

# An hypothesis scratchpad as a supportive instrument in simulation learning environments

**Citation for published version (APA):**

Joolingen, van, W. R., & Jong, de, A. J. M. (1991). *An hypothesis scratchpad as a supportive instrument in simulation learning environments*. (OCTO-report; Vol. 9103). Technische Universiteit Eindhoven.

**Document status and date:**

Published: 01/01/1991

**Document Version:**

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.tue.nl/taverne](http://www.tue.nl/taverne)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[openaccess@tue.nl](mailto:openaccess@tue.nl)

providing details and we will investigate your claim.

**An hypothesis scratchpad as a  
supportive instrument in simulation  
learning environments**

Wouter R. van Joolingen  
Ton de Jong

Address:  
Eindhoven University of Technology  
HG 9.35  
P.O. box 513  
5600 MB Eindhoven  
The Netherlands  
Tel: +31 40 47 42 10  
Fax: +31 40 44 98 75  
Email: [wmorwvj@urc.tue.nl](mailto:wmorwvj@urc.tue.nl)

Department of Philosophy and Social Sciences  
Eindhoven University of Technology

OCTO-report 91/03

This report was published earlier as Section 4 of SAFE report SIM/23: Prototype/mock-up of an Integrated Simulation Based Learning Environment under the title "A prototype scratchpad for hypothesis formation". Only some minor editing changes have been made to the text of this chapter.

### © **The SAFE project**

The research reported was conducted in the project SIMULATE. SIMULATE is part of SAFE, a R&D project partially funded by the CEC under contract D1014 within the Exploratory Action of the DELTA programme. In total 17 partners are involved in the SAFE consortium of which Philips TDS is the prime contractor. The partners directly involved in SIMULATE are: Philips TDS (Germany), University of Leeds, University of Lancaster (UK), TIFSA (Spain), University of Amsterdam, Eindhoven University of Technology, Courseware Europe (The Netherlands).

## Table of contents

SUMMARY .....	4
1 Introduction .....	5
2 Support of hypothesis generation and experiment design in learning with interactive computer simulations .....	5
2.1 Learner instruments as supportive elements in an ISLE .....	5
2.2 Hypothesis generation and testing .....	6
2.3 Hypotheses and experiments in relation with computer simulations .....	9
2.4 Support during the discovery process. ....	10
2.4.1 Support on specifying hypotheses .....	12
2.4.2 Support on designing experiments .....	14
2.4.3 Discussion of existing hypotheses instruments .....	15
3 4SEE: an interactive simulation for teaching error diagnosis .....	16
3.1 Domain description .....	16
3.2 Computer program and mock-up .....	17
3.2.1 The model .....	17
3.2.2 Learner interaction .....	18
3.2.3 Hypothesis scratchpads .....	19
4 Pilot study .....	19
4.1 Method .....	19
4.2 Data .....	20
4.3 Results .....	22
5 Main experiment .....	25
5.1 Method .....	25
5.1.1 Hypothesis scratchpads .....	26
5.2 Data .....	27
5.3 Results .....	27
5.3.1 Activity level .....	27
5.3.2 General functioning of the hypothesis forms .....	28
5.3.3 Assessment of hypotheses, related to the conceptual model .....	30
5.3.4 Students' lines of reasoning .....	32
6 Conclusions .....	33
6.1 The hypothesis scratchpads as supportive instruments .....	34
6.2 The scratchpads as source of information for the ISLE .....	35
6.3 Analysis of the learner hypotheses .....	36
6.4 Recommendations for the design of hypothesis scratchpads to be used in ISLEs .....	36
7 Acknowledgement .....	38
8 References .....	38
APPENDIX Analysis scheme for hypotheses forms .....	42

## SUMMARY

The present report presents a study on a supportive instrument for the generation of hypotheses by learners working with a computer simulation. This instrument will be called *hypothesis scratchpad* and can be used within an Intelligent Simulation Learning Environment (ISLE). The main objective of this research is to derive recommendations for the structure of such hypothesis scratchpads as part of a supportive environment around computer simulations.

Definitions of the concepts of *hypothesis* and *experiment* are derived from the structure of domain knowledge, based on previous work within the SIMULATE project (Van Joolingen & De Jong, 1991). An hypothesis is a statement that a relation holds between two or more variables or variable types. An experiment is a manipulation of one or more variables and the observation of other variables affected by these manipulations.

These definitions lead to a descriptions of *hypothesis space* and *experiment space*, concepts introduced by Klahr and Dunbar (1988) to describe scientific discovery as a search process in these two spaces. The properties of these spaces are investigated and the possible problems that may arise related to the generation of hypotheses and the design of experiments are described. This investigation lead to a draft structure of hypothesis scratchpads which were tested in a pilot experiment. The structure consists of the offering of information on *variables* that are present in the simulation, which the learner can use to construct his/her hypotheses.

The results of the pilot study show that learners are poor constructors of hypotheses, especially the description of relations is a problem. Also little generalisation of hypotheses is observed. This lead to a new design of scratchpads where the variables of the runnable simulation model have been replaced with the variables from the conceptual model and a relation construction aid was added to one version of the scratchpad. The choice of variables and relations that appeared on the scratchpads was based on a conceptual model of the domain.

In the main experiment three different versions of the hypothesis and experiment scratchpads have been compared. One 'structured', offering support on both the variable selection and relation construction, a 'partially structured', offering support on only the variable selection and an 'unstructured' scratchpad. It is found that the students using structured scratchpads construct more well-formed hypotheses than students using one of the other two scratchpads, but that they tend to state hypotheses at a more global level (less precise).

The results of this experiment lead to recommendations on the design of hypotheses scratchpads. The concept of *dynamical scratchpads* is introduced. Dynamical scratchpads should adapt to the learner in specific ways and, in that sense, become a means of system initiated support.

## 1 Introduction

Learning with computer simulations is a form of discovery learning. Discovery learning is a complex process involving a large number of specific learning processes. Therefore learners often experience discovery learning as difficult and demanding. A way to tackle this is to embed the simulation in a instructional/supportive environment. This study will explore the possibilities of supporting learners by providing *learner instruments* for performing the learning processes of 'hypothesis generation and 'experiment design'. Focus will be on the first of these two learning processes.

We start by describing the learning processes that play a role in learning with computer simulations as well as the problems that learners may have while performing the learning processes: generating hypotheses and designing experiments. Also an outline is given of a part of a supportive environment for computer simulations that may be helpful for the learner to overcome these problems.

The types of support that will be outlined are based on the typical structure of computer simulations, described in earlier SIMULATE reports (e.g. Van Joolingen and De Jong, 1990, 1991) and the consequences of this structure for the domain representations for Intelligent Simulation Learning Environments (ISLEs).

Next we will describe the interactive computer simulation and the mock-up of a supportive environment that has been developed and used for the experimental part of this study. Sections 4 and 5 describe an experimental study that has been performed with this prototype. In Section 6 some general conclusions are

drawn which result in recommendations for the design of hypothesis instruments.

## 2 Support of hypothesis generation and experiment design in learning with interactive computer simulations

### 2.1 Learner instruments as supportive elements in an ISLE

As was described in several SIMULATE documents, the learning processes involved with exploratory learning and learning with computer simulations in particular have a complex nature. In Goodyear Njoo & Hijne (1990) (see also Njoo & de Jong, 1989; De Jong & Njoo, 1990; Reigeluth and Schwartz, 1989) these learning processes investigated. They distinguish the following main categories of learning processes (apart from regulative processes, concerned with planning or processes involved with the operation of the simulation system):

- Orientation
- Hypothesis generation
- Hypothesis testing
- Evaluation

The first of these four learning processes is necessary for the learner to get his/her first (often vague) ideas about the elements in the model underlying the computer simulation. These ideas can be used to invoke the second learning process: the generation of an hypothesis about the simulation. This hypothesis must be put to the test to become a part of the learners (mental) model of the simulation. This testing includes the design of an *experiment* which will be performed with the simulation, *predict-*

ing the outcomes of the experiment, performing the experiment and interpreting the results. These results are then evaluated. This evaluation may lead to a rejection or a suggestion of more evidence for the hypothesis and may give rise to the generation of a new hypothesis or a reformulation of an old one. With this new hypothesis the process of testing may start over again. Also the learner can choose to investigate another part of the simulation and state hypotheses about that part. This process can continue until all parts of the simulation are investigated and the learner has discovered the complete model.

Crucial in the discovery process is the *generation of hypotheses*. Research has shown that this part of a discovery process is both the most important and the most problematic part of discovery learning (Mynatt et al. 1977; 1978). Also *designing the right experiment* to test an hypothesis has been shown to be a problematic issue. (Wason, 1960; Gorman, & Gorman, 1984; Gorman, Stafford, & Gorman 1987).

Since the complexity of the learning processes described above is acknowledged, one may search for possible ways to support them. In general, one should strive to ways of support that are non-directive, i.e. which do not destroy the free exploratory nature of the ISLE. A possible realisation of such support can be the offering of *learner instruments*, tools that the learner can use to ease the performance of certain learning processes.

The objective of the present study, performed within SIMULATE activity III, is to investigate the possible characteristics of a learner instrument for the

generation of hypotheses. For this purpose a computer simulation has been developed, together with some prototypical learner instruments (*hypothesis scratchpads*). The effects of the scratchpads on the performance of learning processes have been investigated by means of an experimental study.

The present section of this report gives some theoretical background on the subject and serves as a rationale for the design of the used hypothesis scratchpads.

## 2.2 Hypothesis generation and testing

Klahr and Dunbar (1988, see also Langley & Klahr, 1987; Shrager & Klahr, 1986) studied the forming of hypotheses and the design of experiments to test these. They describe the scientific discovery process as *dual space search*. The two spaces that are to be searched are the *hypothesis space*, containing all possible hypotheses about the system under study, and the *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 going through this space. One of these strategies, used by, what they call, 'experimenters', consists of a first phase in which an hypothesis is tested, followed by a phase where the subject searched the experiment space without explicitly stating an hypothesis. The main characteristic of the experimenters is that they have to perform an experiment which rules out all other possible hypotheses before they actually state the correct hypothesis.

