tests for the definitional and what-if pre-test when the experimental and the control condition are compared. In the experimental condition, the definitional pre-test knowledge does not correlate significantly with any of the post-tests. In the control condition, to the contrary, the definitional knowledge pre-test correlates significantly with all of the post-test measures. For the what-if knowledge pre-test, the pattern is reversed. In the experimental condition, the what-if pre-test correlates significantly with the post-tests. In the control condition, the what-if pre-test correlates significantly with the what-if post-test, but not with any of the other post-tests. In summary, it seems that in the experimental condition the *what-if pre-test* is the main predictor of all post-test scores whereas in the control condition the *definitional pre-test* is the main predictor.

Table 4-3. Correlations between pre-test scores and post-test scores on the knowledge tests.

		Definitional Post-test	What-if Post-test	What-if-why correctness	What-If-Why Explanation
Definitional pre-test	Experimental	.30	.39	.33	.15
	Control	.64**	.46*	.46*	.59**
What-If pre-test	Experimental	.61**	.62**	.76**	.37
	Control	.20	.68**	.21	.32

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

Table 4-4. Relations between pre-test score on the definitional knowledge test and the scores on the post-tests.

	Experimental			Control		
	LD (9.33)	HD (14.11)	t (df) p	LD (8.11)	HD (13.67)	t (df) p
Definitional Post test	13.56	14.11	-.48 (16) .635	11.56	14.33	-3.04 (11.8) .010*
What-If Post test	17.88	18.89	-.67 (15) .535	16.44	18.67	-1.64 (16) .121
What-If-Why Correctness	11.11	11.44	-.53 (16) .640	9.89	11.89	-2.84 (16) .012*
What-If-Why Explanation	7.00	7.22	-.32 (16) .819	5.39	8.28	-2.83 (16) .012*

Note. * means $p<0.05$. The p-values are from a T-test comparing the high and low scoring groups on the definitional pre-test. Mean scores on the definitional pre-test for each group are between parentheses.

Table 4-5. Relations between pre-test score on the what-if test and the score on the post-tests.

	Experimental			Control		
	LW (11.27)	HW (18.18)	t (df) p	LW (10.62)	HW (17.30)	t (df) p
Definitional Post test	12.36	15.0	-2.97 (20) .008**	13.75	13.6	.13 (11.4) .900
What-If Post test	16.72	20.1	-2.79 (19) .012*	16.25	19.2	-2.47 (16) .025*
What-If-Why Correctness	10.27	12.09	-3.81 (20) .001**	11.0	11.2	-.30 (16) .768
What-If-Why Explanation	5.95	8.0	-2.61 (20) .017*	6.5	7.45	-.80 (16) .433

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

To further investigate these differences an extra analysis was performed. If the conditions differ with respect to the way their pre-test scores predicted their post-test scores, this should become apparent by comparing groups based on their pre-test scores. Therefore, both the experimental and control groups were divided into two groups, based on their score for the definitional pre-test (using median split, with the median left out). In Table 4-4, these groups are labeled LD and HD respectively (Low / High on Definitional test). These groups were compared for their scores on the four post-tests, using T-tests. The results are displayed in Table 4-4. The results from the correlational analysis reoccur on a more detailed level in this analysis. In the experimental condition, the low definitional knowledge group can not be distinguished from the high definitional knowledge group on the basis of the post-test results, not even on the post-test version of the definitional test itself. In the control group, the difference between the groups becomes smaller but is still present on the definitional post-test, and differences are also found on the other post-tests, although the difference for the what-if post-test is not significant.

A similar analysis was done based on the scores for the what-if pre-test, yielding LW and HW groups. These results can be found in Table 4-5. Again, a pattern emerged in line with the correlational analysis. Now the low and high group in the experimental condition could still be distinguished in all the post-test scores, whereas in the control group only a difference was found on the what-if test post-test.

Process Measures

Actions that students performed while interacting with the learning environments were registered. This provided data on the use of the environments including time distribution over the four levels of complexity, assignments, feedback on hypotheses, experiments, and the variability of experiments. Table 4-6 summarizes data on time spent in general and on the levels. Table 4-7 summarizes the data on the assignments, experimentation, and feedback.

As shown in Table 4-6, the students in the experimental group spent considerably more time than the control group on the first level ($t = 3.73$, $df = 39.79$, $p = .001$). Comments by students, such as, “It does not say what the correct answer is” indicated that the extra time spent on this level was spent there to get acquainted with the feedback. Time spent on the second and third level was almost equal between the two groups. The experimental group spent less time on the last level although, due to the large variance, this difference is not significant.

Table 4-6. Mean and standard deviations of time spent on the complexity levels in minutes.

	Condition			
Time measures (**minutes**)	**Experimental**		**Control**	
Level 1 (non-accelerated mov.)	22:15	(9:28)	12:46	(7:25)
Level 2 (collisions against wall)	20:17	(6:09)	21:40	(7:36)
Level 3 (elastic collisions)	25:30	(5:47)	26:17	(6:23)
Level 4 (inelastic collisions)	18:16	(13:05)	23:33	(9:57)
Total time	86:18	(4:48)	84:16	(4:44)

Independent samples T-tests for the process measures showed a significant difference on the total number (multiple use allowed) of assignments used (t = -2.04, df = 33.44, p = .049), with the control group using more assignments. Furthermore, as seen in Table 4-7, significant differences were found on the overall number of experiments performed (t = 2.34, df = 28.65, p = .027), the number of experiments performed during an assignment (t = 3.32, df = 43, p = .002), the total amount of feedback (t = 4.23, df = 35.22, p = .000), the average amount of feedback in an assignment (t = 5.69, df = 26.87, p = .000), and the number of unique experiments performed (t = 2.60, df = 27.01, p = .0.15), all with the larger numbers in the experimental group. On average, students in the experimental condition also spent more time on an assignment (t = 3.63, df = 43, p = .0.001). Only the number of unique assignments did not show a significant difference between the two conditions.

Table 4-7. Means and standard deviations of process measures within conditions.

	Condition			
Process measures	**Experimental**		**Control**	
Assignments (total number used)	33	(8)	40	(16)
Unique assignments (max. 41)	30	(8)	34	(9)
Experiments	111	(50)	84	(22)
Experiments during assignments	89	(39)	55	(28)
Feedback	94	(38)	54	(24)
Average feedback per assignment	3.2	(1.3)	1.6	(0.5)
Unique Experiments	65	(31)	46	(12)
Average time per assignment (min)	2:32	(1:44)	1:44	(1:43)

Relations between process and knowledge measures

The relations between the learners' activities and the results on the various knowledge tests (i.e., whether the behavior in the learning environment could be related to the scores on the post-tests) were also investigated. A good measure from the learners' activities is the number of different assignments that were used as this provides an indication of the level coverage. The number of unique experiments learners conducted within each level is also an appropriate measure, because the number of unique experiments indicates the amount of evidence gathered that could be utilized for understanding the simulations' underlying principles. In Table 4-8, the correlations between the number of different assignments and the post-test scores are shown.

Table 4-9 displays the correlations between the number of unique experiments and the post-test scores. In both tables, the correlations are computed for the overall number, but also for the number on each of the four levels that were present in the learning environment.

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

Table 4-8. Correlations between the number of assignments used on the different levels and the results on the different post-tests.

Number of assignments	Level 1	Level 2	Level 3	Level 4	Total
Definitional post test					
Experimental	-.21	.21	.34	-.01	.26
Control	.42*	.54**	.47*	.29	.51*
What-if post-test					
Experimental	-.04	.22	.22	.06	.24
Control	.38	.50*	.53**	.33	.52*
What-if-why correctness					
Experimental	.16	.38	.22	-.07	.17
Control	.56**	.65**	.38	-.04	.40
What-if-why explanation					
Experimental	-.02	.32	.21	.04	.27
Control	.40	.42*	.45*	.17	.42*

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

Table 4-9 shows a significant correlation between the number of unique experiments on level 3 and the results on the what-if post-test and the what-if-why explanation test for the experimental group. Significant correlations for the experimental group were also found between the number of unique experiments on level 2 and the what-if, and what-if-why correctness post-test. The total number of unique experiments correlates significantly with the what-if post-test. Only for the definitional knowledge test, there was no significant correlation between the number of experiments and the scores on the post-test. In the control condition, no significant correlations were found between unique experiments and post-tests.

Table 4-9. Correlations between the number of unique experiments performed with the simulations, and the results on the different post-tests.

Unique experiments	Level 1	Level 2	Level 3	Level 4	Total
Definitional post test					
Experimental	-.15	.01	.36	-.11	.08
Control	-.03	-.28	-.03	.16	-.06
What-if post-test					
Experimental	-.11	.52*	.60**	.18	.47*
Control	-.06	-.18	.07	-.06	-.09
What-if-why correctness					
Experimental	.23	.46*	.42	-.08	.29
Control	.29	-.03	-.13	-.39	-.20
What-if-why explanation					
Experimental	.07	.34	.48*	-.05	.28
Control	.14	-.15	.32	-.01	.16

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

4.7 Conclusion

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

The effects of this different behavior did not directly show in the scores on the knowledge tests that were administered after the experimental session. Both conditions gained from pre- to post-test on both the definitional and what-if tests. Contrary to the predictions, no significant differences in favor of the experimental condition were found for the what-if test or the what-if-why test. The average scores show no difference between the conditions, so it seems that there is no influence of the experimental treatment on the definitional and intuitive domain knowledge that learners gained during the interaction with the learning environment.

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

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

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

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

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

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

The above are stereotypes to contrast the learning in the two conditions. Even if these were true, a number of questions remain unanswered after this study. The main question is why there are no differences between the two conditions on the knowledge tests. Three explanations come to mind.

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

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

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

As this is indirect evidence, it would be interesting to try to substantiate this claim in a follow-up study. This might also provide an answer to the issue of the relation between intuitive knowledge and discovery skills. In this study, it was found that in the experimental condition intuitive knowledge was the main predictor for the post-test results. This raises the question of the exact role of the intuitive knowledge in the discovery processes. Is domain specific intuitive knowledge a necessary condition for discovery learning unrelated to the discovery skills, or is there interdependency between the discovery skills and the domain specific intuitive knowledge?

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

5

Study 2: The effect of heuristics on discovery learning

The study described in the previous chapter investigated the effect of support for hypothesis testing and drawing conclusions on learning outcomes, and discovery behavior. One of the findings was that the learning outcomes were related to the students' intuitive domain knowledge on the pre-test for students in the experimental condition, and to the definitional knowledge for students in the control condition. One explanation for this finding could be that the support for the discovery learning processes was not enough for all students. Students in the control condition could compensate for this by adopting a more traditional learning style in which they drew heavily on the assignments and the feedback that was given in these assignments. For students in the experimental condition this was not possible, since they would only get feedback about the hypotheses in relation to their experiments. In the beginning of Section 3.4 it was described that this could have been problematic for students that are not proficient in discovery learning. If the intuitive knowledge test does not only measure domain specific intuitive knowledge, but also more domain general intuitive knowledge that could be used in discovery learning, this could explain the relation between the intuitive knowledge and the learning outcomes in the experimental condition.

In the present study a decision was made to alter the design of the learning environment in a way that learners would not have to revert to a more traditional way of learning, and that the support would also be sufficient for learners that are less proficient at discovery learning. In Section 2.4.2 it was argued that heuristics could play a role in discovery learning by providing a scaffold at the moment that the ideas behind discovery learning are not yet well established in the learner, by triggering good practices, and by extending the scope of an intelligent tool. In Section 3.4 it was already described how heuristics were used to design the second version of the tool. In the next sections it will be described how heuristics were used in the design of the rest of the learning environment that is the subject of the present chapter.

5.1 How to include heuristics in a learning environment

In order for a heuristic to be helpful for a learner it must either be part of his or her repertoire or be incorporated in the learning environment. Even when the learner's repertoire contains a heuristic, the learner may not use it because he or she fails to make the connection to the present context. This makes that it can be useful for all learners to include heuristics in a learning environment.

Salomon (1992) distinguishes two effects for support, effect *with* and effect *of*. Effect *with* means that the support helps learners to accomplish the current task, effect *of* means that the learner will also be able to accomplish the task when the support is no longer present. Depending on the main goal of the learning environment (effect *with* or effect *of*), there are two ways that heuristics can be incorporated in the design of learning environments.

If the main goal is effect *with*, the heuristics can be incorporated implicitly, by building in structures and/or stimuli that trigger behavior in the learner that is in line with the heuristics, without communicating the rationale behind these structures and/or stimuli. In this case, one could say that the tool takes over the responsibility for the heuristic from the learner. By doing this, the effect *with* the tool, will most probably be enhanced since the learner can not make mistakes with regard to that heuristic, but for the effect *of* the tool this might have a detrimental effect. The fact that the responsibility for the heuristic is taken from the learner reduces the chances that the learner will employ a mindful abstraction (Salomon & Perkins, 1989) on the heuristic to make it his/her own. This means that if the effect *of* is also considered an important goal of the learning environment one should look for a way to communicate this to the learner. The constraints and/or stimuli might still be included, but not necessarily.

Suppose a situation in which learners should investigate a relation between variables and conduct some experiments to find out the nature of this relationship. Heuristics that are helpful in designing experiments for this situation are for instance ”simple values”, ”Votat” and “canonical values”. One could implicitly include these heuristics in this situation by presenting the learner with a set of experiments in which the experiments use simple values for the input variables, the increment between experiments is equal, and only one variable is changed from one experiment to the other. By constraining the possible experiments for the learner, it is assured that the learner will exhibit behavior that is in line with these heuristics. This means that for the effect *with* the simulation the desired behavior is obtained and the chances of obtaining the desired learning results are increased. For obtaining an effect *of* presenting this set up is not favorable since the learner is not aware of the choices that led to the experiment set that is presented. The heuristics are only implicitly included in the environment, thus, the learner has to infer the heuristics from the environment before they can be internalized. In the example it is unlikely that the learner will be able to extract this information from the situation and use it in a mindful abstraction. To enhance these chances, the heuristics that led to the choice of the experiments and the rationale behind this choice should be communicated to the learner. Only then the heuristic will become explicit, and available for the learner to be included in a mindful abstraction. In the example this could mean not only presenting the set of experiments, but also the heuristics that led to choosing this set. An alternative would be to present the heuristics only, and leave the selection of the experiments to the learner.

Table 5-1. Heuristics used in the design of the learning environment.

