Supporting hypothesis information by learners exploring an interactive computer simulation

Please check the document version of this publication:

Supporting hypothesis generation by learners exploring an interactive computer simulation[*]

WOUTER R. VAN JOOLINGEN[1] & TON DE JONG[2]
[1]*Eindhoven University of Technology, Department of Philosophy and Social Sciences, P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands*
[2]*University of Twente, Department of Education, Division of Instructional Technology, P.O. Box 217, NL-7500 AE Enschede, The Netherlands*

Abstract. Computer simulations provide environments enabling exploratory learning. Research has shown that these types of learning environments are promising applications of computer assisted learning but also that they introduce complex learning settings, involving a large number of learning processes. This article reports on an instrument for supporting one of these learning processes: stating hypotheses.

The resulting instrument, an *hypothesis scratchpad*, was designed on the basis of a conceptual representation of the simulation model and tested in an experimental study. In this study three versions of the scratchpad, varying in structure, were compared. It was found that support offered for identifying variables, in the form of a selection list, is relatively successful: students who used this list were better in differentiating different types of variables. For identifying relations, a selection list of relations offered to the students proved unhelpful in finding accurate relations: students using this list stated their hypotheses mainly at a very global level.

Introduction: supporting exploratory learning with simulations

Learning with computer simulations is a promising application of instructional technology. A main reason is that computer simulations enable the creation of relatively cheap, safe and well-accessible exploratory learning environments (Alessi and Trollip, 1987; De Jong, 1991; Reigeluth and Schwartz, 1989). These types of environments provide complex learning settings involving a large number of specific learning processes. In a recent study Njoo and De Jong (1991; see also De Jong and Njoo, 1990) made observations of students working with a computer simulation. They distinguished the following main categories of learning processes (apart from regulative processes, concerned with planning, and processes involved with operation of the simulation system):

- analysis
- hypothesis generation
- hypothesis testing
- evaluation

The process of analysis is concerned with identifying variables and global model properties. In this phase the first, often not yet well-articulated, ideas about the underlying simulation model may arise, leading to the generation of hypotheses about the simulation. Hypotheses must be tested to become a part of the learner's (mental) model of the simulation. This testing includes the design of an *experiment* which will be performed with the simulation, *predicting* the outcomes of the experiment, on the basis of the hypothesis, *performing* the experiment and *interpreting* the results (Njoo and de Jong, 1992). This may lead to rejection of or support for the hypothesis and may give rise to the generation of a new hypothesis or a reformulation of an old one. Then the process may start over again. Also the learner can choose to investigate another part of the simulation model and state an hypothesis about that part. This process can continue until all parts of the simulation have been investigated and the learner has discovered the complete model. Research has shown that *generation of hypotheses* and *designing experiments* to test these are both important and problematic parts of discovery learning (Gorman and Gorman, 1984; Gorman, Stafford and Gorman, 1987; Klahr and Dunbar 1988; Mynatt, Doherty and Tweney, 1977, 1978; Njoo and de Jong, 1991; Wason, 1960).

Hypothesis generation and testing

Klahr and Dunbar (1988; Dunbar and Klahr, 1989; Shrager and Klahr, 1986) studied the formation of hypotheses and the design of experiments to test these, with students discovering the operation of a simple device. Their research results in a theory of *scientific discovery as dual search* (SDDS), partially based on general theories of problem solving (Newell and Simon, 1972; Greeno and Simon, 1988). They propose it as "a general model of scientific reasoning, that can be applied to any context in which hypotheses are proposed and data is collected" (p. 32). SDDS describes the scientific discovery process as a search process in an *hypothesis space*, containing all possible hypotheses about the system under study, and in an *experiment space*, consisting of all experiments that can be carried out with the system.

Klahr and Dunbar's findings indicate that there are two types of strategies for searching these spaces (see also Greeno and Simon, 1988, p. 640). The first is a *bottom up* strategy (used by what they call *experimenters*) consisting of a first phase in which an hypothesis is tested, followed by a phase where the subject searches the experiment space without explicitly stating hypotheses. The main characteristic of experimenters is that they perform experiments which rule out all other possible hypotheses before they actually state the correct hypothesis. In other words, experimenters cannot reach certain parts of the hypothesis space without prior experimental validation.

In the second, *top-down* strategy (used by so-called *theorists*) experiments are never performed without the prior statement of an hypothesis. Typically, a theorist states an hypothesis before carrying out an experiment and switches to a new

hypothesis only after sufficient contradicting evidence has been found. The new hypothesis stated will mostly not differ radically from the old one. Typically, only one relevant aspect will have been changed. Theorists do not need to conduct a critical experiment before the correct hypothesis is stated. Generally, theorists require fewer experiments than do experimenters to reach the correct conclusion.