The second strategy, used by so-

called 'theorists', never performs experiments 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 enough contradicting evidence has been found. In most cases this new hypothesis does 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. In general theorists require less experiments than experimenters to reach the right conclusion.

Based on these findings Klahr and Dunbar developed a model of "scientific discovery as dual search" (SDDS). 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). The basic assumption of SDDS is that scientific reasoning requires search in two related problem spaces, as noted above. Search in the hypothesis space is guided by prior knowledge and by experimental results, search in experiment space may be guided by the current hypothesis or by general manipulative knowledge, and it may be used to generate information to formulate or test hypotheses.

The three basic 'components' of SDDS are: SEARCH HYPOTHESIS SPACE, TEST HYPOTHESIS and EVALUATE EVIDENCE. The first component outputs a fully specified hypothesis to the second which makes a prediction and generates evidence which can be evaluated in the third component. Each component consists of a number of subcomponents which includes moves in experiment space, running experiments and using prior knowledge. An important feature

of the SDDS model is that during the first phase (Search Hypothesis space) experiments with the object of study can be carried out in order to form hypotheses.

The three main phases in the SDDS model correspond to three of the four learning processes that are distinguished by Njoo and De Jong (1989), making the two findings consistent at a global level. The SDDS model, however, does neither try to describe the process of *generating* hypotheses nor the choice of experiments to test them. A part of this problem is tackled by a study by Shute et al.(1989)

Shute, Glaser and Raghavan (1989) (see also Shute and Glaser (1990), Shute and Bonar (1986), Shute, Glaser, & Resnick (1986) and Shute & Glaser (1986)) report about a system for learning laws of economics, *Smithtown*. The system is a free explorable computer simulation of an economic system. Students are invited to explore the simulation to discover the laws that determine the underlying model. Shute et al. investigated the scientific research behaviour of students.

One interesting feature of *Smithtown* is the presence of a *Hypothesis Menu*, which supports the students in stating hypotheses about the model. The hypothesis menu offers a structured framework to enter 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 to form well formed sentences and to



specify the hypothesis more precisely respectively. An example hypothesis entered in the hypothesis menu could be: "as price increases then quantity demanded decreases". The system matches the hypothesis entered with its internal database of economic knowledge. The articles by Shute et al. are not clear on how this matching is performed and to what kinds of action the results of the matching may lead.

The instrument offered by Shute and Glaser primarily offers support on the *formation* of hypotheses, making sure that an hypothesis is well-formed. There is little or no support on *finding* hypotheses nor on *designing experiments* to test them.

Shute et al. find the same difference between subjects as Klahr and Dunbar (1988). They refer to "theorists behaviour" as "*hypothesis driven*" and to experimenter behaviour as "*data driven*". Moreover they conclude that the hypothesis driven subjects are more successful than the data-driven subjects.

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

This *confirmation bias* was also found by Mynatt, Doherty and Tweney (1977; 1978). They confronted students with a computer simulated environment of moving particles and fixed objects. The students were asked to discover the rules of motion (which were not the rules of physics in this case). Moreover, they find that students abandoned falsified hypotheses only 30 % of the time. They conclude that students failed to collect and to use falsifying information.

From the research that is described above we may conclude that there is already much knowledge on the general experimentation behaviour of students working in exploratory environments but little is known about the details of the generation of hypotheses. Especially little is known about the hypotheses generation in complex situations, like most educational computer simulations are. Research is needed in order to be able to support students to generate hypotheses about complex simulations.

Also the design of experiments to test hypotheses about a complex simulation requires additional research. Most research that has been performed so far applies to relatively simple situations with binary results of the experiment (the rule that is to be discovered applies or applies not to a certain situation, no other experiments are possible) and a simple experiment space. We need additional knowledge about the search of more complex experiment spaces by students. Until now only general experimentation advice has been given to students like "vary one variable at a time" (Lavoie and Good, 1988).

In the next section the meaning of the concepts 'hypothesis' and 'experiment' will be elaborated for the context of computer simulations.

### 2.3 Hypotheses and experiments in relation with computer simulations

In Van Joolingen and De Jong (1991a) the main characteristics of computer simulations have been investigated. The driving force behind the simulation is a (*runnable*) *model* which consists of *variables* and *rules* which determine the change of the variables.

In Van Joolingen and De Jong (1991bc) an alternative model has been described, which contains additional knowledge to enable reasoning about the domain, teaching about the domain and diagnosis of learner knowledge: the *cognitive model*. An important part of the cognitive model is the *conceptual model*, which contains a conceptual description of domain elements and relations. The basic elements of the conceptual model are *variables* or *variable types* (to be distinguished from 'runnable variables') and *relations*.

It is the conceptual model that must serve as a description of the model that may be presented to a learner or that s/he must discover. That means that hypotheses about the model must be formulated in terms of relations that could be part of the cognitive model, i.e. in terms of a relation between two or more variables that are part of the state description of the model.

More precisely: *An hypothesis about a simulation model is a statement that a certain relation holds between two or more variables or variable types present in the cognitive model and the description of the attributes of that relation.*

This definition of hypotheses about simulation models determines the di-

mensions of the hypothesis space. This space is a superposition of the space of all possible variable (type) combinations and the space of all possible relations between these variables. In practice this space can be limited by a priori excluding irrelevant variable combinations and irrelevant relations.

To generate an hypothesis a learner has to find a node in hypothesis space. To find this node s/he has to:

- identify the variable set on which the hypothesis will be applicable
- identify that there is a relation between the variables
- define the nature of the relation.

This implies that there can be several problems with the generation of hypotheses:

- A failure to identify the variables or variable types that are present in a simulation. In the classification of Njoo and De Jong (1990) the identification of variables is part of the learning process 'model exploration' (which is a subprocess of Orientation). It is clear that for the generation of hypotheses the identification of variables is a prerequisite.
- A failure to select relevant variable combinations that require investigation and thus are potential variable combinations to state an hypothesis about.
- A failure to find possible relations between the variables (apart from the fact if such a relation is the correct one in terms of the cognitive model).
- Stating relations at a too global level, which makes testing an

irrelevant activity (testing the statement "there is a relation" is not useful, most of the times).

- Stating relations that are too specific for one particular variable combination making generalisation difficult.

Another problem can be that the learner tries to understand too much at a time about the model:

- The formulation of an hypothesis that contains too many variables so that it cannot be tested by a small set of experiments.

The possible problems stated above consider the identification of nodes in the hypothesis space, the recognition that a certain statement is indeed a statement about the model that is underlying the simulation. Another kind of problems is related by moving between nodes in the hypothesis space.

Moving between nodes in hypothesis space is triggered by the evaluation of an experimental data coming from the simulation. In an ideal case a learner would reject his hypothesis on the basis of disconfirming data and keep (and possibly refine) it on the basis of confirmatory data. In practice it has been found that this is certainly not always the case (Klahr & Dunbar, 1988; Mynatt, Doherty, & Tweney, 1987).

Another important activity in learning with a computer simulation is to collect data. This data can be collected by performing an *experiment* with the simulation. Such an experiment consists of *varying one or more input variables or parameters of the simulation and collecting data on the behaviour of one or more output variables*. In most cases

an experiment will be performed to test an hypothesis, which may result in a move in hypothesis space.

Since an experiment is performed to investigate a relation (i.e. the relation that is formulated in the hypothesis) in the simulation model, it is important to know *which* relation is being investigated. In general there are two possibilities:

- The relation between the manipulated input variables and the observed output variables is investigated directly. We can call this *direct control* over variables.
- The relation between two or more output variables is investigated. A subset of the output variables then has a known relation to the input variables that are being manipulated. This relation must be known to the learner. This kind of variable control will be called *indirect control*.

An experiment is always performed with single variables of the runnable model, not with variable types or conceptual variables. This means that when the hypothesis is formulated in terms of variable types the learner will have to select representatives of these types.

In general it is wise not to vary too many variables at a time. In most cases the manipulation of only one variable is the best choice but sometimes, especially with indirect control at hand, a controlled variation of more than one variable can be appropriate.

#### 2.4 Support during the discovery process.

In the preceding section the various learning processes that should take place

## *An hypothesis scratchpad for ISLEs*

during the exploration of a computer simulation were discussed. A summary shows that the three most important aspects of hypothesis generation are:

- *identify* variables (or variable types)
- *select* variable (types) for relation
- identify & select the *relation* that is hypothesized to hold between the selected variables;

the design of experiments consists of the following phases:

- *select* variables to manipulate
- *select* variables to observe
- determine *control* (direct/indirect)
- determine *manipulation* of variables

a prediction is characterised by the following aspects:

- *select* the variable(s) about which the prediction is made
- *predict* effect of manipulation

There are several types of problems that learners may encounter when confronted with these learning processes, as noted in Section 2.3. To overcome or prevent these problems an *hypothesis scratchpad* can be offered to the student as a tool to ease the generation of hypotheses. In the present section possible features of such an instrument will be described. In Section 3 a mock-up of this envisaged tool will be presented.

Since the scratchpad may give support on both the hypothesis generation and the experimental design it must help the student during some or all of the six subprocesses listed above. On the other hand it may not be imperative in the sense that it actually forces students to

generate a particular hypothesis and/or carrying out one special experiment. Furthermore it may not give away too much information, such as providing ready made hypotheses, while the actual generation of hypotheses is one of the goals of the self-discovery process.

The bare presence of an hypothesis scratchpad can already have some advantages for the learner in his/her discovery process:

- The fact that there is a scratchpad for noting down hypotheses and that the learner may be stimulated to use it, may motivate the learner to actually form hypotheses. Often learners just experiment without any hypotheses at all.
- A scratchpad may reduce the memory load of the student (De Jong & Njoo 1990).
- A scratchpad may help the learner in making his hypotheses explicit.

From the viewpoint of a learning environment, the scratchpad may be a valuable source of information about the learners learning progress, which can be incorporated in the learner model.