Orientation	Simplify problem	Simplify the problem, or try to solve part of the problem (Polya, 1945; Schoenfeld, 1985)
Hypothesis Generation	Identify Hypothesis	Generate a small amount of data and examine for a candidate rule or relation. (Glaser et al., 1992)
Hypothesis Generation Regulation	Slightly modify Hypothesis	Address slightly modified problems: Weaken or strengthen conditions slightly in reformulating hypotheses (Glaser et al., 1992)
Hypothesis Generation, Regulation	Set expectations	Expectations for a class are used, as expectations for members of the class not previously tested or if a law in one context is found, expect a similar form of law to hold in a new context. (Kulkarni & Simon, 1988; Langley, 1981)
Hypothesis testing, Regulation	Votat	If a variable is not relevant for the hypothesis under, test then hold that variable constant, or vary one thing at a time (VOTAT), or If not varying a variable, then pick the same value as used in the previous experiment (Glaser et al., 1992; Klahr & Dunbar, 1988; Schunn & Anderson, 1999; Tsirgi, 1980)
Hypothesis testing	Simple values	Design experiments giving characteristic results. (Klahr et al., 1993) Choose special cases, set any parameter to 1,2,3 (Schoenfeld, 1979)
Hypothesis testing, Regulation	Equal increments	If choosing a third value for a variable, then choose an equal increment as between first and second values. Or if manipulating a variable, then choose simple, canonical manipulations (Schunn & Anderson, 1999)
Hypothesis testing	Confirm Hypothesis	Generate several additional cases in an attempt to either confirm or disconfirm the hypothesized relation (Glaser et al., 1992)
Hypothesis testing	Extreme values	Try some extreme values to see if there are limits on the proposed relationship (Schunn & Anderson, 1999)
Hypothesis testing	Make a graph	If you have a number of data points with values for variables, then make a graph to get an indication about the nature of the relationship. (Polya, 1945)
Conclusion	Present evidence	If you state a conclusion about a certain hypothesis present evidence to support that conclusion (Schoenfeld, 1985)
Regulation	Keep track	Keep records of what you are doing. (Schauble et al., 1991; Klahr & Dunbar, 1988; Kulkarni & Simon, 1988)

The two alternatives just mentioned illustrate another principle that can be employed to trigger mindful abstractions in learners; the use of scaffolding in relation to the heuristics. Heuristics can be used in an abstract manner, making them more domain independent, or in a specific manner, tying them closely to the domain. This dimension can be utilized in a scaffolding approach in which at the start the heuristics are incorporated in the more specific manner, and then gradually change this into a more abstract manner. Learners can then use the specific heuristics as examples of the abstract heuristics. This too might trigger mindful abstraction in the learners.

Two learning environments were designed for the present study, based on the ideas described. The first learning environment only incorporates the heuristics implicitly in the choice of assignments, the content of the assignments, the feedback on assignments, and the monitoring tool. The second learning environment incorporates the heuristics both implicitly and explicitly. The scaffolding approach was used in both learning environments. Table 5-1 lists the heuristics that were used in the design of the learning environments.

5.2 Design of the study

In the current study students using a learning environment with implicit heuristics are compared with students using an environment in which the heuristics are not only implicitly, but are also explicitly presented.

The learning environments used in the study are built around simulations. Apart from the simulations, both contain the same set of cognitive tools to support the learners. This includes the use of model progression levels, assignments, feedback on assignments, explanations, and the use of a monitoring tool (as described in Section 3.4). The heuristics from Table 5-1 are included in these cognitive tools The model progression levels for instance include the ”simplify the problem” heuristic by dividing the domain into parts that can then be investigated, in turn by the learners. This also implements the ‘keep track’ heuristic at the meta level. Within the model progression levels assignments are offered to the learners and within these assignments a number of heuristics find their place. If possible the ”simplify the problem” heuristic is used at the start of a model progression level to create an assignment that focuses on a simpler problem, which is easier to investigate. Later assignments use the ”slightly modify hypothesis” and ”set expectations” heuristic to address a broader range of situations and see whether the findings from the simpler problem pertain in this broader range of situations. The ”vary one thing at a time” heuristic is included in the assignments by focusing on a relationship between one input and one output variable within an assignment, and by stressing that other variables should be kept the same over series of experiments within such an assignment. The series of experiments in turn are set up in a way that they comply with the

”simple values”, ”equal increments” and occasionally ”extreme values” heuristics. At the same time the ”keep track” heuristic is included at a micro level by stressing the necessity to keep records of their experiments. The “draw graph” and “confirm hypothesis” heuristics are used to support identifying and testing a hypothesis. The “present evidence” is included in feedback on incorrect answers. The feedback in the assignments also contains references to other heuristics if these can be related to the answer. When students open an assignment they also receive a monitoring tool in which they can store their experiments, create new variables according to the inductive discovery heuristics and draw graphs based on the experiments that are stored.

Both learning environments employ a scaffolding approach with specific heuristics at the beginning and abstract heuristics towards the end. A specific heuristic fully specifies a guideline that is derived from that heuristic. An abstract heuristic specifies the guideline on a more abstract level. This distinction is treated the same in both environments. At the beginning all heuristics are specific, and gradually one after the other is transformed into an abstract heuristic. The timing of changing from specific to abstract differs for the heuristics. Experiment design heuristics (see Table 5-1) are changed first other heuristics follow later. This change from specific heuristics to abstract heuristics coincides with the notion of scaffolding students in a learning environment. In the beginning they are equipped with a full scaffold, but gradually the scaffold will become smaller, until in the end they will be on their own. The removal of the scaffold is the same in both environments. For instance, whenever the heuristic changes from specific to abstract in the implicit condition it also changes in the explicit condition. At the same place in the learning environment the two learning environments are always in the same column in Table 5-2.

Table 5-2. An Example of specific and abstract heuristics in implicit and explicit form.

		Specific	Abstract
Explicit Heuristic	Heuristic Name	Equal increments	Equal increments
	Heuristic Rationale	When you choose equal increments when you are changing a variable, it is usually easier to compare the results of the experiments.	
	Guidelines derived from the heuristic	Make F equal to 20 N ..... Make F equal to 30 N ..... Make F equal to 40 N	Choose equal increments between experiments
Implicit Heuristic	Guidelines derived from the heuristic	Make F equal to 20 N ..... Make F equal to 30 N ..... Make F equal to 40 N	Choose equal increments between experiments

The difference between the environments is that in the implicit environment only the guidelines derived from the heuristic are presented to the student, whereas in the explicit condition the heuristic and the rationale behind the heuristic precede these guidelines. In Table 5-2 this difference is illustrated with an example from the ”equal increments” heuristic. Students in the implicit condition only see the last row: the guidelines derived from the heuristic. Students in the explicit condition also see the heuristic name and the rationale behind the heuristic. In the learning environment this difference will show in the content of the assignments, the feedback on assignments and the feedback on the monitoring tool (for the tool description see Section 3.4). The difference in the assignments is that in the implicit condition the student will only receive the guidelines that define a proper way of working to successfully complete the assignment. The students are told what steps have to be taken in order to obtain enough information to make a conclusion in relation to the assignment goal. In the explicit condition these steps are also presented, but they will be accompanied by the heuristic that they were derived from. For instance, instead of just stating make v2 equal to zero, the student will also be confronted with the simplify the problem heuristic, its rationale, and how this heuristic relates to making v2 equal to zero. Feedback on assignments also reflects the difference between the implicit and the explicit condition. In the implicit condition the student only receives feedback that reflects the implicit nature of the heuristics in this condition. In the feedback there are only references to the guidelines that were presented in the assignment. In case of an incorrect answer a student is redirected to the guidelines in the assignment and asked if all guidelines were appropriately dealt with in the process of doing the assignment. In the explicit condition students are confronted not only with the guidelines, but again in combination with the heuristics. The feedback on finding a function or testing a function in the monitoring tool is done in similar fashion. In the implicit condition only references to guidelines for finding a function or testing a function are available, in the explicit condition guidelines accompanied by heuristics.

In the terms of Salomon (1992) one could say that in the implicit heuristics learning environment the focus is on learning effects *with* the tools, whereas in the explicit heuristics learning environment the focus is on effects *with* and effects *of* the tools.

5.2.1 Participants

The participants were 30 Dutch students from two schools. The students took part in the study on a voluntary basis. They were in their fifth year of pre-scientific education (16-17 year-olds). All students attended physics classes and had reasonable computer experience. Seventeen students participated in the experiment in the first school, and thirteen students in the second school. The students within a school class were distributed evenly over the two conditions.

One participant did not complete the knowledge pre-test. This participant was excluded from analyses where this test score was included.

5.2.2 The Learning environment

The learning environments in this study are adapted versions of the Collision environment that was used in the study described in Chapter 4. The learning environment uses model progression, assignments, a monitoring tool, background information, and feedback explanations to support learners.

Model progression levels

There are four model progression levels in the learning environment. Only the first level differed from the version that was used in the first study. In this version, students could change the force exerted on a ball or the mass of the ball and investigate the relation between the force and the resulting velocity, and between the force and the resulting momentum. Students had prior knowledge about force, and this prior knowledge was used to introduce momentum, a key concept in collisions. The focus of the second level was elastic collision of a ball against a fixed wall. The third level also dealt with elastic collisions, but now between two balls. The last level was similar to the third level, except that on this level the collisions were inelastic.

Simulation Interface

Each level contained a simulation of the phenomenon of the level. In the simulation the students could see an animation of the movement of the ball(s). The movement and velocity of the ball(s) were shown in position-time and velocity-time graphs next to the animation. On the last three levels the kinetic energy was also shown in a graph. Under the animation and graphs, values of the properties of the balls were shown in numbers. Input variables were located towards the left side of the simulation window, output variables towards the right side. Students could increase or decrease the values of the input variables by clicking on the arrows next to the value, or by typing in a new value. The students could start the simulation with the buttons on the bottom of the simulation window.

Assignments

Every level contained assignments to support the students. The first level contained seven assignments, the second level five, the third level fourteen, and the fourth level thirteen. The first one or two assignments are an introduction to the level. They are followed by assignments that investigate relationships between input and output variables set up along the lines of the heuristics described earlier. The final assignment on a level wraps things up, or draws the attention to the important issues related to that level. Students were not forced to do assignments in the given order, they were free to choose any assignment at any moment in time. However, the names of the assignments, starting with a number, at least suggested a preferred order.

As an example, the assignments of the first level will be described here shortly. On the first level the first assignment introduces the student to the relation between the animation and a position-time graph. The second assignment introduces a velocity-time graph. The third assignment connects the position-time graph with the velocity of the ball, and it is also the first assignments in which the students are confronted with the heuristics. In the assignment the students have to do a number of experiments in which they change the force ("vary one variable at a time") applied to the ball to 20, 30, and 40 ("simple values" and "equal increments"). They are asked to store these experiments ("keep track") and draw a graph with the tangent of the position-time graph on the x-axis and velocity on the y-axis ("draw graph"). The question that they have to answer in this assignment is about the relationship between these two variables (linear). In the fourth and fifth assignment (see Figure 5-1) the same heuristics are used to investigate the relationship between mass and velocity, and applied force and velocity. Assignment 6 puts the results of these investigations together in an assignment in which the students are asked to come up with two formulas. One formula for velocity in terms of force and mass, and one for push in terms of mass and velocity. Assignment 7, the last assignment of this level, asks the students a question about a bowling ball and a lightweight ball. Students have to tell which of the two is easier to stop, and why this is the case. The aim of the question is to put momentum in a context.

Monitoring Tool

The monitoring tool that was used in this study is the second version of the tool as it was described in Section 3.4. The monitoring tool is presented to the student on all assignments that investigate relationships between input variables and output variables. All user controlled input variables are shown, together with the output variables that are of interest for the current assignment. Apart from its function of keeping track of the experiments, the monitoring tool also has some extra functionality that supports the students in their investigation of the domain. Students can draw graphs based on the experiments, fit functions on the experiments, and construct new variables.

In line with the general set up of the study, two versions of the monitoring tool were used in the study. In the implicit condition the tool will present the graphs and the estimated fit of the functions. They are accompanied by a short text that tells the learner to look at fit estimation and the graph to see if the function fits nicely through the experiments. The feedback from the heuristics is not included in this version. The explicit condition uses full version of the tool including the feedback from the heuristics.

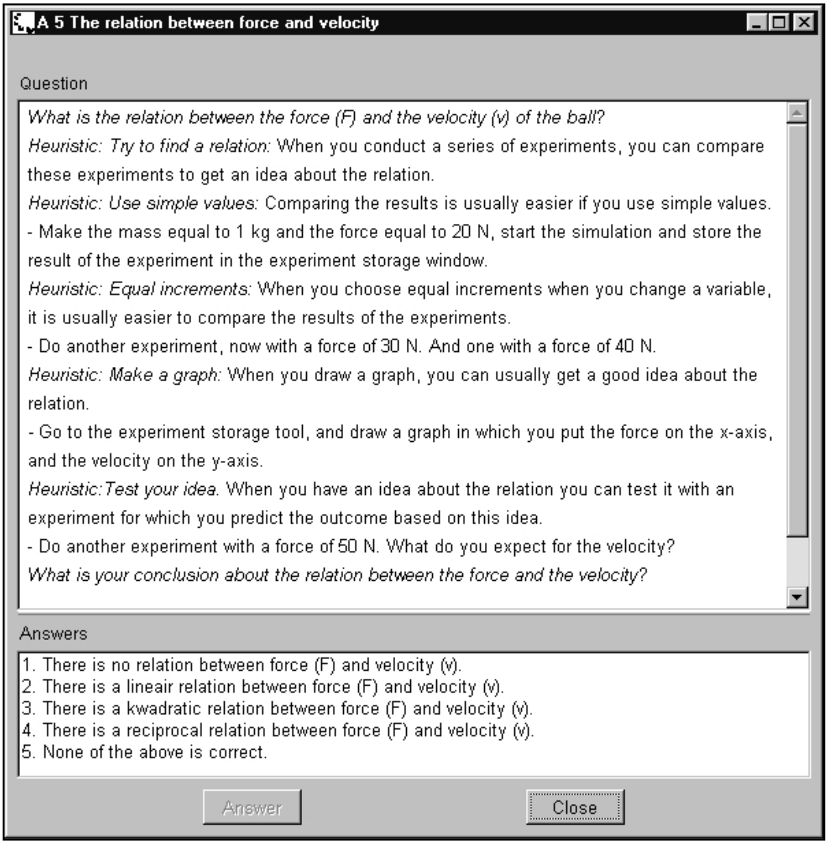


Figure 5-1. Example of an assignment with explicit heuristics.

5.2.3 Tests

Three tests were administered to assess the students' knowledge. Two for assessing the students' domain knowledge (the same tests that were used in the first study described in Chapter 4), and one for assessing scientific reasoning knowledge. The scientific reasoning test consisted of a multiple choice part which will be referred to as the scientific reasoning test, and an experiment design question, which will be referred to as the experiment design test. The scientific reasoning test was added to see whether the differences between the learning environments would result in differences in scientific knowledge, and / or experiment design skills.

Definitional knowledge test

In the definitional knowledge test the students have to answer questions about the formal/static properties of the domain. The test consisted of three-answer items in which students had to choose a correct formula, a general law, or the unit for a certain quantity. The same definitional test was given both as a pre- and as a post-test. Whenever students selected an answer, the item disappeared from the screen and the next item popped-up. Students were allowed to return to previously answered items. The definitional knowledge test consisted of 20 items.