Similar distinctions are reported by Shute, Glaser, and Raghavan (1989) (see also Shute and Glaser, 1986, 1990; Shute, Glaser and Resnick, 1986). They have studied the use of a system for learning laws of economics, *Smithtown*. The system is a freely explorable computer simulation of a model of an economy. Students are invited to explore the simulation to discover the laws that determine the underlying model. Shute and others refer to theorists' behaviour as *hypothesis driven* and to experimenter behaviour as *data driven*. Moreover, they conclude that hypothesis driven subjects are more successful than data-driven subjects.

A study by Wason (1960) and two related studies by Gorman *et al.* (Gorman and Gorman 1984; Gorman, Stafford and Gorman, 1987) showed that students, once they have formed an hypothesis (in their case in a simple domain: discovering regularity in sequences of three numbers), tend to seek confirming evidence for this hypothesis, i.e. they design experiments which are aimed at obtaining data in support of the hypothesis and not at obtaining data, able to discriminate between hypotheses. This may result in long series of fruitless experiments. When the hypothesis space is reduced to a small set of conflicting hypotheses, by offering a small set of rules to the students from which they could choose, the search for the right rule proved to be far more successful.

Recent research by Njoo and De Jong (1992) showed that the formation of hypotheses about a simulation is one of the most problematic parts of the exploratory learning setting. Their research shows that students spontaneously generated very few hypotheses and that they confuse hypotheses and predictions.

From the research described above we may conclude that hypothesis generation is a problematic issue in discovery learning contexts, notably simulations, leading to a search for ways to support this study process. Based on SDDS (Klahr and Dunbar 1988, 1989) we can offer this support by elucidating the structure of hypothesis space to the learner. In the present study this is done by providing learners with a mock-up *hypothesis scratchpad*, a software instrument, or *learner instrument*, on which the student can note down hypotheses.

Hypothesis scratchpads for simulations

Hypothesis scratchpads offer learners a structured overview of hypothesis space. This implies that for designing these scratchpads we first need to investigate the specific properties of hypothesis space for simulations.

In Van Joolingen and De Jong (1991b), it is argued that hypotheses about simulations are formed in principle about a *conceptual model* describing the properties of the simulation model, that is: *An hypothesis about a simulation model is a statement that a certain generic relation holds between two or more conceptual*

variables (see also Reimann, 1989). The term *conceptual variable* stands for a generalisation of the variables present in the computer simulation; the term *generic relation* is a generalisation of the traditional relation concept, allowing for fuzzy and incomplete descriptions of a certain relationship (Van Joolingen and De Jong, 1991a; 1992).

The definition above determines the dimensions of the hypothesis space, being a superposition of the space of all possible combinations of conceptual variables and the space of all possible relations between these variables. This implies that the process of forming hypotheses consists of the following subprocesses:

– *identifying* variables

– *selecting* variables

– defining the *(generic) relation* that is hypothesized to hold between the selected variables.

In the literature, several learner instruments supporting one or more of these subprocesses of hypothesis generation are described. *Smithtown* (Shute and Glaser, 1990) contains a so-called *Hypothesis Menu,* that supports students in stating hypotheses about the model. The hypothesis menu offers a structured framework for entering hypotheses. The most important elements are the *Objects* and *Verbs.* The objects correspond to *variables* in the simulation and verbs express the behaviour of the objects under conditions, expressed in the same hypothesis. The other two elements in the hypothesis menu are *connectors* and the *direct object menu* which are used respectively to produce well formed sentences and to specify the hypothesis more precisely. A sample hypothesis entered in the hypothesis menu could be: "As price increases then quantity demanded decreases" (Shute and Glaser, 1990, p. 292).

Michael, Haque, Rovick and Evens (1989) use an hypothesis menu in a learning environment for pathophysiology problems. The goal that the learner is to achieve is to locate a malfunction in a patient on the basis of given symptoms. The hypotheses that can be entered take the form of an area of the model where the defect may be located. The learner may select his/her hypotheses from a menu of ready-made hypotheses. The system offers the learner a set of nested menus to select from. After the hypothesis has been chosen the learner can collect data to support his/her hypothesis.

In general we can define an *hypothesis scratchpad* as a learner instrument that can support some or all of the subprocesses of forming hypotheses mentioned above by offering the elements needed for hypotheses (variables and relations) and by adding some structure to these elements.

The scratchpad may not be imperative in the sense that it actually forces students to generate a particular hypothesis and/or to carry out one specific experiment. Furthermore, it may not give away too much information, such as providing ready made hypotheses, since the actual generation of hypotheses is one of the goals of the self-discovery process.

In the present study we investigated the effect of offering an hypothesis scratchpad on the various subprocesses of hypothesis generation. Learners were provided with one of three versions (differing in the amount of information and structure) of an hypothesis scratchpad.