In the following two subsections the support the tool may give at each of the subprocesses that are listed above will be described. This description does not imply that all possible support that is described should be present in a learner instrument for specifying hypotheses. Research will have to show which kinds of support are the most effective and which kinds are unwanted.

Section 2.4.3 will discuss some instruments that are already present in existing simulation learning environ-

ments.

It will appear that most kinds of support that can be given to a learner rely heavily on the conceptual model. Therefore this model is the basic source of information for the supportive instrument.

#### 2.4.1 Support on specifying hypotheses

In identifying variable types the tool could give some help by stating which *variables* or which *variable types* are present in the simulation. Listing the variables can be of help when not all variables that are important for the cognitive model are 'visible at the surface'. A list of variables can draw the learners attention to variables that could remain unnoticed.

A list of variable types can be important to stimulate the learner to generalise his/her hypotheses from hypotheses about single variables to hypotheses about variable classes. Listing the variable classes that are present provides the learner with a framework which he can use to classify the variables that s/he recognizes to be present in the simulation.

In *selecting* the variables or variable types a listing as described above could also be helpful (when a learner knows where to choose from the choice is easier). Extra help could be offered by providing possible variable combinations to choose from e.g. by indicating which variable combinations are worth investigating. The variable combinations that would be offered should be combinations that form important relations in the cognitive model. However, offering this choice in a restrictive way, i.e. by forbidding other choices would restrict the learners freedom.

It is also possible to offer the learner a less rigid structure on the set of variable types. Often different levels of abstraction can be distinguished in the variable types. For example in *Smithtown* (Shute, Glaser & Raghavan, 1989, see Section 2.2) on one hand there exist variables like price, supply and demand, which are rather low level concepts; on the other hand variables like supply curve and equilibrium point exist which correspond to higher levels of abstraction. By making the difference in levels clear to the learner and indicating which levels might interact, the identification of important variable pairs may be supported.

The *selection of relations* could (again) be eased by offering a list of possible relations that the learner may use to construct his/her hypotheses. The choice of relations that will be on the list offered is a difficult one. The list must provide at least the relation that is the correct one but also some incorrect but plausible possibilities. Also the list must provide relations at several levels of preciseness. It is to be expected that while investigating a certain relation a learner will start with stating more or less global relationships and move to more precise relationships as s/he proceeds with his/her investigation. Some structure in this list, indicating the different levels of globality may help the learner to state his/her hypothesis at the desired level of preciseness.

The list of relations that is offered may be dependent of the variables selected, so that the learner is offered a choice, only from relations that are thought to be likely to be selected. One must be aware that there is quite a great danger of overrestricting the learner in his/her freedom when the number of

## *An hypothesis scratchpad for ISLEs*

relations to choose from is too small. The learner should have some alternatives to choose from.

Until now the possible support that has been described has a rather passive or non-directive nature. The help described consists of offering possible choices to the learner to help him/her to structure his/her thoughts about the model. With the help of these choices the learner should be able to construct a *well-formed hypothesis*. Some types of support (those involved with variable selection) also may affect the *relevance* of the hypothesis. A selection of an important variable pair for an hypothesis automatically increases the relevance of the hypothesis. Neither well-formedness nor relevance does imply that the hypothesis is correct, but only that it is syntactically valid and comparable to the cognitive model. Well-formedness also means that the hypothesis can be tested, at least in principle.

On the basis of the cognitive model the hypothesis that has been entered can be analyzed. The information that results from this assessment can be used for tracking the development of the learners ideas about the model and therefore contribute to the learner model that is present in the simulation. The assessment can emphasise following aspects of the hypothesis:

- The *relevance* of the hypothesis. Despite support given as above the learner may still select irrelevant variable combinations for his/her hypothesis.
- The *correctness* of the hypothesis. The hypothesis could be matched against the cognitive model and incorrect hypotheses could be traced.
- The *complexity* of an hypothesis. A well-formed hypothesis may contain many variables. An hypothesis that takes too many variables may be too complex and can therefore not be tested by a small set of experiments. This analysis could, for example, lead to feedback to stimulate the learner to try more simple hypotheses first. In order to be able to make this kind of analysis the learning environment needs an algorithm to determine the complexity of an hypothesis. Another form of complexity is related to the cognitive model. In the cognitive model two variables can have a direct relationship with each other or an indirect, via 'intermediate' variables. The more variables serve as an intermediate, the more complex a relation/hypothesis is.
- The *relation of the hypothesis with previous hypotheses and the results of previous experiments*. The system could compare the hypotheses stated with previous hypotheses and previous experiments and signal things like:
  - hypothesis inconsistent with previous one, but no disconfirming evidence
  - sticking to an hypothesis after having received disconfirming evidence
  - going from a precise to a less precise hypothesis.

Of course also positive learning indicators could be signalled such as:

- generalising hypothesis
- going from global to precise hypotheses.

Shute and Glaser (1990, Shute, Glaser and Raghavan, 1989) present an overview of *learning indicators* which includes several of the above mentioned indicators.

The assessment as described above could lead to a more active or directive support from the hypothesis instrument, i.e. feedback on the hypothesis entered. This active support goes a little further than the passive support described at the start of this section. Whereas the passive support only gives guidelines to the learner the active support actually interferes with the discovery process.

#### 2.4.2 Support on designing experiments

After an hypothesis has been generated an experiment must be designed to test it. In *selecting* the variables that will be observed or manipulated in the next experiment the learner can again be offered a choice for the variables to select. It is important that s/he chooses representatives of the variable classes that take part in the hypothesis. The learner could be assisted in this choice in two ways: by restricting the possibilities of choice or by giving feedback on the choice.

Also some general experimentation knowledge could be useful to generate feedback on the choice of variables. Such knowledge includes, for example, that it is not wise to vary too many variables at a time (see e.g. Rivers and Vockel, 1988).

In determining the *control* of the vari-

ables the learner may be confronted with direct or indirect control. In the case of direct control there is no problem: the learner can manipulate the variables chosen directly. In the case of indirect control the learner has the task to select the variables to manipulate in order to vary the variables that are part of the hypothesis under investigation. In this case the learner must make use of knowledge about the relation between the variables manipulated and the variables observed. This knowledge may be prior knowledge, but also discoveries made earlier about the simulation. If the system has knowledge about the learners knowledge it can remind the learner of useful relations for indirect control of variables. Also warnings for *side effects* of indirect control may be given.

In determining the *manipulation* of the variables the instrument could offer support by listing possible manipulations. Examples of such manipulations are 'increasing' and 'decreasing', as general terms, and changing with specific amounts, dependent of the relation that is investigated.

Manipulation can take place at different levels of preciseness and in different types. Examples of the different levels of preciseness are (in increasing order): 'change a variable', 'increase a variable' and 'add a fixed amount to a variable'. Different types of manipulation can be: 'multiply a variable with a factor', 'increase it with a fixed amount' or 'take it over a threshold'.

There is also help possible from a learner instrument as far as *prediction* is concerned. To state a prediction it is also necessary to select the variable(s) for which the prediction will be made. Of course, in good experimental design

these variables will be the observed variable(s) that is(are) defined for the experiment. After this selection a statement should be made what will happen with the selected variable(s). This prediction can be stated in the same terms as the manipulation of the variables that will be changed by the learner. A similar list of possible choices could therefore be offered as support for the learner.

In principle the simulation learning environment should be able to generate a prediction of the outcomes of the next experiment and test the learner prediction against its own prediction and give feedback if both predictions are not compatible.

Just like the hypothesis scratchpads, the information that the learner enters on the experiment/prediction scratchpads can be assessed to generate learning indicators. Several possibilities are (see Shute et al 1989, 1990):

- the number of variables in an experiment
- the relevance of an experiment to the stated hypothesis (are the changed variables the ones occurrent in the hypothesis).
- the manipulation of variables (e.g. is the change great enough).

In the research reported the emphasis is laid on the analysis of hypotheses, the analysis of experiments has only been done for the purpose of data interpretation. The results of experiment analysis were not systematically interpreted, they only served for a better understanding of learner hypotheses.

### *2.4.3 Discussion of existing hypotheses instruments*

In literature many simulation learning environments are described. In some of these environments a learner instrument is available that can be thought of as an hypothesis instrument. In this section some of these existing instruments will be discussed. Their features will be matched against the possible support facilities that were listed in the previous sections.

The hypothesis menu in *Smithtown*, designed by Shute et. al. (1989, see Section 2.2) contains some elements for the support of the generation of hypotheses. In the terminology used in this chapter the list of objects is a list of variable types. In the same terminology the verbs in the hypothesis menu are relation constructors.

In this sense Shutes hypothesis menu offers support to design a well-formed hypothesis, no support is being given for the selection of important variable combinations or relations that are likely to occur in combinations with certain variable combinations. This means that the hypothesis menu does not give support on the selection on possible important parts in the model.

*Smithtown* compares the hypotheses that are entered with a database of economic knowledge. This database can be considered as being part of a cognitive model. However, the reports of Shute et. al. are not clear about what is being done with the results of this matching, for example if any feedback is being derived from it. Also the precise nature and structure of the database of domain knowledge remain unclear.



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 defect 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 hypothesis 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.

Plötzner, Spada, Stumpf and Opwis (1990) use an instrument for experimental design and prediction in a micro-world for collision mechanics (*DiBi*). They provide a kind of direct manipulation interface for this instrument. First the learner can adjust the variables to be manipulated (forces) by adjusting arrows. Second they can state a prediction using the same interface. The microworld is capable of simulating both the 'real' movement and the movements predicted by the learner so that the learner can make a comparison to adjust his/her hypothesis.

The scratchpads used by Plötzner et al. preselect the variables to manipulate in the experiments with *DiBi* and the variables to make predictions about. By using the arrows as representation for values also the range of possible values is somewhat reduced. Both preselections for the student imply a significant reduction of the experiment space. It will be strongly dependent of the simulated domain if such a reduction will be possible and/or admissible.

The next section will describe the simulation that has been developed and the hypothesis and experiment scratchpads that have been used in connection with this simulation.