Intuitive knowledge test

The intuitive knowledge test is intended to measure a different type of knowledge than the definitional knowledge. It aims at measuring knowledge about the informal/dynamic properties of the domain. A special test was created to measure this knowledge (Swaak & de Jong, 1996). In this test, each item contained three parts: conditions, actions, and predictions. The conditions and predictions were possible states of the system. The conditions were displayed in graphs. The action was presented in text. The predicted states were, like the conditions, presented in graphs. In the instructions the students were asked to decide which state would follow from a given condition as a result of the action. The items of the task were kept as uncomplicated as the domain permitted. The items had a three-answer format. In order to prevent memorization effects, two parallel versions of the intuitive knowledge test were developed (however 9 of the 24 items were identical in both versions because no parallel item could be constructed). Whenever students selected an answer, the item disappeared from the screen and the next item popped-up. Students could not go back to previously answered items.

Scientific reasoning test

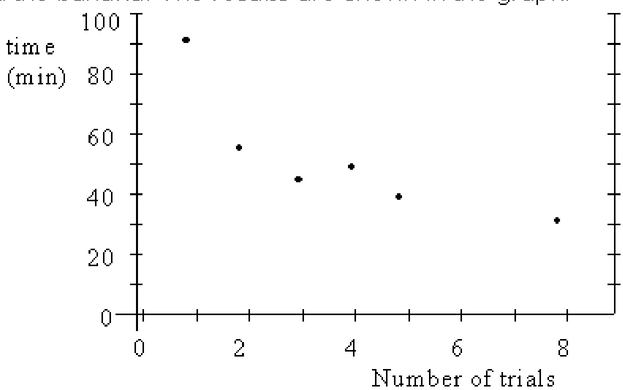
In the scientific reasoning test students received fragments about research done by others. They had to read these fragments and answer one or more multiple-choice questions about each fragment. Each of the questions had 4 answer alternatives. In total there were 15 multiple-choice questions. Figure 5-2 shows an example question from the scientific reasoning test.

Experiment design question

The last question was a question in which students had to design an experimental research setup. The topic was plant growth in a greenhouse, and the students were asked to set up a research design to find out how amount of water, temperature and light influence plant growth in a greenhouse. The idea was that in answering this question students could use the heuristics that they had encountered during their interaction with the learning environment. It was decided to restrict the number of variables to three to make it a non-trivial task that was at the same time not too time consuming.

Question 15

A scientist investigates learning effects in monkeys. He has the monkey's search for a banana in a maze, and measures how long it takes before they find the banana. The results are shown in the graph.



How long will the monkeys need on the 10^{th} trial?

A. 50 minutes
B. 31 minutes
C. 28 minutes
D. 10 minutes

Figure 5-2. Example question from the scientific reasoning test

5.2.4 Process measures

The students' interaction with the learning environment was gathered in a logfile. The content of the logfile was used to extract information about the way students interacted with the learning environment. The amount of time that students worked with the learning environment, model progression levels and assignments were calculated and used for comparison of the students in the two conditions. On the model progression levels scale, time spent on the level, the number of experiments, the number of unique experiments, and the number of graphs drawn, were extracted from the logfiles. On the assignment scale, the time spent on the assignments was calculated, it was checked whether the first answer to assignments was correct, whether students followed the experimentation guidelines, whether students drew a graph, whether students did experiments during the assignment, and if so, the number of unique experiments was counted. These figures were aggregated over the model progression levels to provide a global picture of the students' interaction with the learning environment.

5.3 Procedure

The session lasted approximately three hours, and included the following sequence of events:

1. *Introduction and pre-tests (40 minutes).* Students were welcomed and given an overview of the activities in the session. After this short introduction, the definitional knowledge and intuitive pre-test were administered electronically.
2. *Interaction with Collision (set at 1 hour and 40 minutes).* After the introduction, participants individually worked with the Collision environment. Two experiment leaders were present and available for answering questions concerning operating the environment, but not for answering questions concerning subject matter. Students were encouraged to use the full time available for the interaction. If they wanted to finish earlier they were asked to explore more of the environment, but were not forced to do so.
3. *Post-tests (40 minutes).* After the interaction with the learning environment the post-tests were administered. The definitional knowledge test was administered first, followed by the intuitive test, and the scientific reasoning test. The scientific reasoning test was administered using paper and pencil, the other two tests were given electronically.

5.4 Predictions

Students in both conditions were supported in the learning process by the heuristics that were incorporated in the learning environment. The difference between the conditions is that heuristics remain implicit for the student in the implicit condition, and are implicit as well as explicit for the students in the explicit condition. Even though the students in the implicit condition will not receive explicit information about the heuristics, they are still supported by the implicit presence of the heuristics. It is expected that this support allows the students to acquire knowledge about the domain, but it is less obvious that they will learn about the heuristics themselves.

For acquiring knowledge about the heuristics they have to be aware that there is something to be learned from the implicit heuristics, and to process them in a meaningful way. The first prerequisite is met because the fading of the heuristic support over the levels triggers the idea that something can be learned. The second prerequisite, processing the heuristics in a meaningful way, is less likely to occur. It is therefore expected that this knowledge will be more of a procedural nature as prescribing things to do, and not integrated with their already existing knowledge, that can be used to make context related decisions about what to do.

Because the students in the explicit condition are also supported by the heuristics, the expectation is that they too will be able to acquire knowledge about the domain. Apart from that it is expected that, because of the explicit presence of the heuristics, the students in this condition will also learn about the heuristics.

These ideas lead to the following three predictions.
The first prediction is that differences between the conditions are expected to be found on the scientific reasoning test, and on the experiment design question. The assumption is that knowledge about the heuristics makes it easier for students to interpret experiments and answer questions about these experiments, and that it provides a framework for setting up a research design in the greenhouse in the experiment design question.

The second prediction is that students in both conditions will acquire both definitional and intuitive domain knowledge. With respect to differences between the two conditions no strong predictions will be stated. The reason is that predictions can be stated in both directions. Making the heuristics explicit to the students might lead to a better understanding of the goals of the assignments, and the purpose of the experiments. This would lead to better learning outcomes for the students in this condition. It could also be that learning about the heuristics draws the attention from the learners away from learning about the domain, and lead to better learning outcomes for the students in the implicit condition.

The third prediction is that students are also expected to differ in the way they interact with the learning environment. It is expected that students in the explicit heuristics condition will be spending more time on the first level where they first encounter the heuristics. The students in this condition will be more likely to make mindful abstractions on these heuristics on this level and these are expected to take time. Differences in behavior are also expected on the later levels.

5.5 Results

The results of the analyses of the knowledge and process measures will be presented in the next sections. The general results of the knowledge measures will be presented first, followed by relations between the post-test results, and relations between pre-test results and post-test results, then the process measures will be presented, and finally relations between the process and knowledge measures.

5.5.1 The Knowledge Measures

Two independent raters rated the experiment design question on a scale from 1 to 5. Afterwards they reached agreement on their ratings to come up with the final score. Inter-rater reliability was computed between the raters ($k = .81$) and between each rater and the final score (rater 1-final, $k = .84$, rater 2-final, $k = .90$). These can all be considered high. The results on the scientific reasoning test and the experiment design question were not in line with the expectations. It was predicted that this test would show a difference between the two conditions, but this expectation was not sustained. Students in both conditions scored high on the scientific reasoning test with no difference between the two conditions. The experiment design question did not show any differences. Table 5-3 presents an overview of the mean scores and standard deviations for the different knowledge tests in the different conditions.

The *definitional knowledge* test was administered before and after the session. Reliability analysis (Cronbach's α) resulted in a reliability of .37 (N = 29; n = 20 items) for the pre-test and .62 (N = 30; n = 20 items) for the post-test.

The *intuitive knowledge* test was administered in two parallel versions as pre- and post-test. Reliability analysis of the pre- and post-test resulted in reliabilities of .70 (N = 30; n = 24 items) and .66 (N = 30; n = 24 items).

It was predicted that students in both conditions would gain both definitional knowledge and intuitive knowledge, while working with the learning environment. Table 5-4 shows paired samples T-test and effect sizes for the comparisons between the pre-test scores and the post-test scores for the definitional test and intuitive knowledge test. It also lists the effect sizes. The knowledge gain on both definitional and intuitive knowledge was quite large in both conditions with effect sizes of 1.46 (implicit) and 2.18 (explicit) for definitional knowledge, and 1.15 (implicit) to 1.40 (explicit) on intuitive knowledge. For both definitional and intuitive knowledge the effect size is larger in the explicit heuristics condition than in the implicit heuristics.

No strong predictions were stated about differences between the conditions on the definitional knowledge test or the intuitive knowledge test. No differences were found between the two conditions on the definitional post-test ($t = -.60$, $df = 27$, $p = .56$). or between the two conditions on the intuitive post-test ($t = -.29$, $df = 27$, $p = .77$).

The definitional knowledge test and the intuitive knowledge test are assumed to measure different types of knowledge in students (Swaak, 1998). The low correlation between the two pre-tests (implicit condition .11, explicit condition -.03, both conditions .08) supports this assumption. Further evidence for this assumption comes from the pre-test scores on definitional knowledge test and the intuitive knowledge test when the two schools are compared. On the definitional knowledge there was no difference between the students of the schools ($t = -.02$, $df = 13.2$, $p = .99$), but on the intuitive knowledge test students from one school scored higher than students from the other ($t = -2.89$, $df = 27$, $p = .007$). There was no difference between the conditions because the students were randomly assigned to the conditions within each school.

Table 5-3. Mean scores and standard deviations for the different knowledge tests in the different conditions. Standard deviations are given within parentheses.

	Condition					
	Experimental		Control		Total	
	Pre	Post	Pre	Post	Pre	Post
Definitional (max = 20)	10.9 (2.1)	15.6 (2.4)	10.6 (3.1)	15.2 (2.8)	10.8 (2.6)	15.4 (2.6)
Intuitive (max = 24)	16.9 (3.8)	21.1 (1.9)	16.8 (3.9)	20.7 (2.9)	16.9 (3.8)	20.9 (2.4)
Scientific reasoning test (max = 15)		12.6 (1.1)		12.4 (1.1)		12.5 (1.1)
Experiment design (max = 5)		3.1 (1.5)		3.0 (1.0)		3.0 (1.2)

Table 5-4. Paired samples T-test between pre- and post-tests and effect sizes (d).

	Definitional test				Intuitive test			
	t	Df	p	d	t	df	p	d
Explicit	5.87	13	.000	2.18	4.63	13	.000	1.40
Implicit	6.12	13	.000	1.46	6.04	14	.000	1.15
Total	8.55	27	.000	1.78	7.50	27	.000	1.28

Based on the overall results one might jump to the conclusion that the two learning environments in this experiment are interchangeable. It is, however, not sufficient to look at mean test scores only to draw this conclusion. The inter-relations between these scores and the way that students work with the learning environments must also be examined. The next sections will investigate whether the two learning environments are indeed interchangeable, or if there is a more complicated relationship between the learning environments and the outcomes on the post-tests. This will be done by looking at relations between pre- and post-test results, the way that students work with the learning environment, and relations between the way that students work and the post-test results.

5.5.2 Relations between pre- and post-test knowledge measures

The relations between the pre-test scores and the post-test scores tell something about the way the students respond to the learning environments. When students respond to the learning environment in a similar way, the ranking of the students will remain more or less the same. In this situation a high correlation between (similar) pre- and post-tests would be expected.

Table 5-5. Correlations between the pre-test scores and the post-test scores on the knowledge tests.

		Definitinal post-test	Intuitive post-test	Scientific reasoning test	Experiment design
Definitional pre-test	Explicit	-.11	.40	.50	.09
	Implicit	.59*	.17	-.10	.09
Intuitive pre-test	Explicit	.44	.45	.61*	.26
	Implicit	.52*	.57*	-.13	.43

Note. All correlations are Spearman correlations; * means p<0.05, ** means p<0.01.

As seen in Table 5-5, in the implicit heuristics condition the correlations between the definitional pre- and post-test (.59), and the intuitive pre- and post-test (.57) both suggest that the students responded to the treatment in more or less the way that one might expect. In the explicit condition this is not the case. In this condition the pre-test score is not related with its post-test counterpart on the definitional knowledge test, and only moderately on the intuitive knowledge test. The scientific reasoning test is moderately related with the pre-test score on the intuitive knowledge test, and somewhat less with the definitional pre-test score in the explicit heuristics condition. The experiment design question is weakly related with the pre-test score on the intuitive knowledge test in the implicit condition.

These results can be highlighted with a regression analysis in which the pre-test scores are used to predict the post-test scores. The result of the regression analysis will show how well the post-test scores can be predicted from the pre-test scores and to what extent the different pre-tests contribute to the prediction.

Table 5-6. Regression analyses predicting definitional knowledge post-test scores based on pre-test scores on the knowledge tests.

	Explicit					Implicit				
Definitional test	Sum Sq.	Df	Mean Sq.	F	Sig.	Sum Sq.	df	Mean Sq.	F	Sig.
Regression	19.7	2	9.87	1.92	.193	50.4	2	25.2	4.74	.033
Residual	56.6	11	5.15			58.5	11	5.31		
Total	76.4	13				109	13			

Note. The regression functions for predicting the post-test score are:
Explicit: definitional post = 9.63 + 0.06 * definitional pre-test + 0.33 * intuitive pre-test.
Implicit: definitional post = 5.67 + 0.52 * definitional pre-test + 0.23 * intuitive pre-test.

Table 5-6 shows the results of the regression analysis for the definitional knowledge. It shows that the definitional post-test result can be predicted based on the pre-test results in the implicit condition. The regression function shows that this prediction is mainly derived from the definitional knowledge, although the intuitive knowledge contributes to the prediction as well. In the explicit condition the regression analysis is not significant, but it is striking that the definitional pre-test knowledge hardly contributes to the prediction.

Table 5-7 presents the results of the regression analysis for the intuitive knowledge. In the implicit condition the results intuitive post-test can be predicted based on the pre-test results. This time the prediction is almost solely derived from the intuitive pre-test. Again, in the explicit condition the regression analysis is not significant.

Table 5-7. Regression analyses predicting intuitive post-test scores based on pre-test scores on the knowledge tests.

	Explicit					Implicit				
Intuitive test	Sum Sq.	Df	Mean Sq.	F	Sig.	Sum Sq.	df	Mean Sq.	F	Sig.
Regression	14.6	2	7.28	2.64	.116	70.8	2	35.4	9.61	.004
Residual	30.4	11	2.76			40.5	11	3.68		
Total	44.9	13				111	13			

Note. The regression functions for predicting the post-test score are:
Explicit: intuitive post = 13.97 + 0.30 * definitional pre-test + 0.23 * intuitive pre-test.
Implicit: intuitive post = 10.32 + 0.06 * definitional pre-test + 0.57 * intuitive pre-test.

5.5.3 Process measures

It was expected that the students in the two conditions would differ in their interaction with the learning environments, and that these differences would be most clear on the first level, when the students are confronted with the implicit, and / or explicit heuristics for the first time. Students in the explicit condition were expected to spend more time on this level.