Method

Domain

The domain involved in our study was error analysis in chemistry. As part of the experiment, students worked (individually) with an interactive computer simulation called 4SEE (Statistics Simulation System as a Supportive Exploratory Environment). 4SEE is a simulation of a titration experiment, emphasizing the various types of measuring errors. These errors are simulated and the learner can use the program to investigate which factors contribute to the total measuring error. The experimental computer lab was integrated in the normal curriculum. Before using the simulation, students had received an introduction to error analysis.

Subjects

Subjects were 31 first-year students of chemistry at a University of Technology. There were three experimental groups of 10, 10 and 11 students. These groups were the regular lab groups students were used to working in. Assignment of students to groups was done randomly, as was assignment of groups to experimental conditions.

Conditions

In the experiment, subjects were offered hypothesis scratchpads, which were mocked up as a set of (paper) forms. Three different hypothesis scratchpads were used, one in each experimental group. These versions will be referred to as the structured, partially structured and unstructured scratchpad.

The *structured scratchpad* consisted of three tables (see Figure 1). Each table contained elements for the construction of hypotheses. One table contained variables, one contained conditions and one contained relations. A relation could be constructed by selecting two or more variables, a relation and (optionally) a condition to limit the scope of the relation.

The variables presented in the variable table were derived from a conceptual model of error analysis that was constructed, representing conceptual knowledge at a higher level of abstraction than the simulation model (see the Introduction). The

relations on the scratchpad were all relations present in the conceptual model, completed with a number of relations that were plausible alternatives for these relations. The list of conditions was constructed in a similar way.

The *partially structured scratchpad* shared the variable table with the structured scratchpad but did not contain a relation construction part. The students had to describe the relation in natural language, using the variables as listed in the table.

The *unstructured scratchpad* did not contain any information, neither on variables nor on relations. The forms contained only one area in which the student could express his/her hypotheses, in natural language.

Our prediction was that both the relation selection table and the variable selection table would enlarge the hypothesis space the learner can use and enable him/her to explore this space, resulting in less experiments done without prior statement of hypotheses (theorist search strategies). Also, it was expected that learners using a structured scratchpad would have better formulated and better articulated hypotheses, resulting in more consistent exploratory behaviour.

Table 2: Select a condition here

Table 1: Select variables here

If ..... becomes greater than .....

Table 3: Select a relation here

If ..... increases then ...... also increases

D
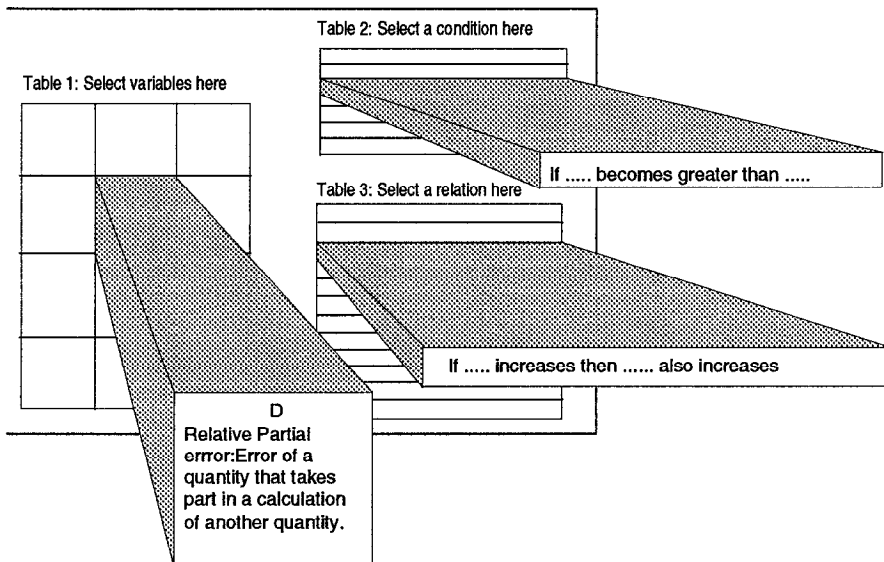Relative Partial error:Error of a quantity that takes part in a calculation of another quantity.

Figure 1. A structured hypothesis scratchpad as used in the experiment

Procedure

After a short introduction to the experiment, the computer program and the use of the scratchpad, the students received a written instruction and worked with the computer program for two hours. The written instruction was the same for all students except for a small part concerning the use of the hypothesis forms.

After this session the students received a post test which consisted of one open question: "Describe on this sheet what you have learned from the computer simulation. Mention the central elements in the simulation: the most important variables and relations. For example, indicate which factors contribute to the total random error and how. It is also important to mention when a certain factor does not contribute when you would expect it to."

In addition to the hypothesis scratchpad, *unstructured experiment/prediction* scratchpads for noting down experimental plans and predictions of experimental outcomes were offered to the students. The purpose of these scratchpads was to obtain information about the students' experimental design, in order to support the interpretation of the students' hypotheses and to assess the consistency of the experimental behaviour of the subjects.