### 3 4SEE: an interactive simulation for teaching error diagnosis

In this section the computer simulation that has been developed will be described. First a description of the domain (error analysis in chemistry) will be given, followed by a description of the simulation program itself.

#### 3.1 Domain description

A well known problem in chemical experimentation is that measurements never are completely accurate. Every measurement results in a value that is not the real value of the measured quantity but a value more or less close to it. In general two types of sources may be the cause of this *measuring error*:

- A systematic error in the measuring apparatus, for example a ruler that is too short or a balance which always generates values that are too small. Errors resulting from this kind of cause are called *systematic* or *determinate errors*. Their characteristic is that they always have the same size and sign for every measurement that is performed.
- Errors resulting from random, uncontrollable events, such as noise. These are called *random* or *indeterminate errors*. Their size and sign varies for each mea-

surement that is performed.

When a large number of measurements of the same quantity is performed the measurements will be spread around a mean value in a normal distribution, due to the random error. The mean value will be the real value of the measured quantity, minus the systematic error. The standard deviation of the distribution is a measure for the occurring random errors.

A value that depends on a measured quantity (by calculation) is also not known precisely. This value also shows systematic and random errors, caused by errors in the values on which it is dependent. There are methods to calculate an estimate for the errors in these kinds of quantities using the errors in the values that take part in the calculation. Sometimes another estimate for the errors can be found by repeated measurements of the dependent quantity. A (significant) difference between the two estimates for the errors is an indication that a contributing error has been overlooked or that the experiment was not carried out in a proper way.

The task of an experimenter (the learner in this case) is to decide:

- which standard deviation estimate can and will be used.
- which measurements contribute to the error.
- if the obtained data set is in concordance with the estimate.
- how the total error can be made as small as possible

The prototype that has been developed explains the occurrence of errors during measurements by simulating a simple

chemical experiment: a titration. The students that will use the simulation are familiar with the titration experiment but not with the statistical aspects of measurement. The chosen experiment (titration) is not important: emphasis will be on the statistical aspects of the measurements and on the ways to achieve optimal experimental results. The main goal of the computer simulation is to provide the learner with a (conceptual) model which s/he can use to make the decisions above.

### **3.2 Computer program and mock-up**

In this section the computer simulation that has been developed for the purpose of the experiment and the mock-up of an hypothesis and experiment learner instrument will be described. First an outline of the simulation model will be given, in Section 3.2.1, followed by a description of the interaction of the learner with the simulation in Section 3.2.2. In Section 3.2.3 the mock-up of the learner instrument will be discussed.

#### *3.2.1 The model*

The model that is underlying the simulation consists of two submodels. Both submodels can be decomposed further into smaller units but that decomposition will not be discussed here explicitly. One submodel describes the titration experiment, the other describes the statistical features (based on the theory of measurement). The titration model is a simple dynamic model for the calculation of the concentrations and other quantities in the given experiment. The statistical model is more important and will be discussed in more detail.

The most important concept that is present in the statistical model is that of standard deviation. According to the theory of measurement the values that will be measured are scattered around a "real value" with a standard deviation  $\sigma$ . In principle this  $\sigma$  is only determinable from an infinite series of measurements (in practice some large finite number is sufficient). In practical situations such a large number of measurements often cannot be performed. Especially in a titration experiment the number of measurements that can be performed is limited. The reason for this is that titration measurements are time-consuming and that they use a relatively large amount of the sample that is being investigated. In practice typically three or four titrations are performed for one sample.

In such situations other methods are available to make an estimate of the standard deviations. One is to make use

of data available from the measurement instruments used. The standard deviations for measurements with each of these measuring instruments are known or can be obtained from calibration measurements, provided that there is a sample of well known properties. These "partial errors" can be combined into an estimate for the total standard deviation.

Another way to make an estimate of the standard deviation is using the small data set of performed measurements. An adapted formula for calculating the "real" standard deviation is available.

### 3.2.2 Learner interaction

The simulation serves as a tool to investigate the properties of the statistical model. The program consists of two parts, an introductory part and an exploratory part. In the introductory part the student is made familiar with the experi-

Overzicht van de experimenten tot nu toe	
1. Gekozen oertiterstof	1
2. Gekozen indicator	A
3. Hoeveelheid oertiter (globaal) (mg)	230.0
4. IJking balans: aantal keer	4
Standaardafwijking balans (mg)	14.39
5. IJking buret: aantal keer	3
Standaardafwijking buret (nl)	0.0063
6. Aantal bepalingen	2
Resultaat stellen (mol/l)	0.0354
Standaardafwijking (gemeten)	0.0024
Standaardafwijking (berekend)	0.0625
7. Keuze pipet	10 nl
8. IJking pipet (aantal keer)	4
Standaardafwijking pipet	0.027
9. Aantal bepalingen	2
Resultaat	2.0861
Standaardafwijking (gemeten)	0.0274
Standaardafwijking (berekend)	0.1095

Geef aan welke parameters u wilt wijzigen door het betreffende nummer in te toetsen. Bevestig uw keuze met <enter>. Type B voor het betrouwbaarheidsinterval.

Figure 1: The control centre of the simulation. An overview is shown of the results of the first experiments. On this screen the learner can enter which variables s/he wants to change.

mental setting, the variables s/he can control and the operation of the computer. During this guided tour the learner enters a first choice of the several parameters and variables that play a role in the simulation.

In the exploratory part the learner can control nine variables that are important for the experiment. Figure 1 shows the screen on which the learner can indicate which variables s/he wants to change. After the learner has confirmed this choice the s/he is asked to enter new values for the selected variables. With this new values those parts of the experiment that depend on the changed variables are repeated. After this repeated experiment the program returns to the screen shown in Figure 1, with the results of the new experiment displayed. Up to four experiments can be listed at the same time.

The interaction (Van Joolingen & De Jong, 1991a) with the model is fixed. The order of access to the variables is controlled by the computer, even in the exploratory part, where the repeated experiment is performed in the order prescribed by the computer program.

### 3.2.3 *Hypothesis scratchpads*

During the exploration of the computer simulations the learners were offered a set of forms which they could use as an hypothesis instrument. After each run of the simulation they were asked to fill in one or more forms, prompted by a message on the screen.

The scratchpads that were used were different for a pilot experiment and the main experiment that will be described in Sections 4 and 5. Therefore the

precise description of the scratchpads used will be presented together with the description of each experiment. In this section only the general characteristics of the scratchpads will be described.

The forms that were used always consisted of two parts: an *hypothesis part* and an *experiment/prediction part*. In the version used in the pilot experiment these two parts were located on the same form, in a later version the parts were separated into two different forms.

In the next two sections two experimental studies are reported. In Section 4 a pilot study is described which was designed to define the experimental settings for the main experiment. This main study is described in Section 5. The basic question for the main study was: "How do structures of an hypothesis scratchpad stimulate the generation of hypotheses by students and how do these structures influence the attributes of the hypotheses generated and the experiments designed with the simulation".

## 4 **Pilot study**

With the simulation that was developed and two first versions of the scratchpads a pilot experiment was conducted. The main goal of this pilot experiment was to find if students understood the idea of hypothesis scratchpads and to have a first investigation of hypotheses generated by students, in order to design the hypothesis scratchpads for the main study.

### 4.1 **Method**

Subjects were six first-year students of chemistry. These students had received

some formal education in error analysis, and almost a year of experience with chemical experimentation, as part of their normal study program. Three of the students received unstructured scratchpads, three received structured ones. Each of the students worked alone with the computer simulation. An observer was present with each session. The observer did not give any domain information to the students, only some hints on the operation of the computer program and the use of the scratchpads were given.

All students received a written introduction, with an explanation of the use of the hypothesis forms. On this instruction were some general examples of hypotheses. The instructions were different for the two versions of the scratchpads. Apart from this written introduction they also were instructed orally.

The unstructured scratchpads contained areas for hypothesis and experiments as described above. No hints were given on the scratchpad about the form the hypotheses should have. The scratchpad further had areas where the student could state if there was supporting or contradicting evidence for his/her hypothesis and a question how certain the student felt about the hypothesis stated.

The structured scratchpads had the same basic structure as the unstructured ones, but as an extra they contained in both the hypothesis and experiment part a set of eighteen variables which the learner could select to construct his/her hypothesis and experiment. The variables that were offered in the list were variables that were directly present in the simulation: all variables that were present on the overview screen (see Figure 1) were repeated in the list. The

two lists for the hypothesis part and the experiment part were exactly the same.

## 4.2 Data

The data that was collected consisted of the forms filled in by the students and the interaction of the students, collected in a log-file. Also thinking-aloud protocols were collected from each student.

The data was divided into meaningful units. A meaningful unit could be:

- An hypothesis entered on a scratchpad
- An experiment entered on a scratchpad
- A prediction on a scratchpad
- A remark on the analysis of experimental data on a scratchpad
- An physical action with the simulation, reconstructed from the log-file.
- A protocol utterance which could be classified in one of the categories hypothesis, experimental plan, prediction or data analysis, or an utterance which was directly related to physical operation of the simulation.

The meaningful units were classified in five categories: hypothesis, experiment, prediction, performance and data analysis. The choice of these categories was based on the learning processes identified by Njoo and De Jong (1990).

The data was carried to data sheets which had columns for each of the five classification categories. An example of such a sheet is shown in Table 1. The vertical dimension of the sheets represents time. When units appear on one line in the table it could not be determined which event happened first,

*An hypothesis scratchpad for ISLEs*

**Table 1** *Extract of a datasheet, the texts in the sheet have been translated to English. Italics represent protocol utterances. Other cells are filled with 'hard' data (log-file and filled-in forms).*

Hypothesis	Experiment	Prediction	Performance	Data-Analysis
<i>If the first becomes less accurate the second also will.</i>				
If the error in the volume of the pipette is inaccurate, the result will also be inaccurate	I will change the number of measurements of the pipette and watch the result of the determination	It will be more accurate		The measured standard deviation is big but smaller than that of the pipette
If the error in the balance decreases then the error in the burette increases and the error in the result decreases	Change: the number of calibrations of the balance and the burette, watch the result	All will become more accurate and the result will change a little		The inaccuracies of all three variables support the hypothesis but the burette is not so inaccurate as you might expect from the balance.
<i>If one s.d increases all others increase too</i>				
			9 calibrations of the balance	
<i>I expect that the more you read a measurement the more accurate it will be</i>				
			20 calibrations of the burette	
				<i>The result did chance, that is what I expected.</i>
				<i>The standard deviation increased, that is strange</i>
				<i>The calibration of the burette has become worse, I did not expect that. Perhaps I did something wrong. lets try it again.</i>

**Table 2** Outline of the analysis scheme used for the pilot experiment. Items marked with \* relate to the cognitive model of the domain.