Table 5-8 shows some indicators of how the students worked with the learning environment on the first level. There was no difference between the time that students worked on this level, and the time that they worked on assignments on this level. There was a difference in favor of the implicit heuristics condition in terms of the number of assignments that were answered correctly on the first attempt. The total number of experiments with the simulation during the level was not different. The number of unique experiments seems higher in the explicit condition, although this difference is not significant. Students in the implicit heuristics condition drew graphs in more assignments then the students in the explicit heuristics condition. They also literally followed the experimentation heuristics in assignments 3 to 6 more often than the students in the explicit heuristics condition. If the unique experiments within assignments are aggregated over the level, no difference between the two conditions is found.

Table 5-8. Process data from the first level. Means, standard deviations and T-test comparison.

	Implicit		Explicit		T-test		
	Mean	Sd	Mean	Sd	t	df	p
Total time (s)	2289	(373)	2292	(558)	-0.02	27	.987
Total experiments	29.7	(10.2)	32.4	(9.6)	0.75	27	.459
Total unique experiments	12.4	(4.79)	16.3	(6.83)	-1.78	27	.086
Time on assignments (s)	1552	(300)	1521	(443)	0.22	27	.827
Correct on first answer (of 5)	4.6	(0.62)	3.9	(0.63)	2.89	27	.007**
Unique exp. in assignments	14.4	(2.4)	16	(3.4)	-1.46	27	.155
Standard experimentating (of 4)	3.27	(0.96)	1.79	(1.37)	3.35	23.2	.003**
Assignments with graphs	3.13	(0.64)	2.36	(1.22)	2.49	27	.046*

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

After the first two introductory assignments, the students are confronted with the heuristics for the first time in assignments 3 to 6. In each of these assignments the students are asked to do a series of experiments to investigate a certain relation between two variables. In the overall results it appeared that there was a difference between the students in the two conditions in terms of following the experimentation heuristics literally. Table 5-9 shows these results in more detail. Students were assigned standard or non-standard groups for analyses based on strict criteria. A student who conducted exactly the same experiments as proposed in the assignment was scored as standard. A student who conducted more, less, or different experiments was scored as non-standard.

Table 5-9. Number of students following instructions in assignments on the first level.

	Explicit		Implicit	
Assignment	Standard	Non-stand.	Standard	Non-stand.
A 3 tangent x(t)-velocity	8	7	13	2
A 4 mass-velocity	2	13	8	7
A 5 force-velocity	6	9	14	1
A 6 formula velocity	10	5	14	1

What can be seen from these results is that almost all students in the implicit condition followed the experimentation heuristics literally on assignment 3, 5 and 6. Only on assignment 4 about half of the students deviated from the heuristics in the assignment. The students in the explicit heuristics showed quite different behavior, with about half of the students following the heuristics literally on assignment 3, 5 and 6, and almost no one on assignment 4.

In assignments 3 to 5 students are also given the “draw a graph” heuristic. They are asked to draw a graph of the results of their experiments, making the interpretation of their results easier. Table 5-10 shows these results in more detail.

Table 5-10. Number of students that drew graphs during assignments on the first level.

	Explicit		Implicit	
Assignment	Graph	No graph	Graph	No graph
A 3 tangent x(t)-velocity	10	5	12	3
A 4 mass-velocity	12	3	15	0
A 5 force-velocity	12	3	15	0
A 6 formula velocity	2	13	5	10

Although it is less strong for the graphs, a pattern similar to the pattern in experimentation is found. Students in the implicit heuristics condition are using this heuristic more than students in the explicit heuristics condition.

These results suggest that there is an effect on the way that students work with the learning environment from the learning environment condition (explicit or implicit). The students in the explicit condition seem to be more self-regulating, whereas the students are more regulated by the environment. This difference does however not propagate clearly to levels three (elastic collisions) and four (inelastic collisions)[1]. The students can no longer be differentiated when compared on similar process measures on these levels.

5.5.4 Relations between process and knowledge measures

In the previous section the students from the two conditions were compared to their behavior while working with the learning environment. The results showed that there were differences in the way students worked at the first level, but that these differences did not propagate through to the next levels. Does this mean that these initial differences only exist in the beginning, and that the students in both conditions are working and learning in the same way in later levels? This section focuses on the relations between the process

[1] In the explicit condition of the first session students could not change the value of one of the variables on the second level as a result of a technical problem. This makes the results of the second level not fully comparable in both sessions. The results will therefore only be shown for levels three and four.

measures and the knowledge measures in an attempt to answer this question. The general idea is that if the students are working and learning in a similar way on these levels, the relations between their way of working and the results on the tests should be comparable. Table 5-11 and Table 5-12 show the relations between the post-test results and the process measures on levels three and four respectively.

Table 5-11. Correlations between the post-test scores and process measures of the third level.

		Total exp.	Unique exp.	Assign. graphs	Assign. exp.	Unique exp in ass
Definitional post-test	Explicit	.17	.45	.50	.57*	.28
	Implicit	-.15	-.28	.09	-.64**	**-.52***
Intuitive post-test	Explicit	-.06	.25	-.01	.34	.37
	Implicit	-.20	-.05	-.36	-.64*	-.30

Note. All correlations are Spearman correlations; * means p<0.05, ** means p<0.01.

The results in Table 5-11 show a number of differences between the two conditions. Most striking is the difference in the relation between the number of assignments in which students did experiments, and the definitional post-test score. In the implicit condition this relation is negative, and in the explicit condition it is positive. Negative relations are also found between the number of assignments in which students did experiments, and the intuitive post-test score, and between the aggregated unique experiments and the definitional post-test score in the implicit condition. In the explicit condition both these relations are slightly positive. This is also the case for the total number of unique experiments on this level, and the number of assignments in which students make a graph.

Table 5-12. Correlations between the post-test scores and process measures of the fourth level.

		Total exp.	Unique exp.	Assign. graphs	Assign. exp.	Unique exp in ass
Definitional post-test	Explicit	.49	.04	.62*	.41	.41
	Implicit	-.32	-.58*	-.10	-.43	-.65**
Intuitive post-test	Explicit	.21	-.17	.18	.19	-.01
	Implicit	-.33	-.45	-.37	-.21	-.36

Note. All correlations are Spearman correlations; * means p<0.05, ** means p<0.01.

On the fourth level there are also a number of differences between the two conditions as can be seen in Table 5-12. Again a negative relation can be seen between the aggregated unique experiments within assignments and the definitional post-test score in the implicit condition. In the explicit condition this relation is positive, although not significant. The number of unique experiments and the definitional post-test score are also negatively related in the implicit condition. In the explicit condition there is a positive relation between the number of assignments in which students draw a graph, and the definitional post-test score. In general relations between process measures and definitional post-test score tend to be negative in the implicit condition, and positive in the explicit condition. For the intuitive knowledge post-test scores, these relations are present as well, but not as strong. A tendency towards a negative relation can be seen in the implicit condition only.

5.6 Conclusions

Two learning environments were compared in this experiment. One that supported students with heuristics in an implicit manner, and one in which this support was also presented in an explicit manner. With respect to the main prediction the expected result was not obtained. It was expected that students who received explicit heuristic support would be more likely to make mindful abstractions about the heuristics and integrate them with their own knowledge. As a result these students would be able to draw on this knowledge when they had to interpret experiments, answer questions about these experiments and also when they had to design an experiment. It was therefore expected that they would score higher on the scientific reasoning test and on the experiment design question. The results could not sustain these expectations. No differences were found on the scientific reasoning test, or the experiment design question.

One of the reasons is that the scores on the scientific reasoning test were high in both conditions. It seems that the students in both conditions had no trouble interpreting experiments. The high scores together with the low variance suggest a ceiling effect. For the comparison of the students in the two conditions this means that, even if the students differed, the test used in this study was not suitable to detect a difference.

The experiment design question did not show any differences either, but it seems to be a much more viable candidate for testing heuristic knowledge in students. Even though there was only one such question, this one question already generated as much variance between students as all multiple-choice questions did. One difficulty, however, was that some students gave very short answers, whereas others gave more elaborate answers. This made it difficult to assess the quality of the answers. Using more questions and, especially, questions that focus on part of the research process might help to overcome this problem. For future research it is suggested not to use a multiple-choice test, but a set of experiment design questions, that ask students not only for an answer, but also for an explanation why they came up with that particular answer.

Before and after working with the learning environment the students were tested on their knowledge about the domain. They were tested on both definitional and intuitive knowledge about the domain. These two types of knowledge are assumed to be of a different nature (Swaak, 1998). The definitional knowledge is about the static relations in a domain (knowing the underlying definitions and formula's in a domain) and intuitive knowledge about the dynamics of a domain (knowing what will happen in a certain situation). The results from this experiment provide further support for this claim. At the start of the experiment the students from both schools had similar scores on the definitional knowledge test, but not on the intuitive knowledge test. This is a clear sign that the two tests measure different types of knowledge in the students. At the same time the correlation between the two pre-tests was rather low. This is not a surprising result, with a difference between schools on the intuitive knowledge test, and not on the definitional knowledge test, but when the schools are separated, thus removing the difference on the intuitive knowledge test, a low correlation between the pre-tests is also found in each school. It can therefore be safely concluded that the knowledge measured with the definitional knowledge test is different from the knowledge measured with the intuitive knowledge test.

At the end of the experiment the students were again tested on their definitional and intuitive knowledge to see whether there were differences between the students in the two conditions, and also to see how much they learned from working with the environment. The results show no differences between the two conditions on the post-test scores, and a considerable gain (large effect sizes) on both definitional and intuitive knowledge from pre- to post-test for the students in both conditions. In both cases (definitional and intuitive knowledge) the effect size was larger in the explicit condition than in the implicit condition.

Based on the results on the knowledge test one might conclude, that the two learning environments were not really different, or in other words, that they were interchangeable. Before drawing this conclusion it is necessary to look beyond mean scores on the post-tests alone. Even though these are similar, there might still be differences on the individual level, with respect to how students respond to the learning environments.

One indication for these kinds of differences comes from the relations between the pre-test scores and the post-test scores. When these relations are compared differences between the two conditions appear. The most striking being the strong relation between the definitional pre-test score and the definitional post-test score in the implicit condition, and the absence of such a relation in the explicit condition. The relation between the intuitive pre-test score and the intuitive post-test score is also stronger in the implicit condition than in the explicit condition. A similar relation was found between the intuitive pre-test score, and the definitional post-test score.

The regression analyses that predict the post-test scores based on the pre-test scores show that these predictions are better in the implicit condition. This is related to the fact that in the explicit condition the students who score low on the pre-test are not necessarily the students who score low on the post-test, and the students who score high on the pre-test are not necessarily the students who score high on the post-test. Students from the lower ranges on the pre-test are gaining more knowledge than students from the higher ranges. This is not so strange since students that score higher on the pre-test have less knowledge to gain, but it is remarkable that the effect is so strong in the explicit condition that students scoring lower on the pre-test are passing students scoring higher on the pre-test.

If the definitional knowledge is taken as the discriminator between strong and weak students, the explicit condition changes this ranking from pre- to post-test more than the implicit condition thus favoring at least part of the weaker students. Even if the intuitive knowledge test is taken as the discriminator between strong and weak students, it is the case that the explicit condition favors the weaker students more than the implicit condition.

The students were also compared with regard to the way they work with the learning environments, to see if there are differences between students in one condition and students in the other. Since the learning environments are equivalent with the exception of the implicit or explicit heuristics, any difference in behavior can be attributed to this difference. The analysis of the behavior of the students shows that there are some differences in behavior at the start, but not the predicted difference in time spent on the first level. Students in the implicit condition are carefully following the implicit heuristics at the start. Students in the explicit condition are much more likely to deviate from the implicit heuristics making their own plans. They make their own decisions about experimenting (more, less, different), and about whether they need to draw a graph. Even though at the start the learning environments seem to trigger different behavior in the students, this difference is not clear anymore later on. When the behavior of the students in the two conditions on the third level and fourth level is compared, there seem to be no differences. This could mean that the students in the implicit condition start working according to their own plans as well, that they just needed a bit more time to find their own way of working, but that after that they are comparable to the students in the explicit condition. The relations between the process measures and the post-test scores do not support this idea. The general tendency seems to be that in the explicit condition students that are still active on the third and fourth level, by experimenting, and drawing graphs of their experiments, are scoring high on the definitional knowledge test. The implicit condition shows a relation in the other direction; here the active students are students that score low on the definitional knowledge post-test. This could indicate that there are students in the implicit condition that do not really know how to deal with the implicit heuristics. They follow them, but they do not really grasp the ideas behind them, resulting in lower scores on the definitional knowledge test at the end. The relations between the process measures and the intuitive knowledge are not that clear, although there seems to be a tendency in the same direction. One reason for this less strong relation could be related to the relatively large differences between students on the intuitive knowledge before the start. It could be that even though there is a relation between the process measures and gaining intuitive knowledge, that this relation is partly hidden by the differences that already existed at the start.

In general it can be concluded that the learning environments used in this study both worked very well in supporting the acquisition of domain knowledge. Using heuristics in the design of a learning environment seems to be a promising approach. With respect to scientific reasoning, the conclusion is that the students either already possessed this knowledge, or that both learning environments do equally well on teaching them to the students. With respect to the heuristics, no firm conclusions can be stated. There is slight evidence that the explicit heuristics triggered more self-regulation, which would mean that the heuristics are incorporated in already existing knowledge structures. Whether this is really the case could be investigated in a study in which students with a discovery learning transfer task without support.

6

Conclusion

6.1 Conclusions

This thesis started from the idea that there are two potentials in simulation-based discovery learning environments. The potential of discovery (learning domain knowledge) and the potential of transfer (learning discovery skills).

The conditions for realizing these potentials are that students employ discovery learning activities while interacting with the simulation and, that this interaction will result in learning domain knowledge, and discovery learning skills. The history of simulation-based discovery learning shows that these conditions are not easily met. One finding was that working with the simulation did not always result in learning domain knowledge. An important reason for this finding is that the processes needed for discovering knowledge in a simulation are quite complex. The goal for this thesis was to investigate whether it was possible to develop intelligent support for learners in a simulation-based discovery learning environment that could be included in an authoring environment for developing these environments. The latter implies that the principles used should be general, and that no strong assumptions can be made about the content of the domain.

The SIMQUEST authoring environment already contained means to augment the learning environment to support learners (by using model progression, assignments, feedback, and the control structure). However, assessment *of* and feedback *on* the learner's experimenting in relation to testing a hypothesis was not yet supported.

Assessing experimentation is not trivial, nor is giving feedback, since there is no single correct solution as to which experiments should be done in a certain situation. Apart from having more than one possible solution, the kinds of possible solutions also change in the presence or absence of a hypothesis that they are supposed to test. To make the task more tractable it was decided to focus on the process of testing predefined hypotheses and drawing conclusions about these hypotheses.

The first approach that was taken was to explore the possibility to use ideas from ITS to support discovery learning, as these kinds of systems aim at adapting instruction to an individual learner, which is what is needed for assessing a learner's experimenting and giving individualized feedback.