Each time, after completion of an experiment, students were prompted (by the simulation program) to fill in an experiment form. They were also instructed to fill in an hypothesis form at the same time if they wanted to express an idea. Table 1 summarizes the experimental set-up.

Table 1. Overview of the experimental design

Group	Relation support	Variable support	Experiment scratchpads	Post-test
I (N = 10)	•	•	•	•
II (N = 10)		•	•	•
III (N = 11)			•	•

Data

The data collected consisted of the forms (hypothesis and experiment) filled in by the students, the log-files of the interaction with the computer simulation and the results of the post test.

The results of the post test were assessed using the conceptual model that was created of the domain as the criterion. The number of relations that was treated in the essay was counted and each relation was matched against the conceptual model. This was done by two independent domain experts (teachers from the Chemistry Department).

Results

To assess the influence of the presence of an hypothesis scratchpad on exploratory study behaviour, we examined the activity level of the students, the way in which hypotheses are formulated and the agreement between hypotheses and experiments, the mapping of hypotheses entered by the students to the conceptual model and, finally, the consistency of the students' exploratory behaviour.

Table 2. Activity level of the students, as indicated by the average number of experiments performed, experiment forms filled in, and hypotheses stated.

| | Group | | | |
	I	II	II	
Number of hypotheses	3.6	5.5	6.1	$(F(2,28) = 2.1, \text{n.s.})$
Number of experiment forms	5.3	8.9	9.7	$(F(2,28)= 7.7, p<0.005)$
Number of experiments	6.6	10.6	11.6	$(F(2,28) = 8.8, p<0.005)$

Activity level. The activity level of the students was measured by the number of hypotheses they stated, the number of experiment forms they filled in, and the number of experiments they performed. Table 2 summarizes this data, together with the results of an analysis of variance applied to these three variables. From Table 2 it is clear that students using structured scratchpads (group I) showed a lower activity level.

General functioning of the hypothesis forms. For a general indication of the functioning of the hypothesis forms two dependent variables are important:

(1)   The (relative) number of well-formed hypotheses (i.e., hypotheses with a correct syntax).

(2)   The agreement between hypotheses and experiments (i.e., are the experiments designed in such a way that they can test the hypotheses?).

For *syntax analysis* we looked separately at variables and relations. An hypothesis was judged to have a correct syntax if both the variables about which the hypothesis was stated and a relation between them were present. For each hypothesis stated, and for each of both aspects separately, we determined whether it was correct, incorrect, or interpretable, the latter meaning that the variable or relation on the scratchpad was incorrectly stated on the scratchpad but that the variable or relation the student intended to refer to could be inferred.

Figure 2 displays the results of the syntax analysis, expressed in relative scores. This figure shows that the students using the structured scratchpads (Group I) use a better syntax for their hypotheses than the other students (for variables $\chi^2=53.9$, d.f.=4, $p<<0.001$; for relations $\chi^2=17.6$, d.f.=4, $p<0.005$).
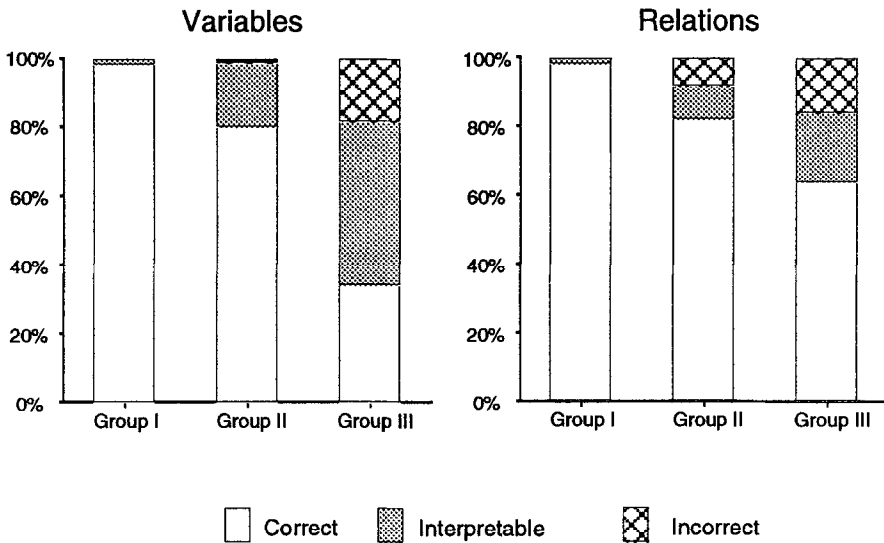
Figure 2. The syntax of the hypotheses stated on the scratchpads, for variables and relations entered separately

As such this is not unexpected, since the structure of the scratchpads more or less forces them to use a correct syntax, but from the fact that for the other groups, especially the unstructured one (Group III), the percentage of syntactically correct hypotheses is substantially lower than for the structured groups, one may conclude that the support offered by the scratchpads is successful here. One noticeable aspect of Figure 2 is that one would expect that Group I and II would have the same score on the choice of variables, since they used the same variable selection table. However, it appeared that in Group II some students did not always use this variable table and invented their own variables instead. This explains the slightly lower score on variable selection syntax in Figure 2.