---

**Hypothesis**

Variable types in relation\*

Relevance of relation\*

*Is there a relation between the variables chosen in the cognitive model?*

Relation type

Preciseness

Generality

Correctness\*

Relation with previous data analysis

**Experiment**

Relevant for testing hypothesis\*

*(changing one of the input variables in the hypothesis)*

Confirmatory/Disconfirmatory

*(Ability of experiment for giving confirmatory or disconfirmatory evidence)*

**Prediction**

Consistent with hypothesis and experiment

Confirmatory/Disconfirmatory

**Data analysis**

Result

Selection of relevant data

Conclusion correct?

---

mainly because they were entries on one sheet. The data-sheets give a nice global overview over the students' learning processes.

The data was further analyzed using an assessment scheme, based on the general framework for cognitive modelling that has been introduced in Van Joolingen and De Jong (1991). This assessment scheme makes explicit reference to the conceptual domain model, which has been developed for this purpose. The assessment scheme was revised several times during the analysis. In Table 2 a summary of this

assessment scheme is presented.

The assessment scheme gave as output a set of *learning indicators* for each hypothesis or experiment. Conclusions could be drawn on the basis of these indicators.

### 4.3 Results

The average duration of a session was 1h 40m. During this time the students filled in an average number of 6.0 forms for the structured group and 5.7 forms for the group using unstructured

**Table 3** Summary of the use of hypothesis scratchpads

Protocol	No. Forms	No. Hypotheses	Well-formed Hypotheses	Relevant relations	No. experiments	Relevant experiments	R.E. + Rel. Pred.	Irrelevant experiments
35-I <sup>1</sup>	7	7	0	0	7	0	0	6
41-I <sup>2</sup>	5	6	6	3	4	4	4	0
56-I	6	7	5	0 <sup>3</sup>	6	6	6	0
Total structured	18	20	11	3	17	10	10	6
35-II	5	5	0	0	4	2	1	2
41-II <sup>4</sup>	6	5	5	2	2	2	0	0
56-II	6	6	0	0	4	1	0	3
Total unstructured	17	16	5	2	10	5	1	5

Notes:

- <sup>1</sup> Subject 35-I did not seem to catch the idea of a relation. In 5 of his 7 "hypotheses" he only checked one variable.
- <sup>2</sup> Subject 41-I did not check any variable but all his hypotheses were well-formed, using the variable names from the sheet.
- <sup>3</sup> Subject 56-I was confused by the term "Result" on the sheet. She stated that the result was more "accurate", a relation that should be applied to the standard deviation (that becomes small). When the use of "Result" would be replaced by the standard deviation she would address an important relation three times.
- <sup>4</sup> Subject 41-II twice entered "No idea" in the hypotheses area. These two forms were not counted in this table.

scratchpads. In Table 3 a summary of the use of the scratchpads is given. Some effects are apparent<sup>1</sup>:

*Many hypotheses are not well-formed.*

Only 55% of the hypotheses on the structured scratchpads and 31% of the hypotheses on the unstructured ones were well formed, in the sense that they consisted of two or more variables and a relation that could be applied to those variables. Especially the formulation of a relation seems to be problematic.

*Often non-relevant relations were chosen.*

In only 3 of 11 (well-formed) hypotheses for the structured scratchpads and 2 of 5 for the unstructured scratchpads a variable pair was selected about which a relevant hypothesis could be formulated. In the other cases pairs were selected for which it was apparent (from the structure of the model) that there would be no relation. Of course a learner may have another view on this matter (and s/he should explore this view) but on the other hand one does not want the



learner to spend too much time in exploring such a relation.

*Often non-relevant experiments are designed.*

Table 3 shows that quite a large number of the experiments described on the experiment part of the forms were not relevant for testing the hypothesis, e.g. because the variables that were manipulated were not the variables about which the hypothesis was stated. Especially students using the non-structured forms often designed non-relevant experiments. Also noticeable is the fact that they also did not produce many relevant predictions.

The analysis of the datasheets gave some further results:

*Hypotheses are often stated at a global level, students tend not to investigate relations at a deeper level.*

The students seldomly showed behaviour which leads to a deeper investigation of a certain relation. The relations that were formulated were all on a general or qualitative descriptive level.

*No generalisations of hypotheses by the students are observed.*

Both the students working with the structured and those working with unstructured scratchpads did not try to generalize their hypotheses. Hypotheses were stated only about variables, not about variable types. For example an hypothesis about a specific partial error-total error relation was not generalised for all such relations.

*A confirmation of experimenter/theorist behaviour.*

The outlines of datasheets presented in Figure 2 show that the two strategies (theorist and experimenter strategies) as identified by Klahr and Dunbar (1988, see Section 2.2) also can be observed within this experiment. In this picture, boxes represent meaningful units, obtained from the filled-in scratchpads, the log-files or the think-aloud protocols. Arrows mark which boxes represent filled in-forms (Subject 41-II did not fill in the first three forms).

The fill patterns of the boxes indicate the relation in the conceptual model to which the represented activities refer. It can be read from the figure that subject 41-I first states his hypothesis about a relation and furthermore tests this hypothesis. Subject 41-II first carries out several experiments before stating his first hypotheses about a certain relation. (The fact that he didn't state predictions was due to the specific lay-out of the unstructured scratchpads, no conclusions should be drawn from this fact).

The preliminary findings of the pilot experiment were used to adjust the scratchpads and the computer program for the main experiments. The simulation was not changed essentially, apart from an automation of experiments to make the time needed for each experiment shorter.

The scratchpads were adjusted in a more dramatic way. In Section 5.1.1 the final results of this changes will be presented for the three versions used in the main experiment. Here the considerations for these changes will be discussed.

To assist students in forming well-formed hypotheses a relation selection

aid (as proposed in Section 2.4.1 was considered since it appeared that in many cases hypothesis were not well-formed because of a bad described relation.

To help students to select relations that were more relevant it was considered to be useful to supply extra information about the variables, so the variable selection list was commented with extra information about the meaning of the variables.

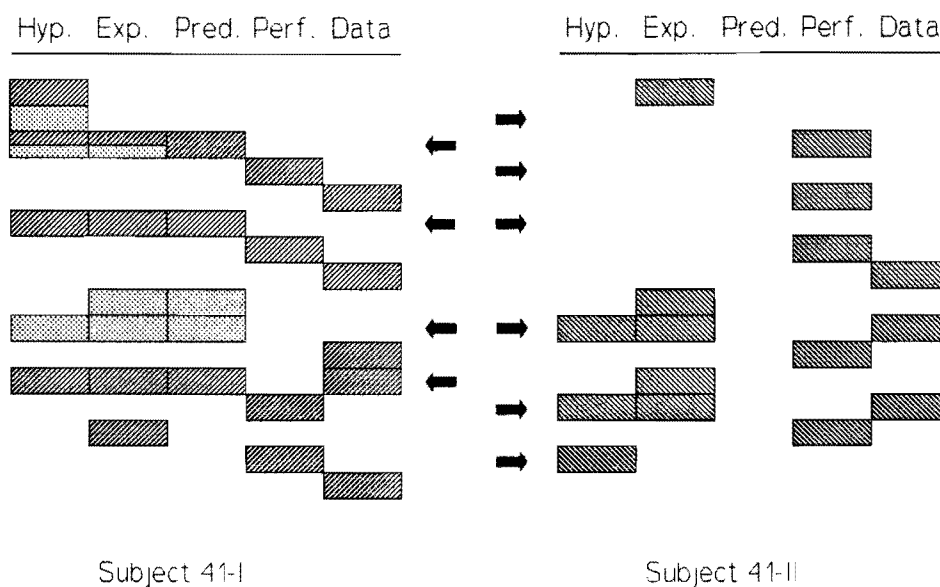
To support generalisation of the learner hypotheses it was considered to put variable types (as occurrent in the cognitive model) in the list of variables instead of variables in the runnable model. By offering both global and more precise relations in the relation aid support and stimulus was given to state hypotheses at a more precise level.

## 5 Main experiment

### 5.1 Method

The main experiment was performed with 31 first-year students of chemistry at the Eindhoven University of Technology. The students had received an introductory instruction in error analysis as part of their normal study program. The students were divided into three experimental groups of 10, 10 and 11 students. Each group received a different version of the hypothesis scratchpads.

For this experiment the scratchpads were split into an hypothesis scratchpad and an experiment/prediction scratchpad. The latter was the same for each group. The students were instructed to use one experiment form each time they were asked to (i.e. after the completion of each experiment) and to use an hypothesis form at the same time, if they had any idea to express on this form.



**Figure 2** Two outlines of datasheets (only the activities are depicted, not the contents) Sheet 41-I can be seen as theorist behaviour 41-II as experimenter behaviour.

**Table 4** Overview of the experimental design

	Relation support	Variable support	Experiment scratchpads	Posttest
I (N = 10)	•	•	•	•
II (N = 10)		•	•	•
III (N = 11)			•	•

After a short introduction about the experiment, the computer program and the use of the scratchpads, 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 posttest which consisted of one open question: "Describe on this sheet that 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 does."<sup>2</sup>

### 5.1.1 Hypothesis scratchpads

Three versions of the hypothesis scratchpads were developed, one for each group. These versions will be referred to as the structured, partially structured and unstructured scratchpads. Each one will be described in more detail in this section.