The usual way to assess learners in an ITS is to construct a model of the learner and to compare this model to an expert model. The result of this comparison is then used to generate instruction that can be presented to the learner. In the case of experimenting by learners this would mean that the system builds a model of the experiments that the learner has done, and has an expert model of how experiments should be done, and that these two are then compared.
This leads to two problems.

The first problem concerns the model of the learner, just like there is no single set of ‘correct experiments to test a certain hypothesis, there is also no single hypothesis related to a set of experiments. The problem becomes even harder when the set might not be the correct set for testing a hypothesis. Without knowledge of the hypothesis that the set is supposed to test, there is no way of constructing a comprehensive model of the learner’s experimentation; there are too many unknowns.

The second problem concerns the expert model. Even if we know the hypothesis that the experiments are supposed to test, there is no single expert experiment set. Many experiment sets can be constructed that are equivalent in their power to test the hypothesis.

A solution to the first problem that was adopted in this thesis was to provide learners with hypotheses, and to focus on hypothesis testing. Hypotheses that learners could test in the learning environment were specified in advance, and assessment of a learner’s experiments was done in relation to a specific hypothesis chosen by the learner. This made the task easier in two ways. Firstly, there was a guarantee that hypotheses would be syntactically sound, and secondly, the hypothesis that these experiments were meant to test would be available in the analysis of the experiments.

A solution for the second problem was found in reversing the problem. The problem was that there is no single correct set of experiments to test a hypothesis, so no single set can be identified that can be compared to the learner’s experiments. If the problem is reversed, it is possible to check whether the learner’s experiments could be one of these many sets of experiments. If it is not one of these many sets; this is valuable information that can be communicated to the learner. If it is one of these sets then the experiments can be analyzed to see whether the hypothesis should be rejected or not.

Based on these ideas, a tool was designed that provides support for testing hypothesis and drawing conclusions. The effects of the tool on discovery learning and on the learning outcomes were investigated in an experimental study. This study raised some issues on the design of the first version of the tool.

The first version of the tool used principles of induction and deduction to analyze the learners’ experiments in relation to their hypothesis. In order to give specific feedback about the learners’ experiments, hypotheses had to be stated in a way that a formal analysis could distinguish between experiments that were suitable to test a hypothesis and experiments that were not. This was realized by using semi-quantitative hypotheses that have a condition part that can be matched against the experiments. Although this choice allowed for more specific feedback, it may not have been ideal for the learners. A problem with the semi-quantitative hypotheses is that they are less familiar to learners, and that an extra transformation is needed to convert them into quantitative hypotheses or formulas that can be mapped onto the underlying model of the domain.

Another issue was the absence of definite feedback about the correctness of the hypotheses. Definite feedback would only be provided when a hypothesis was proven wrong by their experiments. In other cases the feedback remained uncertain, stating that the hypothesis can not be rejected on the basis of the experimental evidence. In the absence of definite feedback about the correctness of a hypothesis, the learners had to rely on their own assessment of their experiments, and whether they were sufficient to believe in the correctness of a hypothesis. It is likely that this made it more difficult for the students to acquire domain knowledge. It is also likely that the absence of feedback made the acquisition of discovery skills more difficult. Feedback about the correctness of the hypothesis could help learners to assess their experiments in the light of a correct or incorrect hypothesis. It could help them realize when and why their experiments were insufficient to prove an incorrect hypothesis wrong. In this case the feedback provides an external validation criterion that they can use to assess their discovery learning processes.

Furthermore, the tool was maybe too formal in the analysis of the experiments and in the feedback that was given to the learner. This might have given the students the impression that this was *the* correct way of investigating relations between variables, choosing hypotheses, designing experiments to test these hypotheses, and drawing conclusions. This is not the message that students should get from working with an environment such as this. Even though they would learn a procedure that they may be able to apply it would not be meaningful, as they probably do not know when and why it can be applied.

In an attempt to deal with these issues the design of the tool and the learning environment as a whole were reconsidered and heuristics related to discovery learning were used to redesign both the tool and the learning environment.

The monitoring tool, which was first only a storage place for experiments, was extended with functionality to support the interpreting of these experiments. Learners were given the possibility to draw graphs, fit qualitative and quantitative relations on the experiments, and create new variables based on existing ones. Support for testing a hypothesis was provided after drawing a graph by analyzing the experiments using both heuristics and formal analysis, and using the results to generate feedback. In the feedback the emphasis was on the heuristics rather than on the formal analysis. This resulted in feedback that had a more informal character. The learner was reminded of heuristics related to hypothesis testing, but not prescribed what to do, merely stimulated to make an assessment of the value of the feedback and what to do with it.

The learning environment was redesigned using the heuristics to guide decisions about the content of the learning environment. The “simplify the problem” heuristic was used to present simpler collisions at the beginning of the level. The in combination with the “slightly modify hypothesis”, the “set expectations” could than be used to connect these simpler situations to more complex ones, but also to connect the elastic and the inelastic collisions. Within the assignments that came out of this process, heuristics were used to set up a structure that started with the “identify hypothesis” heuristic, used the hypothesis testing heuristics to set up suitable experiments, the “draw graph” heuristic to interpret the results, and the “confirm hypothesis” to test if the hypothesis was correct. A scaffolding approach was used for the heuristics in the assignments. They were fully specified in the beginning and specified in general terms in the end.

In the next section the results of the studies in which the two versions of the tool where tested will be discussed. The discussion will be structured around the goals of the tools and the learning environments, and will describe the results of the studies, and try to compare these results.

6.1.1 The learning goals

There were two main goals with respect to learning for the tools and the learning environments that were used in the studies that were described in Chapters 4 and 5. The first is related to learning domain knowledge and can be divided into learning the static definitional domain knowledge in the form of laws and formulas, and learning the intuitive domain knowledge that reflects a qualitative, more dynamic understanding of the domain. After working with the simulation environment the learners should have acquired knowledge in both respects.

The second is related to learning discovery learning skills, the skills that are required for obtaining definitional and intuitive knowledge through exploring a certain domain. Learners should acquire some of these skills while working with the environment.

In order to obtain these results the learners should have the opportunity to acquire this knowledge and skills during their interaction with the learning environment.

To be able to learn discovery skills, students should engage in the discovery learning processes. This means that the learning environment should stimulate the students to explore the domain. They should explore relations between variables in the domain by investigating hypotheses about these relations, setting up experiments to test the hypotheses, and drawing conclusion about the hypotheses in the light of the experimental evidence. If learning does not involve these activities, then the chances of obtaining results in this direction will not be very high.

For the acquisition of definitional knowledge of the domain it means that the learner should be able to discover the laws and formulas of the domain. Although it might be the ultimate goal that learners explore the whole domain by themselves, it is not likely that all learners will be able to do so. To enhance the chance that learners explore all the relations in the domain, the support in the environment should be organized in a way enables the learners to cover the whole domain.

In relation to the acquisition of intuitive knowledge it means that the students should be exposed to many different situations in order to be able to develop a more intuitive understanding of the domain. If students do not encounter many different situations it will be harder for them to develop a more intuitive understanding of the domain as intuitive understanding is closely related to experience.

If the above conditions are met, the learners should be able to acquire definitional and intuitive domain knowledge, and discovery skills, while interacting with the learning environment. This is the ideal situation, and it is of course no guarantee that these results will actually be obtained. The next section will look back at the two studies and the conditions in these studies, and will try to describe the studies and conditions in terms of the two goals.

6.1.2 Discovery learning skills

The first study compared a learning environment that provided learners with intelligent feedback with one that provided learners with pre-defined feedback. It was investigated whether giving intelligent feedback influences the learner's discovery behavior and / or learning outcomes.

In this study there were indications for differences between the experimental condition and the control condition with respect to the way the students were working with the learning environment, were learning from the learning environment.

The two conditions had the same model progression levels, and assignments, and only differed in the feedback that the students received on their answers to the assignments. The students in the control condition received pre-defined feedback that identified a hypothesis to be correct or incorrect. The students in the experimental condition received feedback that evaluated their experiments in the light of the hypothesis and stated whether the hypothesis should be rejected based on these experiments, or that the hypothesis could not be rejected. When a hypothesis could not be rejected, it could either be that the experiments were not sufficient to test the hypothesis, in which case the students received feedback that provided support for designing experiments to test the hypothesis, or that the experiments were in line with the hypothesis.

It could be argued that for the students in the experimental condition it was necessary to do experiments in order to be able to acquire knowledge about the domain, and that this was not the case in the control condition. Students in the control condition could also acquire knowledge by answering assignments, and learn from the answers and the feedback without experimenting.

This means that the students in the control condition had two options; learn through experimenting in the same way as the learners in the experimental condition, or learn from the assignments and answers. If learners choose the first option they have to engage in the discovery learning processes, and possibly acquire knowledge and/or skill related to these processes while working with the learning environment. If they choose the second option it is much less likely that they will do so.

The results showed that in the control condition the post-test results correlated with assignment use, and not with experimenting behavior. This could be an indication that at least part of the learners learned from the assignments and answers, and not from exploring the domain. In the experimental condition correlations were found between experimenting and the post-test results and not between assignment use and the post-test results.

Another finding was that the students in the experimental condition experimented more in general, more during assignments, did more unique experiments, and tested more hypotheses, compared to students in the control condition. These figures are also indications that the students in the experimental condition employed more discovery learning activities than students in the control condition.

A speculative conclusion based on these results is that: If there were a difference between the students in the two conditions it would most probably be in favor of the students in the experimental condition.

The second study compared two learning environments that included heuristics to support the learners on the discovery processes. In one environment the heuristics were only implicitly included, in the other environment both implicitly and explicitly.

The scientific reasoning test and the experiment design question, which were used in the second study, were meant to reveal more about the discovery skills of the students. The test was given as a post-test only, and therefore no data on change of this knowledge as a result of working with the learning environment were available. The results on the scientific reasoning test were that students scored very high on this test in both conditions. A ceiling effect was found on this test in both conditions. One explanation for finding this result could be that this was a result of working with the learning environment, which would be a very good result. Another explanation could be that the students already possessed the knowledge, and that the results would have been just as high without working with the learning environment.

No ceiling effect was found on the experiment design question, but no differences between the conditions either. In general the students' designs were fairly good. From the 30 students, only four came up with a design that can be classified as confounding. After a real experiment according to the set-up of these students no conclusions could be drawn about the influence of the three factors. In all other cases it would be possible to draw conclusions about the influence of the factors. Here too, it would be good if this was a result of working with the learning environment, but of course it could be that the same results would have been found without working with the learning environment.

What can not be answered is whether these answers reflect a real understanding of experiment design, or merely copying the examples from the learning environment. Maybe the finding that the more active discoverers in the implicit condition scored low and the more active discoverers in the explicit condition scored high could be interpreted as an indication that there was more understanding in the explicit condition than in the implicit condition.

A comparison between the two studies is not straightforward, but it might be interesting to look at the two studies, and try to compare them.

Students in the experimental condition of the first study needed to do experiments in order to learn about the domain. The students in the control condition could avoid experimenting by using assignments and answers only. There were indications that this indeed happened. In the second study students in both conditions could avoid experimenting by relying on assignments and answers only. In this second study the averages of the number of experiments, and the number of unique experiments were between the averages of the first study. It seems that students were less explorative than students in the experimental condition of the first study, but more explorative than students in the control condition of the first study. Even though students in the second study could have reverted to learning from assignments and answers they seemed less likely to do so in the second study than students in the control condition of the first study.

This does not necessarily mean that students in the experimental condition of the first study acquired more discovery learning skills than the students in the second experiment. Students in the first study did not get feedback telling them whether a hypothesis was true or false. Students had to rely on their own assessment of the correctness of a hypothesis in the evaluation of their discovery learning processes. The students in the second study did receive similar feedback when they drew a graph, but also definite feedback when they answered an assignment. This provided them with external feedback that they could use in the evaluation of their discovery processes.

6.1.3 Definitional and intuitive domain knowledge

One finding that was common among the two studies that were described in this thesis is related to the definitional and intuitive knowledge tests that were used in the two studies. The underlying assumption behind the use of these two tests was that they measure different kinds of knowledge in the learners, (see Swaak 1998 for an elaborate discussion about these differences). The results of the pre-tests on the definitional knowledge test and the intuitive knowledge test support the idea that the tests do indeed measure different types of knowledge. In both the first and the second study a low correlation was found between the scores on these tests. If the two tests were measuring the same type knowledge, a much stronger correlation would be expected. The conclusion after the two studies is therefore that the tests that were used in the studies are indeed measuring different types of knowledge in learners.

No differences were found in mean scores between the conditions within each of the two studies on the definitional knowledge test, or on the intuitive knowledge test. In both cases students were comparable at the beginning, and at the end. In both cases there students gained considerably from pre- to post-test. There was however a difference between the first and the second study. The students in the second study scored higher on both post-tests in comparison to students in the first study. It seems that the learning environments that were used in the second study enabled the students to acquire more definitional and intuitive domain knowledge, than the learning environments in the first study did.

It is difficult to pinpoint exactly what caused these differences. The students in the control condition of the first study did, for instance, receive true-false feedback on the assignments, just like the students in the second study, making that an unlikely cause for the difference. As a result of the absence of predefined feedback, students in the experimental condition of the first study had to engage in discovery learning. Discovery learning alone is therefore also not a likely cause for the differences that were found. One thing that might have contributed is the nature of the hypotheses in the assignments. Hypotheses in the first study were qualitative or semi-quantitative. Students still needed to transform these hypotheses into quantitative hypotheses, before they could be mapped onto formulas and laws of the domain. This might have been a problem for at least part of the learners in this study. Hypotheses in the second study were qualitative or quantitative, and stated in a way that might have been more familiar to the students, and was certainly closer to the laws and formulas of the domain. Using heuristics to guide the design of the learning environment might also have contributed to the differences that were found. The use of the heuristics helped to define goals of the assignments, and guided the specification of their content in terms of discovery learning support for the learners. The result was a learning environment that covered the important relations in the domain in a way that probably allowed students to make connections between these relations and acquire knowledge about the domain.

6.1.4 Supporting the weak?

Even though the mean test scores in experimental conditions did not differ in either of the two studies, there were differences between the students in the two conditions in both of the studies. Differences in behavior were already discussed in relation to discovery skills, here differences in the relations between pre- and post-test knowledge will be discussed.

In the first study the correlations between the pre- and post-test for both definitional and intuitive knowledge were considerable in the control condition. In the experimental condition this was also the case for the intuitive knowledge, but the correlation for definitional knowledge was somewhat lower. Correlations between pre-test scores and the other post-test scores revealed that in the experimental condition the correlation between the intuitive pre-test score and the definitional post-test score was higher than the correlation between definitional pre-test score and intuitive post-test score. In general the intuitive knowledge had strong correlations with all the post-tests in the experimental condition and much less so in the control condition. For the definitional knowledge it was the other way around. Here strong correlations were found in the control condition and much less so in the experimental condition. To explore this result further the groups were split into low and high scoring students (this was done separately for definitional and intuitive knowledge), and then compared for all of the post-test scores. The results of this comparison showed that in the experimental condition the low and high scoring groups from definitional pre-test could not be separated any more after the post-test on any of the tests including the definitional knowledge post-test, and that the high scoring group from the intuitive pre-test outscored the low scoring group on all of the post-tests. The control condition showed a reversed effect, with the high scoring group on definitional knowledge outscoring the low scoring group on all but the intuitive knowledge test, and no differences on any but the intuitive knowledge between low and high on the intuitive knowledge test. These results suggest that there was a differential effect from the condition on the learning results.