The agreement between hypotheses and experiments did not differ significantly between groups: for all groups about 67% of all experiments was appropriate for testing the last-stated hypothesis.

Another general evaluation can be made on the basis of the post test that was taken from the students. In Figure 3 the results of this test are depicted. The figure shows that the total number of statements per student is slightly, but not significantly, lower in experimental Group I but that there are no differences in the ratios between the numbers of correct and incorrect statements.
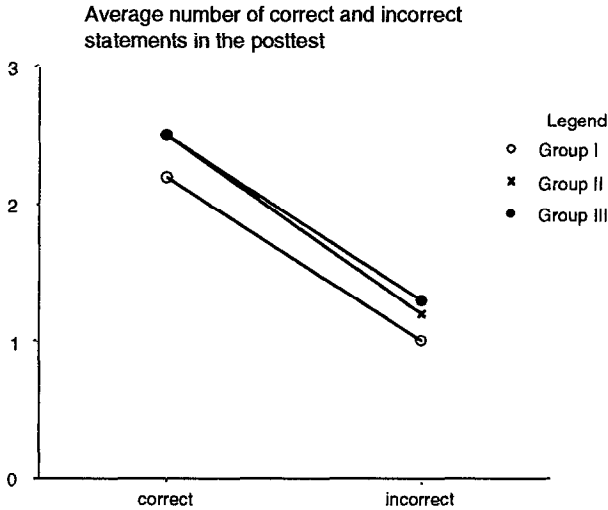
Average number of correct and incorrect
statements in the posttest

Figure 3. Results of the post test, showing the average number of correct and incorrect hypotheses per student for the three experimental groups.

Assessment of hypotheses related to a conceptual domain model. The hypotheses entered on the scratchpads were matched with a conceptual model of the simulated domain. This model was defined in terms of (conceptual) *variables*, representing conceptual units, and *relations*, describing the interdependencies between the variables. It is important to notice that conceptual variables can occur in hierarchies: child variables represent more specific conceptual units and inherit the characteristics of their parents. For more information on the conceptual model see Van Joolingen and De Jong (1992).

As the variables and relations in the conceptual model span the hypothesis space, an analysis of the variables and relations used by the students provides insight in the students' search through hypothesis space.

Variables selected. A measure for hypothesis space students effectively use while exploring a simulation is the number of different variables and relations they use to construct their hypotheses.

Figure 4 shows the use of the variables representing different kinds of measuring error as output variables in relations. The nodes in the networks depict variables, representing different kinds of measuring error. The lines depict hierarchical relations between variables. The top of the graph represents the most general type of error. Further down the graph, various kinds of error (relative, absolute, total error, partial error) are differentiated.

From Figure 4, it is clear that the variable support offered on the hypothesis scratchpads triggers the students to use more different variables in stating their hypotheses. The students using the unstructured scratchpads use, in effect, a smaller hypothesis space to state ideas.
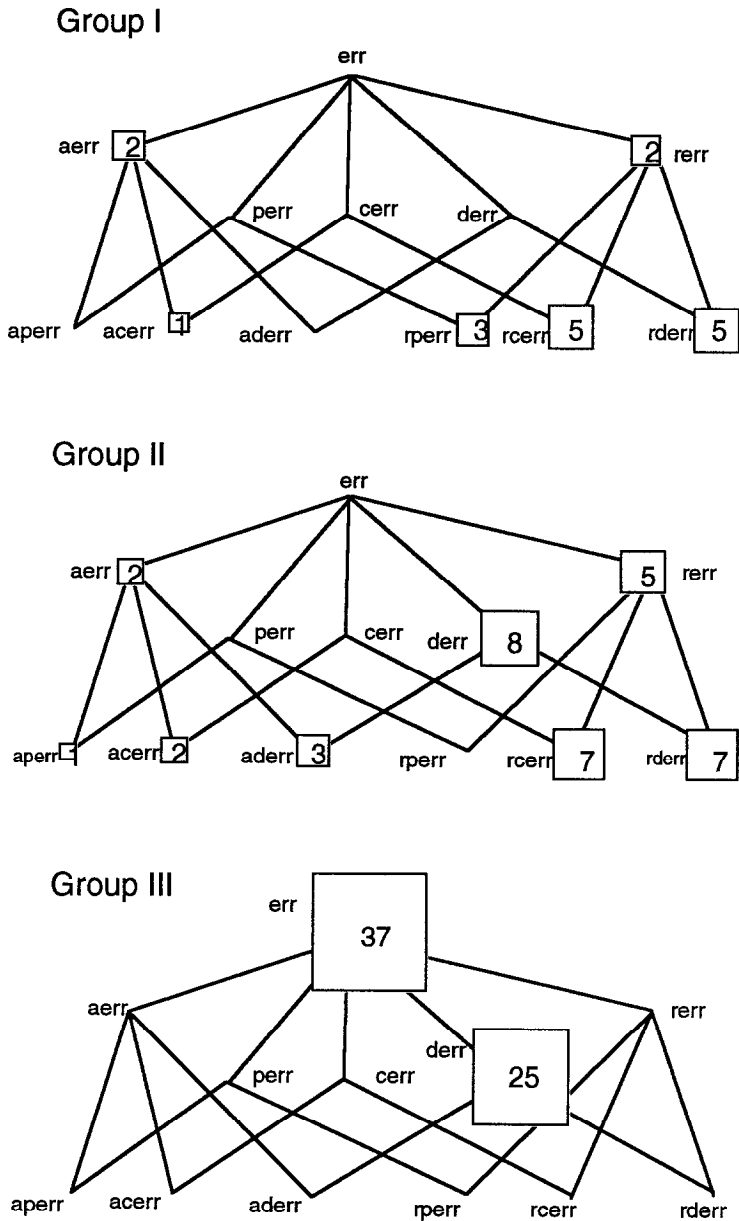
Figure 4. Use of the variables representing different kinds of measuring error. The area of the squares depicts the number of times a certain variable has been chosen.