The *structured scratchpads* consisted of three tables (see Figure 3). Each table contained building blocks for the construction of an hypothesis. There was a table with variables, one with

conditions and one with relation prototypes. A student could construct a relation by selecting two or more variables, a relation prototype and (optionally) a condition to limit the scope of the relation and combining these building block into a complete relation.

The variables that were presented in the variable table were derived from the cognitive model of error analysis that was constructed. Actually the variables were variable *types* when viewed from the runnable model. The variables in the table are generalisations of the 'runnable' variables. Also some new variables, derivable from the other variables were introduced. The main role of these variables is to express general ideas about the model, such as the reliability of a measurement series, something that was not modelled in the runnable model.

The relations on the scratchpads were all relations present in the cognitive

**Table 5** Overview of the data collected in the main experiment.

Exp. Group	No. Experiments	No. Exp forms	No. Hyp. Forms
I	66	53	38
II	106	90	55
III	128	107	73

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 scratchpads* shared the variable table with the structured scratchpads but they 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 scratchpads* did not contain any information, neither on variables nor on relations. The sheets only contained an area in which the student could express his/her hypotheses.

The *experiment/prediction* scratchpads were unstructured. They contained two areas, one for the description of the experiment and one for the prediction. The students were instructed to describe the experiment by stating which variable(s) they were going to change, how they were going to change those variable(s) and which variable(s) they were going to observe. The prediction was described by stating the expected change of the observed variables. Table 4 gives an overview of the experimental design and the scratchpads used.

## 5.2 Data

The data that was collected consisted of the forms filled in by the students, the log-files of the interaction with the computer simulation and the results of the post-test. In Table 5 an overview is given of the data that has been collected. The forms

and log-files were analyzed, using an assessment scheme that is a refinement of the one used in the pilot study (See Table 2). The complete assessment scheme can be found in appendix I. The results of this analysis were collected in datasheets in a similar way as was done in the pilot (see Section 4.2).

Furthermore, the results of the post-test were analyzed. The number of relations that was treated in the essay was counted and each relation was matched against the cognitive model, by two independent domain experts (teachers from the chemistry department).

## 5.3 Results

The present section contains the results obtained from the experiment. Four aspects will be discussed: the activity level of the students (Section 5.3.1), a general functioning of the hypothesis scratchpads (Section 5.3.2), the mapping entered hypotheses onto the cognitive model (Section 5.3.3) and the consistency of the students' exploratory behaviour (Section 5.3.4).

### 5.3.1 Activity level

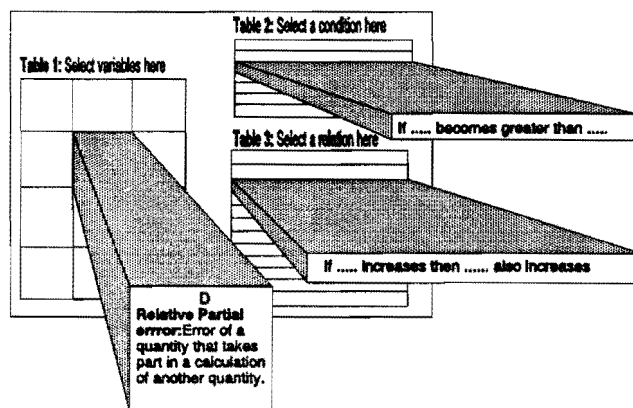
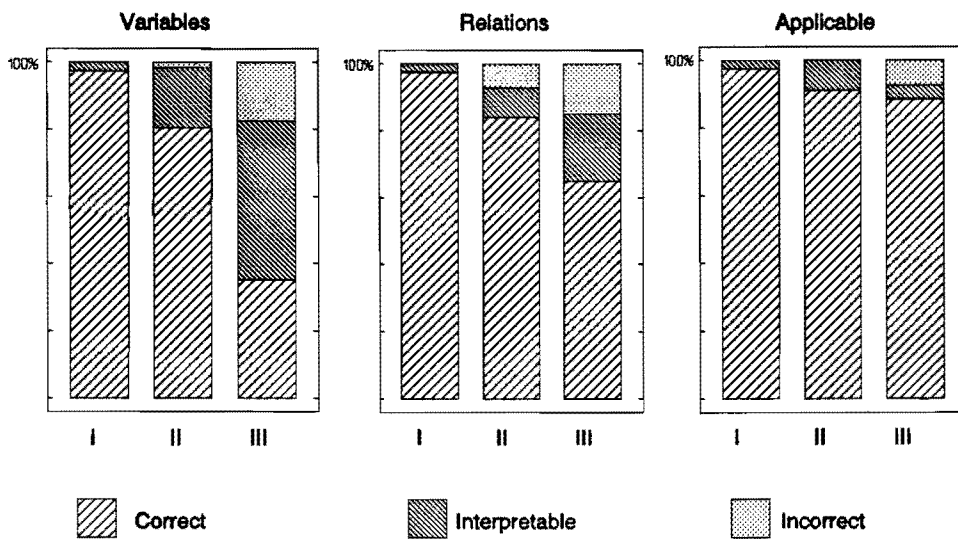


Figure 3 A structured hypothesis scratchpad as used in the experiment



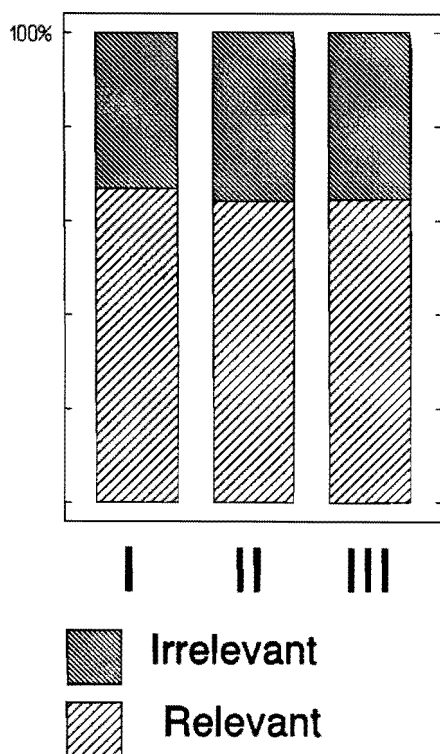
**Figure 4** Comparison of the relation syntax. From left to right the relative scores are drawn for the selection of variables, the selection of a relation and if the relation applicable.

One striking difference between the three experimental groups is the number of scratchpads that were filled in. For group I (using structured scratchpads) the average number of filled in forms is 3.9, for group II (using partially structured scratchpads) the average is 5.6 forms per student and for group III the average is 6.3 forms per student. The number of experiments that was performed per students also differed between groups with approximately the same ratios. This difference can (partly) be explained by the amount of information that was offered to the students. Apparently it took the students a significant amount of time to read and process this information. Also the conversion of the students ideas into the structure offered by the scratchpads may have taken time and effort by the students. If these assumptions prove to be true this effect of offering structured scratchpads should decrease when the simulation

would be used for a longer period and when structured scratchpads would be used more often in different learning environments.

### 5.3.2 General functioning of the hypothesis forms

A general evaluation of the function of the hypothesis scratchpads can be made by trying to answer the following two questions: "Do the structured hypothesis scratchpads indeed offer a framework which can easily be used by students to state their hypotheses and which really supports the generation of hypotheses." and: "Do the structured scratchpads not limit the students too much in their freedom of expression." These questions can be answered by analyzing the following aspects of the data obtained from the experiment:



**Figure 5** Overview of the relevancy of the experiments designed by the students (relative scores). The differences in mean between the groups are not significant (using a *t*-test).

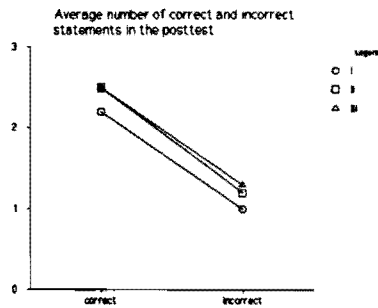
- (1) The (relative) number of well-formed hypotheses (i.e. hypotheses with a correct syntax).
- (2) The agreement between hypotheses and experiments (i.e. are the experiments designed in such a way that they can test the hypotheses).

The combination of these two aspects determines if the structured scratchpads were successful. If (1) is better for the structured scratchpads this can mean two things: or the scratchpad assists the student in forming hypotheses, or the scratchpad limits the freedom of the

student too much so that s/he is forced to formulated hypotheses that s/he doesn't understand. The agreement between experiments and hypotheses should discriminate between these alternatives: if the relation is better for structured scratchpads, these scratchpads help, if the relation is worse the structured scratchpads may limit the students too much in their expression.

Figure 4 displays the results of the syntax analysis. This figure shows that the students using the structured scratchpads (Group I) on average use a better syntax for their hypotheses than the other students. As such this is not remarkable 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 4 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 4.

Figure 5 shows the relation between hypotheses and experiments. Depicted is the relevancy of experiments, designed after the first statement of an hypothesis. An experiment is called relevant if it is capable of producing confirmatory or disconfirmatory evidence for the last hypothesis stated. The picture shows that about two-thirds of all experiments



**Figure 6** Results of the posttest, depicted is the average number of correct and incorrect hypotheses per student for the three experimental groups.

were well defined (there are no significant differences between groups). This leaves the case in favour or against the structured hypothesis scratchpads undecided for the moment. This chapter will produce some extra arguments fore and against structured hypothesis scratchpads.

Another general evaluation can be made on the basis of the posttest that was taken from the students. In Figure 6 the results of this test are depicted. The figure shows that the total number of statements per students is slightly lower in experimental group I but that there are no differences in the ratios between the numbers of correct and incorrect statements.

### 5.3.3 Assessment of hypotheses, related to the conceptual model

In this section a more detailed investigation of the hypotheses forms will be presented. Several aspects of the hypotheses that were stated by the students will be discussed. Of course the main issue will be a comparison between the

groups using the different structures on the hypotheses scratchpads, but also some more general issues will be discussed.