The correlations between the pre- and post-test were also calculated in the second study. The results showed a picture that was slightly different from the results of the first study. The correlations between both definitional and intuitive pre- and post-test, and between intuitive pre-test and definitional post-test were high (and significant) in the implicit condition. In the explicit condition no correlation was found between the definitional pre- and post-test, and only moderate correlations in the rest of the cases.

It was not possible to do the analysis with the low and high scoring students in the second study. Instead a regression analysis was performed to see whether the post-test results could be predicted from the pre-test. Both post-test scores could be more accurately predicted in the implicit condition. Most striking was the finding that definitional pre-test did not contribute anything to the prediction of the definitional post-test score in the explicit condition.

This result tells us that in the implicit condition students with higher scores at the beginning were in general also scoring relatively high at the end. In the explicit condition this was not the case, here the population got mixed with some students with low scores at the beginning ending up scoring relatively high at the end, and some students with high scores at the beginning ending up relatively low at the end.

One might argue that it is not necessarily the weak that are supported in the experimental condition of the first study and the explicit condition of the second study, since the pre-test scores do explain at least some of the variance in the post-test scores. While this might be true, it is the case that different learners are benefiting from the environment than normally is the case, resulting in a smaller differences and changes in the rank order from pre- to post-test.

Learning environments like the ones used in this research might therefore be a welcome addition to traditional education. It favors a different population than traditional education, and hopefully leads to the development of discovery learning skills in learners as well. It might be well suited in the Dutch situation with its “studiehuis” that strives for more independent learning for the learners in schools. The learning environments provide a means for independent learning that would be a valuable addition to the curriculum.

6.1.5 Using heuristics to support learners

In the first study a formal analysis of experiments in relation to a hypothesis was used to generate dynamic feedback to the learners. This resulted in knowledge gain on both definitional and intuitive knowledge. In the second study a less formal approach was used, an approach based on heuristics for discovery. The strength of heuristics is that they can be general and therefore applicable in many different domains. The application of the heuristics is not always trivial. There might be domain specific knowledge needed for the application. Consider for example the “extreme values” heuristic in relation to temperature. An extreme value for temperature will be different in a context where the behavior of molecules is investigated compared to a context where growth of bacteria is investigated. This domain or context specific knowledge is not always available to learners who are new to the domain. It is however available to the designer of a learning environment. The designer can use the heuristics to guide the design of the learning environment presenting the learners with guidelines derived from the heuristics in the present context. By explicating the design decisions and the heuristics learners will get the opportunity to build or extend their own heuristic repertoire. This heuristic repertoire might be seen as a toolbox that is open to transfer to other domains.

The design of the second version of the tool showed that heuristics can also be used in combination with ideas from ITS. Including heuristics could extend the scope of the tool. The heuristics can be used as examples of good practice that can provide support for learners. Their uncertain nature makes that they can not be used in an obligatory way, but only in a non-obligatory way. Using them in a non-obligatory way enhances the chances that learners will not see them as procedures that they should apply, but as guidelines that you can use or not use depending on the context.

6.2 Future directions

In the first part of this chapter we looked back at the two studies that were described in this thesis, and drew some general conclusions. In the second part of the chapter we will face the other way and will try to say some things about the future. We will address some issues in relation to the measurement of discovery learning skills and/or heuristic knowledge, and we will further explore the heuristics as a design principle.

6.2.1 Measuring discovery learning skills

No firm conclusions could be drawn in this thesis about the discovery learning skills that students learned during working with the learning environment. The measurements used in the studies were unable to reveal differences between students in the different conditions in that respect.

Our experiences with the scientific reasoning test were disappointing in this respect. The scores were very high for all students and their answers gave no insight at the process that lead them to the answers. Based on the results of the test it was not possible to distinguish between the students who were exposed to the heuristics explicitly and the students who were exposed to the heuristics only implicitly.

This might be because there were no differences between the students in the two conditions, but the differences that were found on the process measures, could be an indication of differences between the conditions in use of the heuristics. The students in the implicit condition seemed to look at them more as procedures that had to be carried out whereas the students in the explicit conditions seemed to use them more as guidelines that are open to interpretation. The multiple-choice questions of the reasoning test were not able to detect such a difference.

The experiment design question seems to be a better candidate to reveal differences between the use of heuristics. In this kind of question the learners have to generate an experiment design, which provides more information about the design process than an answer to a multiple-choice question. Even though students' answers were sometimes very brief, these answers still yielded information about the capabilities to set up an experimental design. It showed that most students in the two conditions were able to set up unconfounded experiments that could be used to investigate plant growth in a greenhouse.

Structured interviews such as the Nature of Science Interview (Carey, Evans, Honda, Jay, & Unger, 1989; Carey & Smith, 1993) could be used as an alternative for open answer questions, and could also be used to investigate whether working with a learning environment leads to epistemology changes in learners. The advantage of a structured interview is that the interviewer can ask a learner for clarification at the moment that the answer is not yet clear, which is not possible with an open answer question. The disadvantage is of course that learners can not work with the learning environment at the same time since the interviews have to be administered after working with the learning environment.

Another method that could provide more information about the discovery skills, is verbal protocol analysis (Ericsson & Simon, 1984; van Someren, Barnard, & Sandberg, 1994). The advantage of verbal protocol analysis is that it provides a closer view on the interaction with the learning environment, and that it can therefore provide more insight of the use of the learning environment and whether this use is in concordance with the intentions of the designer. The disadvantage as with the structured interview is that learners have work with the learning environment one at a time, and that analysis of the protocols time consuming.

6.2.2 Using heuristics to support learners: extending the scope

In the second part of this thesis we used heuristics as a design principle. From a large set of heuristics that were gathered from literature on problem solving, scientific discovery and artificial intelligence, a condensed list of potentially useful heuristics for simulation-based discovery learning was constructed.

This list was used to guide the design of a simulation-based discovery learning environment that was tested in an experimental study with high school students. The results of the study showed that large effect sizes were obtained for knowledge gain on definitional and intuitive domain knowledge. These results strengthen the belief that heuristics can be used as a design principle.

In 1971, Polya argued that application of heuristics should be extended to education. The design of the learning environment as used in the second study is a step in that direction. For the future, there are two directions that might be of interest to explore.

The first direction is extending the ideas that were used in a way that it allows more learner control over the feedback. In the present situation feedback was given to all students when they were drawing graphs, and although it was individualized, and based on their experiments, it was the same for all learners in the sense that the feedback would always be given when it might be appropriate. The feedback was not telling the students that something was wrong, but mere asking for reflection from their side. The students were free to interpret the feedback as valuable, or superfluous. It could be that the latter is especially the case for “good” learners, and it might lead to a loss of motivation in these learners. For a future design it might be preferable to give learners responsibility over the feedback, allowing them to disable and enable feedback related to heuristics. A learner that thinks that certain feedback is superfluous, can then decide to turn it of, and will not be disturbed by the feedback anymore. One step further would be the ability to turn the heuristics of in assignments as well. Learners could then manage the scaffolding of the support, and turn it on or of whenever they feel they need it or can do without. There is of course a risk that learners will not remove the scaffolds themselves, but research should show whether or not they will, and what the consequences are for learning.

A second direction would be to look at other domains, and other simulations, to see whether these heuristics can be used to guide the design of simulation-based discovery learning environments in other domains. Other domains might require other heuristics, whereas some of the heuristics used here might be superfluous. This is most certainly the case. An example of an additional heuristic is a heuristic that can be used to distinguish noise from a main effect. In the simulations that were used in the research described in this thesis, noise does not exist, making this heuristic is superfluous for these simulations. In a simulation that does generate noisy data this no longer the case, and it would be helpful for the designer to use such a heuristic in the design process and for the learner to learn about it in the learning process.

Extending the set of heuristics that can be used to guide the design of learning environments, and extending the set of heuristics used in tools could be a great help for designing learning environments with a good balance between learner freedom and learner support.

7

References

ANDERSON, J. R., BOYLE, C. F., & YOST, G. (1985). The geometry tutor. In IJCAI-85: *Proceedings of the International Joint Conference on Artificial Intelligence* (pp. 1-7). Los Altos, CA: Morgan Kaufmann.

ANDERSON, J. R., FARRELL, R., & SAUERS, R. (1984). Learning to program in LISP. *Cognitive Science, 8*, 87-129.

BARNARD, Y. F., & SANDBERG, J. A. C. (1994). *The learner in the centre: Towards a methodology for open learner environments.* PhD Thesis. Amsterdam, The Netherlands: University of Amsterdam.

BROWN, J. S., & BURTON, R. R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science, 4*, 155-192.

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

BURTON, R. R. (1982). Diagnosing bugs in a simple procedural skill. In D. SLEEMAN & J. S. BROWN (Eds.), *Intelligent Tutoring Systems* (pp. 157-184). New York: Academic Press.

BUTLER, D. L., & WINNE, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65,* 245-281

CAREY, S., EVANS, R., HONDA, M., JAY, E., & UNGER, C. (1989). "An experiment is when you try it and see if it works": A study of grade 7 students' understanding of the construction of scientific knowledge. *International Journal of Science Education, 11,* 514-529.

CAREY, S., & SMITH, C. (1993). On understanding the nature of scientific knowledge. *Educational Psychologist, 28,* 235-251.

DEWEY, J. (1938). *Logic: the theory of inquiry.* New York: Holt and Co.

DUNBAR, K. (1993). Concept discovery in a scientific domain. *Cognitive Science, 17,* 397-434.

ERICSSON, K. A., & SIMON, H. A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.

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.

GIGERENZER G. (2000). Adaptive thinking: Rationality in the real world. New York: Oxford University Press.

GIGERENZER G., TODD P. M., & THE ABC GROUP (1999). *Simple Heuristics That Make Us Smart* Oxford: Oxford University Press.

GLASER, R., RAGHAVAN, K., & SCHAUBLE, L. (1988). Voltaville, a discovery environment to explore the laws of dc circuits. In G. GAUTHIER & C. FRASSON (Eds.), *Proceedings of Intelligent Tutoring Systems-88* (pp. 61-67). Montreal: University of Montreal.

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* (pp. 345-373). Berlin: Springer-Verlag.

GROOT, A. DE (1969). Methodology: Foundations of inference and research in the behavioral sciences. Belgium: Mouton & Co.

GOUDGE, T. A. (1950). *The thought of C. S. Peirce.* Toronto, On., Canada: University of Toronto Press.

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

HUTCHINS, E. (1995). How a cockpit remembers its speeds. *Cognitive Science, 19,* 265-288.

JONASSEN, D. H. (1991). Objectivism versus constructivism: Do we need a new philosophical paradigm? *Educational Technology Research & Development*, *39*, 5-14.

JONG, T. DE, & JOOLINGEN, W. R. VAN (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68,* 179-201.

JONG, T. DE, JOOLINGEN, W. R. VAN, SWAAK, J., VEERMANS, K., LIMBACH, R., KING, S., & GUREGHIAN, D. (1998). Self-directed learning in simulation-based discovery environments. *Journal of Computer Assisted Learning, 14* , 235-246.

JONG, T. DE, MARTIN, E., ZAMARRO J-M., ESQUEMBRE, F., SWAAK, J., & JOOLINGEN, W. R. VAN (1999). The integration of computer simulation and learning support; an example from the physics domain of collisions. *Journal of Research in Science Teaching, 5,* 597-615

JONG, T. DE, & NJOO, M. (1992). Learning and instruction with computer simulations: learning processes involved. IN E. DE CORTE, M. LINN, H. MANDL, & L. VERSCHAFFEL (Eds.), *Computer-based learning environments and problem solving*. Berlin: Springer-Verlag.

JOOLINGEN, W. R. VAN (1993). *Understanding and facilitating discovery learning in computer-based simulation environments.* PhD Thesis. Eindhoven, The Netherlands: Eindhoven University of Technology.

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

JOOLINGEN, W. R. VAN, & JONG, T. DE (1991). Supporting hypothesis generation by learners exploring an interactive computer simulation. *Instructional Science*, *20*, 389-404.

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-Verlag.

JOOLINGEN, W. R. VAN, & JONG, T. DE (1997). An extended dual search space model of learning with computer simulations. *Instructional Science, 25,* 307-346.

JOOLINGEN, W. R. VAN, & JONG, T. DE (in press). SimQuest: Authoring educational simulations. In T. Murray, S. Blessing & S. Ainsworth (Eds.), *Authoring Tools for Advanced Technology Educational Software: Toward cost-effective production of adaptive, interactive, and intelligent educational software*. Lawrence Erlbaum.

JOOLINGEN, W. R. VAN, KING, S., & JONG, T. DE (1997). The SimQuest authoring system for simulation-based discovery environments. In B. DU BOULAY & R. MIZOGUCHI (Eds.), *Knowledge and media in learning systems* (pp. 79-87). Amsterdam: IOS.

KAHNEMAN, D., & TVERSKY, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237-251.

KARASAVVIDIS, I. (1999). *Learning to solve correlational problems. A study of the social and material distribution of cognition*. PhD Thesis. Enschede, The Netherlands: University of Twente.

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

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

KLAYMAN, J. & HA, Y-W. (1987). Hypothesis testing in rule discovery: strategy, structure, and content. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 596-604.

KULKARNI, D., & SIMON, H. A. (1988). The processes of scientific discovery: The strategy of experimentation. *Cognitive Science, 12*, 139-175.

KULKARNI, D. & SIMON, H. A. (1990). Experimentation in machine learning. In J. Shrager & P. Langey (Eds.), *Computational models of scientific discovery and theory formation* (s. 255-273). San Mateo, CA: Morgan Kaufmann.

KUHN, D., BLACK, J., KESELMAN, A., & KAPLAN, D. (2000). The Development of cognitive skills to support inquiry learning. *Cognition and Instruction, 18*, 495-523.

KUHN, D., SCHAUBLE, L., & GARCIA-MILA, M. (1992). Cross-domain development of scientific reasoning. *Cognition and Instruction, 9*, 285- 327.

LAJOIE, S. P. (1993). Computer environments as cognitive tools for enhancing learning. In S. P. LAJOIE & S. J. DERRY (Eds.), *Computers as cognitive tools* (pp. 261-288). Hillsdale, NJ: Lawrence Erlbaum Associates.

LANGLEY, P. (1981). Data-Driven discovery of physical laws. *Cognitive Science, 5*, 31-54.

LANGLEY, P. (2000). The computational support of scientific discovery. *International Journal for Human-Computer Studies, 53*, 393-410.