Table 3. Number of selected/stated relations, classified according to level of preciseness, absolute scores and relative scores.

	Group		
	I	II	II
Very global (there is a relation)	16 (41%)	6 (11%)	9 (13%)
Qualitative descriptive (if A increases, B decreases)	22 (56%)	46 (82%)	60 (83%)
Conditional relations and other more precise ones	1 (3%)	4 (7%)	3 (4%)

Relations selected. In Table 3, the different kinds of relations that students have selected are listed for the three different groups. The table shows that students using the structured scratchpads stated their hypotheses on a more global level than students using the unstructured or partially structured ones (χ^2=17.7, d.f.=4, p<0.005). This contradicted our expectations. Offering more precise relations on the scratchpads did not stimulate the students to formulate hypotheses in a more precise way. On the contrary, the "relations", "there is a relation" and "there is no relation", were chosen by the students using the structured scratchpads much more often than by students using unstructured or partially structured ones. Apparently, the presence of these, very imprecise, relations on the hypothesis forms triggered the students to use it, in contrast to the more precise relations, but the students do not state these very global relations by themselves. In particular, conditional relations were almost never chosen, despite the fact that this type of relation occurs quite often in the conceptual model used and that a special tool to construct these relations, a condition selection table, was offered on the structured scratchpads.

Students' lines of reasoning. The consistency of the students' lines of reasoning was measured by matching the students' hypotheses with previous ones and by relating hypotheses and experiments.

Matching with previous hypotheses. To investigate whether students followed a more or less consistent line of reasoning hypotheses were matched against their predecessor. We wished to see whether students posed more than one hypothesis concerning the same set of variables or if they turned to other sets of variables. A *hypothesis train* is defined to be a set of consecutive hypotheses concerning one set of variables or related variables (i.e. variables which belong to the same supertype in the variable hierarchy of the conceptual model). The length of an hypothesis train is the number of hypotheses it consists of. A student who has long hypotheses trains can be regarded as a systematic learner (provided that the experiments s/he performs are relevant for testing the hypotheses). No significant differences between groups were found but for each group the average length of the hypothesis trains is very small (1.2). The number of trains that were concluded with an incorrect hypothesis found was quite high: 33% of the last hypotheses in a train were incorrect. Because of the average small length of hypothesis trains it was not possible to obtain a reasonable amount of data to investigate the development of hypotheses during the discovery process.

Table 4. The number of times an hypothesis was stated before an experiment was carried out with the variables involved

Number of hypotheses stated	Group		
	I	II	II
Before experiment	26 (67%)	21 (38%)	30 (45%)
After experiment	13 (33%)	35 (62%)	36 (55%)

Relation with previous experiments. An important question is whether the hypotheses that were entered had a basis in previously performed experiments or whether they were elicited prior to experiments. This indicator can serve as a discriminator for theorist or experimenter behaviour (Klahr and Dunbar, 1988). However, it is dangerous to use this variable as the only indicator of such behaviour, because when an hypothesis is stated at a very global level ("there is a relation") it would be incorrect to conclude that such an event would be a real "theorist event" since such an hypothesis can be merely a statement of intention to investigate a relation. On the other hand a theorist event has some value only if it is followed by an experiment which is able to put the hypothesis to the test. Table 4 shows that the group using structured scratchpads stated relatively more often hypotheses about variables with which they had not experimented before ($\chi^2=8.1$, d.f.=2, p<0.05). We may, however, not conclude that among the students using structured scratchpads the number of theorists was higher, since the hypotheses that were generated were, on average, less precise than for the other groups. Despite this fact, this result indicates that the structure on the scratchpads may contribute to the formations of ideas related to variables that were not manipulated before.