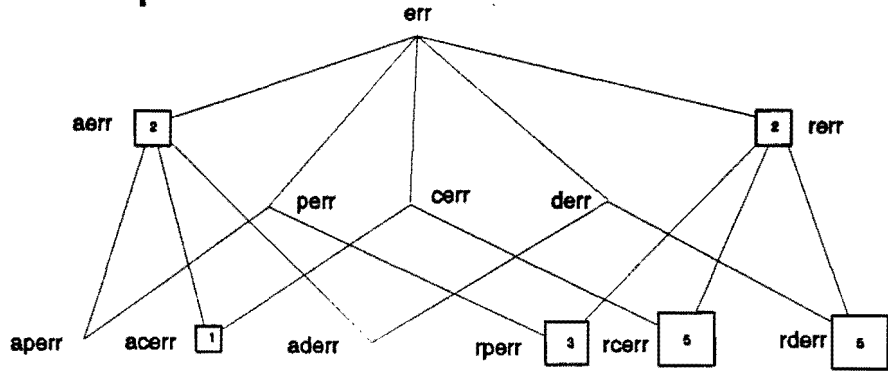
### Variables selected

A measure for the hypothesis space that students effectively use while exploring a simulation is the number of different variables they use to construct their hypotheses. Our expectation was that this number would be larger for the two groups of students who received scratchpads with a variable selection table. Especially it could be expected that the student would differentiate better between different kinds of measuring error, since this differentiation was indicated on their scratchpads.

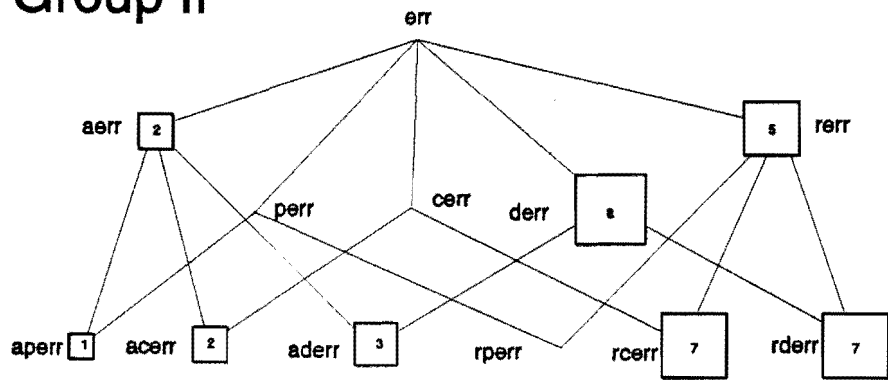
Figure 7 shows the use of the variables representing different kinds of measuring error as output variables in relations. The nodes in the networks represent variables. The lines depict subtype-supertype relations between variables. The top of the graph represents the most general type of error, going down the various kinds of error are differentiated. The names of the variables have been chosen systematically: they all end in 'err', the remaining letters represent the type of error: *absolute* or *relative* or the role they can play in the model: *partial*, calculated from other errors or obtained from measuring data. The squares represent the number of times a certain variable has been chosen by the students.

From Figure 7 it is clear that the variable support offered on the hypothesis scratchpads triggers the students to use more different variables in stating the hypothesis, in this case about measuring error. Effectively the students using the unstructured scratchpads use a

### Group I



### Group II



### Group III

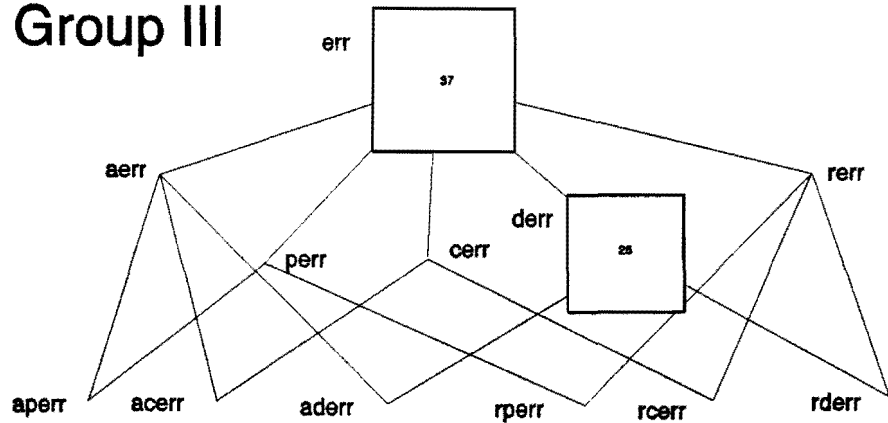


Figure 7 Use of the variables representing different kinds of measuring error.



smaller hypothesis space to state ideas.

#### *Relations selected*

In Table 6 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 which used the unstructured or partially structured ones. This contradicted our expectations. The offering of more precise relations did not stimulate the students to formulate their 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 did the presence of these relations on the sheets trigger the students to use it, in contrast to the more precise relations, but the students do not state these very global relations by themselves. Especially conditional relations were chosen almost never, despite the facts 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.

A conclusion that can be drawn from this finding is that for structured scratchpads there is a need to stimulate the use of more precise relations, e.g. by restricting possibilities of choice for the learner.

#### *5.3.4 Students' lines of reasoning*

The previous section presented an analysis of individual hypotheses. This sec-

tion will deal with the position of a specific hypothesis within the complete process of discovery. One aspect of this analysis, the relation between experiments and previous hypotheses has already been discussed in Section 5.3.2 (see Figure 5). The present section discusses the relation of hypotheses with previous ones, in order to discover logical lines of reasoning in the students' exploratory behaviour. Also the relations of hypotheses with preceding experiments is investigated in this section. This allows for a distinction between theorists and experimenters (Klahr & Dunbar, 1988, see Section 2.2)

#### *Matching with previous hypotheses*

To investigate if students followed a more or less consistent line of reasoning hypotheses were matched against their predecessor. It was investigated if students posed more than one hypothesis concerning the same set of variables or if they turned to other sets of variables. An '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 which has long hypothesis trains can be regarded as a systematic learner (provided that the experiments s/he performs are relevant for testing the hypotheses. Table 7 gives an overview of the average lengths of hypothesis trains for the three groups. It appears that there are no significant differences between groups but that for each group the average length is very small. Important is to determine if the last hypothesis in a

**Table 6** Selected relations classified according to level of preciseness.

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

The differences between group I and groups II and III for the first two categories are both significant for  $\alpha < 0.01$ , using a t-test. The differences between groups II and III are not significant.

train is correct, if compared with the conceptual model. This decides whether a line of investigation is concluded successfully or not. It appeared that about one third of the last hypotheses in an hypothesis train were incorrect, which seems to be quite a high percentage.

#### *Relation with previous experiments*

An important question is if the hypotheses that were entered had a basis in previously performed experiments or that they were elicited prior to experiments. This 'indicator' can serve as a discriminator for theorist or experimenter behaviour (Klahr & 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 this would be a real 'theorist event' since such an hypothesis is merely an statement of intention to investigate the mentioned relation. On the other hand a theorist event only has

some value if it is followed by an experiment which is able to put the hypothesis to the test. Table 8 shows that the group using structured scratchpads relatively more often stated hypotheses about variables with which they had not experimented before. 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. This is in concordance with the finding that students from groups I used a larger hypothesis space.

## 6 Conclusions

In this section some general conclusions will be drawn from the results presented in the previous section. Also some recommendations for the design of hypothesis scratchpads and indications for future research will be presented.

Conclusions regarding the hypothesis scratchpads should address two aspects of the use of the scratchpads: their usefulness as a supportive instrument and their usefulness as a supplier of information for the ISLE. Both aspects should be taken into account: a scratchpad providing more and/or better information to the ISLE contributes to the quality of that ISLE. In an indirect way such a scratchpad can increase the functionality of the ISLE.

### 6.1 The hypothesis scratchpads as supportive instruments

**Table 7** Average length of 'hypothesis trains' for the three groups, the differences are not significant for  $\alpha < 0.05$

	I	II	III
Total number of trains	35	46	56
Average size	1.1	1.2	1.3
Standard deviation	0.4	0.5	0.5
Maximum train length	3	3	3
Number of last incorrect hypothesis	13	15	17
thesis	(37%)	(33%)	(30%)

The basic assumption behind the introduction of a structure on the basis of the conceptual model of the simulation into the hypothesis scratchpads was that such a structure could familiarise the students with the hypothesis space and reveal its structure, so that the formation of hypotheses would be made more easy. The results of the experiment show that the scratchpads used partly worked this way but also that some aspects of the structured scratchpads did not have the expected result.

The number of different variables that was used was greater for the students who used structured or partially structured scratchpads. This implies that these students effectively used a larger hypothesis space than the students using unstructured scratchpads. This enlargement of hypothesis space was primarily the result of a differentiation between different types of measuring error in the variable selection table. This differentiation was not included in the runnable model. The introduction of relations (especially conditional relations) on the structured scratchpads, however, did not have the desired effect. The number of different relations that was chosen did not differ between groups and the students using structured scratchpads even

often selected very global relations, in contrast to the students who used a scratchpad which did not have a relation selection list.

The idea that the structured scratchpads would ease the process of hypothesis formation 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 well 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. When scratchpads would be used more often and when the simulation would be used for a longer period of time one may assume that the time needed to understand the structure of the scratchpad would relatively decrease. The scratchpads did help the students to form well-formed hypotheses: nearly all hypotheses formed by the students were well-formed, opposed to lower scores for the other groups.

Overall one may conclude that the structure offered on the hypothesis scratchpads did contribute to the structuring of hypothesis space by the students, but that the support was ineffective at some crucial points. This will result in some recommendations for future versions of the scratchpads, at the end of this chapter.

## 6.2 The scratchpads as source of information for the ISLE

To decide on the value of the scratchpads as source of information three aspects should be taken into account:

- Do the students enter information

## An hypothesis scratchpad for ISLEs

- in an interpretable format?
- Is the information entered on the hypotheses scratchpads representative for the 'real' learner hypotheses?
- Which kinds of information can the scratchpad provide for the ISLE?