LANGLEY, P., SIMON, H., & BRADSHAW, G. (1987). Heuristics for empirical discovery. In L. BOLC (Ed.), *Computational models of Learning*. Berlin: Springer-Verlag.

LANGLEY, P., SIMON, H. A., BRADSHAW, G. L., & ZYTKOW, J. M. (1987). *Scientific discovery, computational explorations of the creative process*. Cambridge: MIT Press.

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*, 335-360.

LENAT, D. B. (1979). On Automated Scientific Theory Formation: A Case Study Using the AM Program. In J. HAYES, D. MICHIE, & L. I. MIKULICH (Eds.), *Machine Intelligence 9* (pp. 251—283). Halstead, NY: Ellis Horwood.

LENAT, D. B. (1983). The role of heuristics in learning by discovery: Three case studies. In R. S. MICHALSKI, J. G. CARBONELL, & T. M. MITCHELL (Eds.), *Machine learning : An artificial intelligence Approach*. Palo Alto, CA: Tioga.

LEWIS, M., BISHAY, M., & MCARTHUR, D. (1993). The Macrostructure and microstructure of inquiry activities: Evidence from students using a microworld for mathematical discovery. *Proceedings of the World Conference on Artificial Intelligence and Education*. Edinburgh, Scotland.

LEWIS, E. L., STERN, J. L., & LINN, M. C. (1993). The effect of computer simulations on introductory thermodynamics understanding. *Educational Technology*, *33*, 45-58.

MYNATT, C. R., DOHERTY, M. E. & TWENEY, R. D. (1977). Confirmation bias in a simulated research environment: an experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, *29*, 85-95.

NEWELL, A., & SIMON, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

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

NORMAN, D. A. (1993). *Things that make us smart: Defending human attributes in the age of the machine.* New York: Addison-Wesley.

POLYA, G. (1971). Methodology or heuristics, strategy or tactics? *Archives de Philosophie, 34*, 623-629.

POLYA, G. (1945). *How to solve it*. Princeton, NJ: Princeton University Press.

POPPER, K. R. (1959). *The logic of scientific discovery*. London: Hutchinson.

QIN, Y., & SIMON, H. A. (1990). Laboratory replication of scientific discovery processes. *Cognitive Science*, *14*, 281-312.

QUINN, J., & ALESSI, S. (1994). The effects of simulation complexity and hypothesis generation strategy on learning. *Journal of Research on Computing in Education*, *27*, 75-91.

RAGNEMALM, E. L. (1996). Student diagnosis in practice; Bridging a gap. *User Modelling and User Adapted Interaction*, *5*, 93-116.

REIMANN, P. (1991). Detecting functional relations in a computerized discovery environment. *Learning and instruction, 1*, 45-65.

RIJN VAN, H. (2003). *Exploring the limited effect of inductive discovery learning: Computational and model-based analyses.* PhD Thesis in preparation. Amsterdam, The Netherlands: University of Amsterdam.

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

SALOMON, G., & PERKINS, D. N. (1989) Rocky roads to transfer: Rethinking mechanisms of a neglected phenomenon. *Educational Psychologist, 24*, 113-142.

SALOMON G. (1992). Effects with and of computers and the study of computer-based learning environments. In E. DE CORTE, M. C. LINN, H. MANDL, & L. VERSCHAFFEL (Eds.), *Computer-based learning environments and problem solving* (pp. 249-263). NY: Springer-Verlag.

SANDERS, I., BOUWMEESTER, M., & BLANKEN, M. VAN (2000). Heuristieken voor experimenteren in ontdekkend leeromgevingen. Unpublished report.

SCHAUBLE, L., GLASER, R., DUSCHL, R. A., SCHULZE, S., & JOHN, J. (1995). Students' understanding of the objectives and procedures of experimentation in the science classroom. *The Journal of the Learning Sciences*, *4*, 131-166.

SCHAUBLE, L., GLASER, R., RAGHAVAN, K., & REINER, M. (1991). Causal models and experimentation strategies in scientific reasoning. *The Journal of the Learning Sciences*, *1*, 201-239.

SCHAUBLE, L., KLOPFER, L., & RAGHAVAN, K. (1991). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching, 18*, 859-882.

SCHOENFELD, A. (1979). Can heuristics be taught? In J. Lochhead & J. Clement (Eds.), *Cognitive process instruction* (pp. 315-338). Philadelphia: Franklin Institute Press.

SCHOENFELD, A. (1985). Metacognitive and epistemological issues in mathematical understanding. In E. A. SILVER (Ed.), *Teaching and learning mathematical problem solving: Multiple research perspectives* (pp. 361-380). Hillsdale, NJ: Erlbaum.

SCHUNN, C. D., & ANDERSON, J. R. (1999). The generality/specificity of expertise in scientific reasoning. *Cognitive Science*, *23*, 337-370.

SELF, J. A. (1990). Bypassing the intractable problem of student modeling. In C. FRASSON & G. GAUTHIER (Eds.), *Intelligent Tutoring Systems: At the crossroads of articicial intelligence and education* (pp. 107-123). Norwood, NJ: Ablex.

SHEN, W. M. (1990). Functional transformations in AI discovery systems. *Artificial Intelligence, 41*, 257-272.

SHUTE, V. J. (1993). A comparison of learning environments: All that glitters In S. P. LAJOIE & S. J. DERRY (Eds.), *Computers as cognitive tools* (pp. 47-75). Hillsdale, NJ: Erlbaum.

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

SHUTE, V. J., & PSOTKA, J. (1996). Intelligent tutoring systems: Past, present and future. In D. JONASSEN (Ed.), *Handbook of research on educational communications and technology* (pp. 570-601). London: Prentice Hall International.

SIMON, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*, 129-138.

SIMON, H. A. (1973a). Does scientific discovery have a logic? *Philosophy of Science, 12*, 471-480.

SIMON, H. A. (1973b). The structure of ill-structured problems. *Artificial intelligence, 4*, 181-201.

SIMON, H. A. (1990) Invariants in human behavior. *Annual Review of Psychology, 41*, 1-19.

SIMON, H. A., & LEA, G. (1974). Problem solving and rule induction: a unified view. In L. W. GREGG (Ed.), *Knowledge and cognition* (pp. 105-128). Hillsdale, NJ: Erlbaum.

SLEEMAN, D., & BROWN, J. S. (Eds.). (1982). *Intelligent Tutoring Systems*. New York: Academic Press.

SOMEREN, M. W. VAN, BARNARD, Y. F., & SANDBERG, J. A. C. (1994). *The think aloud method: A practical guide to modelling cognitive processes.* London: Academic Press.

SWAAK, J. (1998). *What-If: Discovery simulations and assessment of intuitive knowledge.* Phd Thesis. Enschede, The Netherlands: University of Twente.

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

SWAAK, J., & JONG, T. DE (2001). Discovery simulations and the assessment of intuitive knowledge. *Journal of Computer Assisted Learning*,17, 284-294.

TSIRGI, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development, 51*, 1-10.

TVERSKY, A. & KAHNEMAN, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

VANLEHN, K. (1988). Student modeling. In M. C. POLSON & J. J. RICHARDSON (Eds.), *Intelligent tutoring systems* (pp. 55-78). Hillsdale, NJ: Earlbaum.

VASSILEVA, J. (1990). A classification and synthesis of student modelling techniques in intelligent computer-assisted instruction. In D. H. NORRIE & H.-W. SIX (Eds.), *Computer Assisted Learning: Vol 438. Lecture Notes in Computer Science* (pp. 202-213). Berlin: Springer-Verlag.

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

VEENMAN, M. V. J., ELSHOUT, J. J., & BUSATO, V. V. (1994). Metacognitive mediation in learning with computer-based simulations. *Computers in Human Behavior*, *10*, 93-106.

VEERMANS, K., & JOOLINGEN, W. R. VAN (1998). Using induction to generate feedback in simulation-based discovery learning environments. In B. P. GOETL, H. M., HALFF, C. L. REDFIELD, & V. J. SHUTE (Eds.), *Intelligent Tutoring Systems, 4th International Conference, San Antonio, TX USA* (pp. 196-205). Berlin: Springer-Verlag.

VEERMANS, K., JOOLINGEN, W. R. VAN, & JONG, T. DE (2000). Promoting self directed learning in simulation based discovery learning environments through intelligent support. *Interactive Learning Environments 8*, 229-255.

WASON, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*, 129-140.

WASON, P. C. (1983). Realism and rationality in the selection task. In J. S. B. T. EVANS (Ed.), *Thinking and reasoning: Psychological approaches* (pp. 44 - 75). London: Routledge & Kegan Paul.

WENGER, E., (1987). *Artificial Intelligence and Tutoring Systems*. Los Altos, CA: Morgan Kaufman.

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

WHITE, B. Y., & FREDERIKSEN, J. R. (1989). Causal models as intelligent learning environments for science and engineering education. *Applied Artificial Intelligence*, *3*, 83-106.

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

ZOHAR, A. (1995). Reasoning about interactions between variables. *Journal of Research in Science Teaching*. *32*, 1039-1063.

ZACHOS, P., HICK, L. T., DOANE, W. E. J., & SARGENT, S. (2000). .Setting theoretical and empirical foundations for assessing scientific inquiry and discovery in educational programs. *Journal of Research in Science Teaching*. *37*, 938-962.

Nederlandse samenvatting

Intelligente ondersteuning voor ontdekkend leren:

Het gebruik van een opportunistisch leerling-model en heuristieken voor het ondersteunen van ontdekkend leren met een computer simulatie

Iedereen die ooit het gevoel van opwinding heeft ervaren dat gepaard gaat met het ontdekken van iets “nieuws”, zal de potentie van ontdekkend leren kunnen inzien. “Iets nieuws” in deze zin betekent nieuw voor de persoon die de ontdekking doet. Dit kan duidelijk worden als de ontdekking wordt gedeeld met iemand anders, waarbij die ander opmerkt dat de ontdekking allang bekend was. Hoewel deze opmerking het gevoel van opwinding kan wegnemen, is het gevoel dat overblijft over het algemeen positief. De potentie van ontdekkend leren ligt enerzijds in de ontdekking zelf, d.i. het leren van nieuwe kennis. Anderzijds ligt deze in het proces dat tot de ontdekking geleid heeft, d.i. het opdoen van vaardigheden voor ontdekkend leren die de kans vergroten op het doen van ontdekkingen op een ander tijdstip en een andere plaats. Ontdekkend leren sluit aan bij veranderende opvattingen over leren. Waar leren vroeger gezien werd als het overdragen van kennis aan een passief ontvangende leerling, wordt leren tegenwoordig gezien als een constructief proces, waaraan de leerling actief deelneemt. Het onderzoek in dit proefschrift probeert een steentje bij te dragen aan de verwezenlijking van de potentie van ontdekkend leren.

Mede als gevolg van het toegenomen gebruik van computers in het onderwijs mag het gebruik van simulaties voor ontdekkend leren zich in een toegenomen belangstelling verheugen. Een computersimulatie is een programma dat een bepaald fenomeen simuleert op basis van een model. Een leerling kan tijdens het werken met een simulatie variabelen in het model veranderen en het effect van deze veranderingen bestuderen. De bedoeling van ontdekkend leren is dat de leerling door middel van de interactie met de simulatie inzicht verkrijgt in het achterliggende model van de simulatie.

Uitkomsten van onderzoek naar ontdekkend leren met computer-simulaties komen echter niet altijd overeen met de verwachtingen. Een van de redenen is dat het leerproces in ontdekkend leren voor veel leerlingen moeilijk blijkt. Leerlingen lopen tijdens het ontdekkend leerproces tegen allerlei problemen aan. Daarom worden leerlingen in computersimulaties over het algemeen niet aan hun lot overgelaten, maar wordt binnen de leeromgeving ondersteuning aangeboden om het ontdekkend leerproces beter te laten verlopen.

Een veel gebruikte aanpak om leerlingen te ondersteunen is het “verrijken” van de computersimulatie met “cognitive tools”, die de leerling ondersteunen tijdens het ontdekkend leerproces. Bijvoorbeeld door de leerling informatie aan te bieden, door ze op de taak toegesneden notitieblokken aan te bieden, of door het leerproces voor de leerlingen te structureren. Hoewel deze tools de leerling zeker van pas komen, zijn ze meestal niet ontworpen om het leerproces van een specifieke leerling te volgen en van feedback te voorzien. De algemene onderzoeksvraag in dit proefschrift is daarom: kunnen we een tool ontwikkelen die de leerlingen ondersteunt in het ontdekkend leerproces, waarbij deze ondersteuning gebaseerd is op de interactie van de leerling met de leeromgeving?

Om dit te kunnen doen moeten we iets meer weten over ontdekkend leren, over het leerproces, en over hoe ondersteuning kan worden aangeboden op basis van het gedrag van de leerling. In het eerste deel van Hoofdstuk 2 wordt dieper ingegaan op ontdekkend leren. Daarbij wordt een onderscheid gemaakt tussen de ontdekking, en het proces dat tot de ontdekking leidt. Het ontdekkend leerproces wordt vervolgens nader bekeken, en opgedeeld in een aantal deelprocessen. Deze deelprocessen zijn: oriëntatie, hypothese generatie, hypothese toetsing, conclusie, en regulatie.

Oriëntatie is het in kaart brengen van het probleem door te kijken naar de omgeving, en/of door voorkennis te activeren. In hypothese generatie gaat het om het formuleren van ideeën over de relatie tussen variabelen in het domein. Dit proces kan door de theorie gestuurd worden, door de data, of door een combinatie van deze twee. Een geformuleerde hypothese wordt idealiter ook getoetst. Hiervoor dienen experimenten opgezet te worden. Aan de hand van de hypothese en de experimenten kan er een conclusie aangaande de correctheid van hypothese getrokken worden. Regulatie kan onderverdeeld worden in planning, monitoren en evaluatie. De regulatie processen geven sturing aan de andere processen. Ze houden toezicht op het verloop van de processen en nemen beslissingen aangaande de overgang van het ene proces naar het andere.

Niet alle leerlingen ondervinden dezelfde problemen tijdens ontdekkend leren. Het is dan ook wenselijk dat de ondersteuning aan de individuele leerling aangepast kan worden. Om dat te kunnen verwezenlijken is gekeken naar Intelligente Onderwijs Systemen (ITS). Een ITS probeert op basis van de interactie van de leerling met het systeem een model van de kennis van de leerling te construeren en op basis van dit model de instructie aan de leerling aan te passen. Veel gebruikte toepassingen zijn het bepalen of een leerling voldoende weet om naar een volgend niveau te gaan, het geven van (on)gevraagd advies, en het genereren van problemen die net binnen de mogelijkheden van de student liggen. Voor het gebruik binnen ontdekkend leeromgevingen is alleen de tweede optie, het geven van advies, geschikt. Het advies dat aan de leerlingen wordt gegeven richt zich niet op de kennis van de leerlingen, maar op de leerprocessen die tot het verwerven van deze kennis zou moeten leidden. De bedoeling van het advies is het ondersteunen van de leerlingen tijdens deze processen. Hiervoor is een opportunistisch leerling-model gebruikt, het model probeert niet een accuraat model van de kennis van de leerling bij te houden, maar net genoeg om advies te kunnen geven.