Conclusions

The basic assumption behind the introduction of hypothesis scratchpads was that they could familiarize students with the hypothesis space and therefore support the search in this space. The effect of the scratchpad was supposed to be related to its structure.

The structured scratchpads as used in our study resulted in a larger hypothesis space: learners working with this scratchpad used a larger number of different variables than subjects from the partially structured and unstructured groups. Also search of hypothesis space *before* experiments were performed was encouraged by the structured scratchpads. This is illustrated by the higher number of hypotheses stated by the students using structured scratchpads about the relation between variables before an experiment was carried out with these variables. This implies that the structured scratchpad has a potential of stimulating theorist behaviour, as identified by Klahr and Dunbar (1988).

The introduction of relations (especially conditional relations) on the structured scratchpads did not, however, have the desired effect. The number of different relations that was chosen did not differ between groups and the students using structured scratchpads even selected very global relations more often, in contrast to the students who used a scratchpad which did not have a relation selection list.

The idea that the structured scratchpads would make the process of hypothesis formation more easy was contradicted by the fact that the students using the structured scratchpads conducted less experiments and formed less hypotheses than other students. However, this might be the result of the fact that the structured scratchpads contained more information than the unstructured ones and that the experiment was conducted during a relative short period of time. A more frequent and longer use of scratchpads may result in a more efficient and effective use of this tool. Similar effects were found by de Jong, de Hoog and de Vries (1992) who found a negative effect of the amount of support on the activity level of the students. These findings indicate that support offered can be considered to be an additional task for the learner. This aspect should be taken into account for the design of interactive learning environments. The scratchpads were effective in supporting the generation of well-formed hypotheses: nearly all hypotheses stated by the students using structured scratchpads were well-formed, opposed to lower scores for the other groups.

We may, therefore, conclude that the structure offered on the hypothesis scratchpads contributed to structuring and enlarging of hypothesis space by the students, but that the support was ineffective at some crucial points.

To increase the supporting functionality of the scratchpads we may think of mechanisms to influence dynamically the properties of hypothesis space to be used by the learner. For example one can design a scratchpad which disables the choice of very global relations once these have already been chosen for a certain set of variables. This would force the students to think of more precise relations. A similar mechanism could be used for variables, to stimulate generalisation of hypotheses. It will be more straightforward to implement dynamics in structured scratchpads, as used in the present study, than in unstructured ones. Finally, dynamical scratchpads may facilitate the presentation of different *views* on the domain (White and Frederiksen, 1990).

A second, and not yet discussed, aspect of hypothesis scratchpads is that they can provide the learning environment in which they are integrated with valuable information about the learning processes. An Intelligent Tutoring System (ITS), for example, will have to maintain a learner model, representing (among other learner attributes) the current knowledge state of the learner (Duchastel, 1988; Goodyear, Njoo, Hijne and Van Berkum, 1991; Ohlson 1986; Self, 1988). This knowledge state can be well expressed in terms of the conceptual modelling language, using variables and relations, mentioned above (De Jong, Tait and Van Joolingen, 1992; Van Joolingen and De Jong, 1992). This means that the structured scratchpads as used in the experiment can provide the learner model with an accurate account of the learner's ideas about the simulation, directly expressed in terms of a language the learner model can use to represent the learner's knowledge state. The

represented knowledge state of the learner can be assessed by comparing it to a formalised version of the conceptual model (Van Joolingen and De Jong, 1992). The result of this assessment is a set of *learning indicators* (Shute, Glaser and Raghavan, 1989) which the learner model can pass to an instructional planner for deriving instructional actions.

Thus, hypothesis scratchpads may have a double function: supporting the learner and improving the functionality of simulation learning environments.

Acknowledgement

References

Alessi S. M. and Trollip, S. R. (1985). *Computer based instruction, methods and development.* Englewood Cliffs, NY: Prentice-Hall.

van Berkum, J. J. A. and de Jong, T. (1991). Instructional environments for simulations. *Education and Computing, 6,* 305–358.

Dunbar, K. and Klahr, D. (1989). Developmental differences in scientific discovery processes. In D. Klahr, and Kv . Kotovsky (Eds.), *Complex information processing: the impact of Herbert Simon.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Duchastel, P. (1988). Models for AI in education and training. In P. Ercoli and R. Lewis (Eds.), *Artificial intelligence tools in education* (pp. 17–29) Amsterdam: Elsevier.

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

Gorman, M. E. and Gorman, M. E. (1984). A comparison of disconfirmatory, confirmatory and control strategies on Wason's 2-4-6 task. *Quarterly Journal of Experimental Psychology 36A,* 629–648.

Gorman, M. E., Stafford, A. and Gorman, M. E. (1987). Disconfirmation and dual hypotheses on a more difficult version of Wason's 2-4-6 task. *The Quarterly Journal of Experimental Psychology, 39A,* 1–28.