Hierbij kunnen heuristieken een belangrijke rol spelen. Heuristieken zijn vuistregels voor het nemen van beslissingen in situaties waarin geen goede oplossing voor handen is, of waarin deze oplossing veel tijd en moeite kost. Heuristieken kunnen op twee manieren een rol spelen bij het ondersteunen van ontdekkend leren. Ze kunnen gebruikt voor het ondersteunen van de formele kant van ontdekkend leren: wanneer moet een hypothese verworpen worden? Welke voorspellingen kunnen op basis van een hypothese gedaan worden? Heuristieken kunnen de leerling een houvast te bieden op het moment dat de leerling deze formele kant nog niet goed beheerst. Daarnaast kunnen heuristieken ook gebruikt worden om de informele kant van ontdekkend leren te ondersteunen. Wat voor hypothese kan ik opstellen? Wat zijn goede experimenten om een hypothese te toetsen? Heuristieken kunnen de leerling hierin ondersteunen met richtlijnen voor zogenaamde "good practices".

De specifieke onderzoeksvraag is: kunnen we een tool ontwikkelen die de leerlingen ondersteunt in de hypothese toetsing en het trekken van conclusies, waarbij deze ondersteuning gebaseerd is op de interactie van de leerling met de leeromgeving?

Aan de hand van deze vraagstelling zijn binnen SIMQUEST, een programma voor het maken van simulatie leeromgevingen, twee versies van een tool ontworpen die in Hoofdstuk 3 beschreven worden. De eerste versie van de tool richtte zich voornamelijk op de formele kant van het toetsen van hypotheses, en het trekken van conclusies. De tweede versie van de tool richtte zich ook op de informele kant van het toetsen van hypotheses. De eerste versie van de tool werd gebruikt in een studie die beschreven wordt in Hoofdstuk 4, de tweede in een studie die beschreven wordt in Hoofdstuk 5.

In de eerste versie van de tool werd, gebruik makend van principes van inductie en deductie, op het moment dat een leerling een antwoord gaf voor een hypothese uit een van de opdrachten bekeken of de conclusie die een leerling over de hypothese trok op basis van de experimenten gerechtvaardigd kon worden. Het resultaat van deze analyse werd vervolgens gebruikt om intelligente feedback te geven aan de leerling. Om op formele gronden ook feedback te kunnen geven op experimenteer gedrag werd gebruik gemaakt van semi-quantitatieve hypotheses. In deze hypotheses (bv. "Als de massa 2 keer zo groot wordt dan wordt de kinetische energie ook twee keer zo groot") kan het conditie gedeelte gebruikt worden om te kijken of de experimenten aan deze conditie voldoen, en kan aan de hand daarvan feedback gegenereerd worden voor leerlingen die hier problemen mee lijken te hebben.

Deze eerste versie van de tool werd getest in een studie die wordt beschreven in Hoofdstuk 4. In deze studie werden twee leeromgevingen gebruikt. Een die de eerste versie van de tool gebruikte om intelligente feedback te geven aan de leerlingen en een waarin de leerlingen vooraf gespecificeerde goed / fout feedback te zien kregen na het beantwoorden van een vraag. Het onderwerp dat in de leeromgevingen aan de orde kwam had betrekking op botsingen en bestond uit vier onderdelen. De leerlingen werden vooraf en achteraf getoetst op definitionele kennis (formules en wetten) en intuïtieve kennis (voorspellen wat er in een bepaalde situatie gebeurt). Beide toetsen waren meerkeuze toetsen. Daarnaast werd achteraf nog een toets afgenomen waarin de leerlingen gevraagd werd om naast een voorspelling ook uitleg te geven. Verwacht werd dat de experimentele conditie leerlingen meer tot experimenteren en analyseren van de experimenten zou aanzetten, en dat ze als gevolg daarvan meer intuïtieve kennis zouden opdoen en beter in staat zouden zijn om deze kennis in een uitleg te verwoorden

De verwachtte verschillen ten aanzien van de toets resultaten werden niet gevonden. In beide condities gingen de leerlingen er op vooruit, maar ze bleken niet van elkaar te onderscheiden te zijn op basis van de natoetsen. Er bleken wel verschillen te zijn in de samenhang tussen toets resultaten. De samenhang tussen de natoets resultaten bleek sterker te zijn in de controle conditie dan in de experimentele conditie. Ook de samenhang tussen voortoets en natoets resultaten bleek te verschillen. In de controle conditie bleek definitionele voorkennis niet alleen met definitionele nakennis samen te hangen, maar ook met de andere natoets resultaten. De definitionele voorkennis lijkt in belangrijke mate te bepalen hoe de leerlingen scoorden op de natoetsen In de experimentele conditie is het precies andersom hier hangt de intuïtieve voorkennis samen met de natoets resultaten, terwijl de definitionele voorkennis geen sterke samenhang met de andere natoets resultaten vertoont. Hier lijkt intuïtieve voorkennis te bepalen hoe de leerlingen scoorden op de natoetsen

Uit de analyse van interactie van de leerlingen met de leeromgeving bleek dat leerlingen in de controle conditie meer opdrachten deden, en per opdracht minder tijd besteedden. De leerlingen in de experimentele conditie deden daarentegen meer experimenten. Er was ook een opvallend verschil tussen de twee condities in de samenhang tussen de interactie met de leeromgeving en de resultaten op de natoetsen. In de controle conditie bleek het gebruik van opdrachten samen te hangen met de resultaten op de natoetsen, en het experimenteren niet. In de experimentele conditie was dit eerder andersom.

Het leek er in deze studie op dat, hoewel de leerlingen evenveel leerden, het leren in de experimentele conditie meer samenhing met ontdekkend leeractiviteiten, terwijl het in de controle conditie meer samenhing met meer traditioneel leren van opdrachten en antwoorden Daarnaast leek er in de experimentele conditie een verband te zijn tussen de intuïtieve voorkennis en de natoets resultaten.

Het zou kunnen zijn dat de intuïtieve kennis toets niet alleen domein specifieke intuïtieve kennis meet, maar ook kennis die gerelateerd is aan ontdekkend leren, en dat er in de leeromgeving een soort drempel was om de intelligente feedback goed te kunnen gebruiken.

Dit laatste leidde ertoe nog eens goed naar de tool en de feedback te kijken. Hierbij kwamen twee problemen naar voren. Het eerste probleem was dat de leerlingen alleen feedback kregen die gebaseerd was op hun eigen experimenten. Ze wisten daarom nooit zeker of een hypothese al dan niet waar was. Zonder een externe validatie is het moeilijk om kennis te vergaren en om het eigen leerproces te evalueren. Het tweede probleem was dat de tool misschien wel te formeel was. De semi-quantitatieve hypotheses die gebruikt werden om op formele gronden iets over het experimenteren te kunnen geven bleken niet zo eenvoudig op het achterliggende model van de simulatie te mappen. Het vereist een extra transformatie stap van de leerling en dit maakte het waarschijnlijk moeilijker om domeinkennis te verwerven.

Op basis van deze ideeën werd besloten om een nieuwe versie van de tool te ontwerpen, ditmaal een tool die ook meer informeel naar het hypothese toetsen keek, en daarnaast de mogelijkheid open liet om goed / fout feedback met betrekking tot hypotheses te geven. Een al bestaande tool waarin leerlingen experimenten konden opslaan, werd uitgebreid met de mogelijkheid om de experimenten in een grafiek weer te geven. Het moment waarop de leerling een grafiek tekende werd aangegrepen om advies te geven. Voor het geven van meer informele feedback werd gebruik gemaakt van heuristieken. Er werd gekeken of de experimenten van de leerlingen volgens deze heuristieken uitgevoerd waren, en als dat niet zo was werd daar in de feedback melding van gemaakt, waarbij het aan de leerling overgelaten werd om daar al dan niet iets mee te doen.

Deze tweede versie van de tool werd gebruikt in een conditie van de tweede studie die beschreven wordt in Hoofdstuk 5. Ook in deze studie was het onderwerp botsingen in de natuurkunde. In dit onderzoek werden twee leeromgevingen met elkaar vergeleken waarbij in beide gevallen heuristieken gebruikt werden om de leeromgeving vorm te geven. De heuristieken kwamen terug in de keuze van de opdrachten, de inhoud van de opdrachten, de feedback op de opdrachten en de zojuist beschreven opslag tool. In de impliciete conditie bleven de heuristieken impliciet voor de leerling, en werden alleen de beslissingen die op basis van de heuristieken genomen waren aan de leerling getoond. De tweede versie van de tool werd in deze conditie ontdaan van de feedback op basis van de heuristieken. In de expliciete conditie werden beslissingen op basis van de heuristieken vooraf gegaan door de naam van de heuristieken die eraan ten grondslag lag en een korte beschrijving van de rationale van de heuristiek.

Ook in deze studie werd vooraf en achteraf weer definitionele en intuïtieve kennis gemeten. Daarnaast werd er ook een toets voor wetenschappelijk redeneren en werd leerlingen gevraagd een experiment te ontwerpen in een gegeven situatie om te kijken of het geven van expliciete aandacht aan de heuristieken tot betere resultaten zou leiden. Ook in dit onderzoek gingen de leerlingen er in beide condities op vooruit, en bleken ze na afloop niet van elkaar te onderscheiden te zijn op basis van de natoetsen. De resultaten van de toets voor wetenschappelijk redeneren liet ook geen verschil zien tussen de twee condities, de leerlingen in beide condities scoorden hierop zo hoog dat er sprake was van een plafond effect. Ook de door leerlingen ontworpen experimenten lieten geen verschillen zien. De experimenten verschilden echter sterk in de mate van detail waarin ze uitgewerkt waren.

De samenhang tussen de voortoets en de natoets was sterker in de controle conditie. Regressie analyse wees uit dat in de impliciete conditie zowel de definitionele als de intuïtieve natoets goed voorspeld kon worden op basis van de beide voortoets resultaten. Voor de expliciete conditie was de voorspelling minder goed. Opvallend was dat in de impliciete conditie de intuïtieve voorkennis ook bijdroeg aan de voorspelling van de definitionele nakennis en dat in de expliciete conditie de definitionele voorkennis niets bijdroeg aan de voorspelling van de definitionele nakennis. Bij de analyse van de interactie van de leerlingen bleek dat leerlingen in de expliciete conditie na de eerste kennismaking met de heuristieken meer eigen initiatief ontplooiden dan leerlingen in de impliciete conditie, maar dat dit verschil later niet meer duidelijk terug te vinden waren. Wel was er verschil in de samenhang tussen de interactie met leeromgeving, en de natoets resultaten. In de expliciete conditie hing de interactie positief samen met het resultaat op de definitionele natoets, in de impliciete conditie negatief. Een verklaring voor dit resultaat zou kunnen zijn dat de leerlingen in de impliciete conditie zich geen raad wisten met de impliciete heuristieken. Ze doen wat er gesuggereerd wordt, maar begrijpen het achterliggende idee niet echt.

In zijn algemeenheid kan na deze studie geconcludeerd worden dat beide leeromgevingen het verwerven van domein kennis goed ondersteunen. Het gebruik van heuristieken lijkt een veelbelovende aanpak. Met betrekking tot de heuristieken zelf kunnen geen ferme uitspraken gedaan worden. Er zijn aanwijzingen dat ze tot meer zelf-regulatie leiden, wat zou betekenen dat de leerlingen zich de heuristieken eigen gemaakt hebben. Of dit ook echt het geval is, kan pas blijken in een studie waarin leerlingen na afloop een ontdekkend leertaak krijgen zonder ondersteuning.

Terugkijkend op de twee studies zouden we kijkend naar het experimenteer gedrag van de leerlingen kunnen zeggen dat de leerlingen in de experimentele conditie van de eerste studie het meest exploratief zijn, de leerlingen in de controle conditie van het eerste experiment het minst, en dat de leerlingen in de tweede studie er tussenin zitten. Hoewel de leerlingen in de tweede studie zich net als de leerlingen in de controle conditie van de eerste studie meer hadden kunnen gaan richten op de vragen en de antwoorden lijken zij minder geneigd dit te doen. Het is niet gezegd dat de leerlingen in de experimentele conditie van de eerste studie de meeste vaardigheden hebben opgedaan. Het feit dat de leerlingen in de tweede studie ook feedback kregen over de correctheid van hypotheses en deze informatie konden gebruiken bij de evaluatie van het eigen leerproces kan voor deze leerlingen erg behulpzaam geweest zijn bij het verwerven van vaardigheden voor ontdekkend leren en kan er toe geleid hebben dat deze leerlingen meer vaardigheden opgedaan hebben.

Er waren geen verschillen tussen leerlingen in de condities binnen de twee studies met betrekking tot resultaten op de definitionele en intuïtieve natoets. Wel was er verschil te zien tussen de twee studies. In studie twee waren de scores op de definitionele en intuïtieve natoets hoger en was het verschil tussen voor en natoets groter. Het lijkt er op dat de leeromgevingen in de tweede studie de leerlingen beter in staat stelden om domein kennis te verwerven.

Er waren ook verschillen tussen de condities in de samenhang tussen de voortoets resultaten en de natoets resultaten. In de experimentele conditie van de eerste studie bleek vooral de intuïtieve kennis uit te maken voor de scores op de natoetsen. In de expliciete conditie van de tweede studie bleken de voortoetsen geen goede voorspeller te zijn voor de natoetsen. Deze condities lijken misschien de zwakkere leerlingen, maar in ieder geval een ander soort leerlingen te ondersteunen dan traditioneel onderwijs. De leeromgevingen en de tools lijken dan ook een goede aanvulling op het bestaande onderwijs te kunnen vormen.

Het meten van de ontdekkend leervaardigheden bleek niet gemakkelijk. De antwoorden op de toets voor wetenschappelijk redeneren geven weinig zicht op het achterliggende proces. Leerlingen vragen om een experiment te ontwerpen lijkt meer informatie over deze vaardigheden te kunnen geven. Andere alternatieven die mogelijk tot meer inzicht zouden kunnen leiden zijn gestructureerde interviews, of het afnemen van hardop denk protocollen.

Het gebruik maken van heuristieken om een tool en/of leeromgeving vorm te geven lijkt veelbelovend. Voor de toekomst zijn er twee richtingen waarin deze aanpak verder uitgewerkt zou kunnen worden. Ten eerste zouden leerlingen meer controle over de heuristieken kunnen krijgen. Een leerling zou de mogelijkheid moeten krijgen om heuristieken als het ware uit te zetten als deze voldoende bekend zijn. Ten tweede zou ook naar andere domeinen en simulaties gekeken moeten worden om te zien of deze aanpak daar ook werkt, en of voor deze andere domeinen andere heuristieken nodig zijn. Dit samen zou kunnen helpen bij het ontwikkelen van simulatie leeromgevingen voor ontdekkend leren die een goede balans vinden tussen vrijheid voor de leerling en ondersteuning van de leerling.