Greeno, J. G. and Simon, H. A. (1984). Problem solving and reasoning. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey and R. D. Luce (Eds.), *Stevens' handbook of experimental psychology; Vol. 2: Learning and cognition* (pp. 589–673). New York: Wiley.

de Jong, T. (1991). Learning and instruction with computer simulations. *Education and Computing, 6,* 217–230.

de Jong, T. and Njoo, M. (1992). Learning and instruction with computer simulations: learning processes involved. In E. de Corte, M. Linn, H. Mandl and L. Verschaffel (Eds.), *Computer-based learning environments and problem solving.* (NATO ASI series F: Computer and Systems Series) Berlin: Springer.

de Jong, T., Tait, K. and van Joolingen, W.R. (1992). Authoring for intelligent simulation based instruction: a model based approach. In S. A. Cerri and J. Whiting (Eds.), *Learning technology in the European communities* (pp. 619-637). Dordrecht: Kluwer Academic Publishers.

de Jong, T., de Hoog, R., and de Vries, F. (1992). Coping with complex environments: the effects of navigation support and a transparent interface on learning with a computer simulation of decision support theory. (Submitted for publication.)

van Joolingen, W. R. and de Jong, T. (1991a). Characteristics of simulations for instructional settings. *Education and Computing, 6,* 241–262.

van Joolingen, W. R. and de Jong, T. (1991b). *An hypothesis scratchpad as a supportive instrument in simulation learning environments.* OCTO Report 91/03. Eindhoven: OCTO, Eindhoven University of Technology.

van Joolingen, W. R. and de Jong, T. (1992). Modelling domain knowledge for Intelligent Simulation Learning Environments. *Computers and Education, 18,* 29–38.

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

Lesgold, A. (1990). Tying developments of intelligent tutors to research on theories of learning. In H. Mandl, E. de Corte, S. N. Bennett and H. F. Friedrich (Eds.), *Learning and instruction; European research in an instructional context* (Vol 2.1). Oxford: Pergamon Press.

Michael, J. A., Haque, M. M., Rovick, A. A. and Evens, M. (1989). The pathophysiology tutor: a first step towards a smart tutor. In H. Maurer (Ed.), *Computer Assisted Learning. Proceedings of the 2nd International Conference ICCAL* (pp. 390–400). Berlin: Springer Verlag.

Mynatt, C. R., Doherty, M. E. and 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. and Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Njoo, M. and de Jong, T. (1991). *Learning processes of students working with a computer simulation in mechanical engineering.* In M. Carretero, M. Pope, R. Simons and J. I. Pozo (Eds.), *Learning and instruction: European research in an international context* (Volume III) (pp. 483–499) Oxford: Pergamon Press.

Njoo, M. and de Jong, T. (1992). Exploratory learning with a computer simulation for control theory: learning processes and instructional support. *Journal of Research in Science Teaching,* (in press).

Ohlson, S. (1986). Some principles of intelligent tutoring. *Instructional Science, 14,* 293–326.

Plötzner, R., Spada, H., Stumpf, M. and Opwis, K. (1990). *Learning qualitative reasoning in a microworld for elastic impacts.* Research Group on Cognitive Systems, Forschungsbericht nr. 59. University of Freiburg: Psychological Institute.

Reigeluth, C. M. and Schwartz, E. (1989). An instructional theory for the design of computer-based simulations. *Journal of Computer-Based Instruction, 16,* 1–10.

Reimann, P. (1989). Modeling scientific discovery learning processes with adaptive production systems. In D. Bierman, J. Breuker and J. Sandberg (Eds.), *Artificial intelligence and education; synthesis and reflection. Proceedings of the 4th International Conference on AI and Education* (pp. 218–227). Amsterdam: IOS.

Self, J. (1988). Student models, what use are they? In P. Ercoli and R. Lewis (Eds.), *Artificial intelligence tools in education* (pp. 73–86) Amsterdam: Elsevier.

Shrager, J. and Klahr, D. (1986). Instructionless learning about a complex device. *International Journal of Man-Machine Studies, 25,* 153–189.

Shute, V. and Glaser, R. (1990). A large scale evaluation of an Intelligent Discovery World: Smithtown. *Interactive Learning Environments 1,* 59–77.

Shute, V. and Glaser, R. (in press) An intelligent tutoring system for exploring principles of economics. In R. Snow and D. Wiley (Eds.) *Improving inquiry in social science: a volume in honor of Lee J. Cronbach.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Shute, V., Glaser, R. and Raghavan, K. (1989). Discovery and inference in an exploratory laboratory. In P. L. Ackerman, R. J. Sternberg, and R. Glaser (Eds.), *Learning and individual differences.* San Francisco: Freeman.

Shute, V., Glaser, R. and Resnick, P. (1986). *Discovering and learning to discover: an intelligent microworld for economics.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

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

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