Concerning the first question it is clear that the most structured scratchpads have an advantage to the two other types of scratchpads. The structured scratchpads provide a format of hypotheses which is easily interpretable by an ISLE. The language in which the hypotheses are entered is mappable to the language in which the domain representation can be expressed (See Van Joolingen & De Jong 1991). This means that there is no need for an interpreter of natural language, which would increase the complexity of the ISLE and decrease the reliability of the information. The fact that the syntax of the hypotheses entered was correct in most cases for the structured group indicates that the students did not have trouble with the syntactical part of entering hypotheses in the offered format.

It is more difficult to try to answer the second question. Of course one can never be sure that the relation a student enters is indeed the relation he has in mind as an hypothesis, but some indications may be useful. It is especially important to know how the correspondence is between the learners ideas and the entered hypotheses when the hypotheses are used for diagnosis purposes. There is an important difference between the case that the students enters an hypotheses which does not match the conceptual model because s/he did not understand the variables on the scratchpad (and therefore entered an hypothesis

**Table 8** *The number of times an hypothesis was first stated on the basis of previously obtained experimental data. The differences between Group I and the other groups are significant (t-test,  $\alpha < 0.01$ ).*

	I	II	III
Previous experiment?			
No	26 (67%)	21 (38%)	30 (45%)
Yes	13 (33%)	35 (62%)	36 (55%)

in terms of variables which do not correspond to his/her ideas) or that s/he has a different idea on the relation between variables. Both cases may result in the same relation entered in the hypothesis scratchpad and thus be diagnosed as the same misconception. There are two escapes possible out of this situation. The first is to provide extensive help on the variables that are offered; the second is to take more variables into account for the diagnosis, like information on manipulated variables.

Unfortunately the information as analyzed until this moment provides us with little information on the relations between the 'real' learner's ideas and the hypotheses that were entered on the scratchpads. A further analysis, especially directed at this questions should be made.

The type of information entered into the scratchpads can be used for learner diagnosis. Primarily the hypothesis scratchpads offer information about the learners ideas about the simulation, but also a sequence of filled-in hypothesis and experiment scratchpads can provide

information about certain learner attributes, e.g. if a learner is systematic or not. The latter type of information is richer than information on the same attribute that is generated on the basis of runnable model manipulations, because the scratchpads give information about the learners *conceptual* activities. Two physical actions on the model may seem unrelated, but may be part of some higher order research strategy. On the other hand, two manipulations of the same variable may be conceptually unrelated (e.g. because the learner is watching another output variable).

### 6.3 Analysis of the learner hypotheses

The assessment scheme that has been developed for this experiment provides an instrument to analyze the information entered by learners and derive the desired information from it. An implemented version of this assessment tool could therefore be a useful part of the ISLE. Of course the implementation would be the easiest if structured scratchpads would be used. The analysis scheme has a strong basis in the conceptual model, which is part of the domain representation for the ISLE. Therefore it will be necessary that the structure of the conceptual model is of a kind that will allow the analysis that is proposed by the assessment scheme.

The result of the assessment using the proposed scheme is a set of *learning indicators* which can be incorporated in the learner model. In order to allow for an implementation of the assessment scheme, a formal definition of these learning indicators should be available. This could be one of the goals of SIMULATE activity IV-2.

The assessment scheme also puts demands on the structure of the scratchpad. A scratchpad should be designed in such a way that it can output the learner hypothesis in the language used for the description of the cognitive model, i.e. in terms of variables and relations. In the present study this was solved by offering the conceptual model elements and relations on the scratchpads. Another solution could be to use a domain dependent graphical representation of cognitive model relations. A requirement on this representation would be that it can be translated into the cognitive model language. This solution has been chosen by Plötzner, Spada, Stumpf and Opwis (1990) in *DiBi* (see Section 2.4.3)

### 6.4 Recommendations for the design of hypothesis scratchpads to be used in ISLEs

A problem encountered with the simulation used is that students tend to keep on thinking in terms of the runnable variables instead of the more generic cognitive variables, despite the fact that the templates they have been offered contained only cognitive variables. A possible stimulus to generate more general hypotheses could be to open the possibility to inspect the cognitive variables to see which runnable variables are representants and to actually state the hypotheses in terms of these runnable variables. The rationale behind this is that it is better to have an explicit hypothesis about a runnable variable than to have an implicit one and state an hypothesis about a cognitive variable. Then, by limiting the hypothesis space (by forbidding or discouraging the choice of other variables) to the cogni-

tive variable involved (with the runnable representants of it), the students can be forced (or stimulated when more non-directive support is appropriate) to investigate related runnable variables and, in the end, arrive at a general rule for the cognitive variable.

This mechanism would require a variable inspection mechanism, i.e. a link between the cognitive and runnable models, visualised on screen. The latter could, for example, be realized by clicking on the variable name which brings up a submenu containing the runnable representants of the variable. In this way a *structured browsing* through all variables, cognitive and runnable, becomes possible.

Of course also the structure of the cognitive model must be adapted to allow this possibility, especially the relation between the cognitive and runnable models must be described in a way that the structure needed for the browsing process is well defined within the cognitive model.

We will term a scratchpad that can adapt itself to a changing situation (especially a changing learner model) can be called dynamical. Scratchpads that can limit hypotheses space, on the basis of previous experiments and hypotheses, are a special case of this type of scratchpads.

Dynamical scratchpads can also be used for stimulating the formation of more precise hypotheses. For example, when a learner has stated a very global relationship between two variables, and a more specific relation should be appropriate the scratchpad could limit the selection of relations to the more precise relations, if the student decides to state an hypothesis about the same or related variables.

A third use of dynamics in a scratchpad could be to direct the student to formulate hypotheses about the same variables as in previous hypotheses in order to stimulate consistent research behaviour. The learner should, in that case, explicitly decide to investigate the relation between other variables. That decision can be evaluated and the scratchpad could, for instance, decide to disallow this switching before a correct hypothesis on the first variables has been stated. Also it should be possible to *edit* an old hypothesis instead of stating a new one every time.

Possibly, the consistency of experimentation can also be improved by introducing *dynamical experiment scratchpads*. The experiment scratchpad (which should have a structure) can be used as a tool to structure experiment space. It will be clear that such a scratchpad can be utilised to limit experiment space in order to stimulate the learner to experiment with some specific variables. This dynamics can be used to establish a strong link between the hypothesis and experiment scratchpads.

Another extension of hypothesis scratchpads is to draw the attention of the learner to specific elements on it. For instance when a learner selects two variables on the scratchpad, a third, which is part of a conditional relation with the two selected can be highlighted. Such constructs can possibly ease the construction of more complex relations.

Finally the scratchpads can also be used to present different *views* on the domain. White and Frederiksen (1988) have developed a prototype in which different views of are presented to the learner. In their case these views are of

increasing complexity, but it is also possible that different views are introduced to show alternative conceptualisations of the domain (Van Joolingen & De Jong, 1991) Dedicated scratchpads can be used to express these views. Each view corresponds to a set of (cognitive) variables and relations, which can be represented on the scratchpad.

The above considerations show an new advantage of the use of structured scratchpads: the structure can serve as a basis for a dynamical interaction with the student. It would be far more difficult to implement dynamics of the kinds presented above in partially structured or unstructured scratchpads. Of course the proposed improvements will have to be tested experimentally.

What also becomes clear from the above considerations is that the design of the scratchpads should have a firm basis in the conceptual model of the domain. The dynamics, structures and views need to be derived from the domain representation. Also the analysis of the entered hypotheses should take place on the basis of the conceptual model. More research is needed to determine the exact links and requirements on the domain representation.

Currently a start has been made at EUT with the implementation of scratchpads to be used on line with the computer simulation. This allows for research to the properties of dynamical scratchpads.

## 7 Acknowledgement

The authors would like to thank Dr. J.C. Reijenga and Dr. M.C.A. Donkersloot of the department of

chemical engineering, for their cooperation with this research.

## 8 References

- Alessi S.M., & Trollip, S.R. (1985). *Computer based instruction, methods and development*. Englewood Cliffs, NY: Prentice-Hall.
- Goodyear, P., Njoo, M.K.H., Hijne, H., & Van Berkum, J.J.A. (1990). Learning processes, Learner attributes and Simulations in: Jong de, T. (ed.) (1990). *Computer Simulations in an Instructional Context*. SAFE project report SIM/9 (SAFE/SIM/WPI/EUT-rep/comp.final).
- Gorman, M.E., & 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., & 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.
- Joolingen, W.R. van, & Jong, T. de, (1991a). Characteristics of simulations for instructional settings. *Education and computing* 6, 241-262.
- Joolingen, W.R. van, & Jong, T. de, (1991b). An instruction related domain representation for simulations. In: De Jong, T. & Tait, K. (Eds.): *Toward formalising the components of an intelligent simulation learning environment*. Report SAFE/SIM/WPI-II/EUT-

- rep/formal.upd1. Delta project P7061/D1014. Eindhoven: Eindhoven University of Technology.
- Van Joolingen, W.R. and De Jong, T. (1991c). *Modelling domain knowledge for intelligent simulation learning environments*. OCTO report 91/02. Eindhoven: OCTO, Eindhoven university of technology.
- Jong de, T., & Njoo, M. (1990). *Learning and instruction with computer simulations: learning processes involved*. Paper presented at the NATO workshop "Computer-based learning environments and problem solving", Leuven (Belgium), september 26-28.
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12, 1-48.
- Lavoie, D.R., & Good, R. (1988). The nature and use of predictions skills in a biological computer simulation. *Journal of Research in Science Teaching*, 25(5), 335-360.
- Michael, J.A., Haque, M.M., Rovick, A.A., & Evens, M. (1989). The pathophysiology tutor: a first step towards a smart tutor. In H. Maurer (Ed.), *Computer Assisted Learning. Proceedings of the 2<sup>nd</sup> International Conference ICCAL* (pp. 390-400). Berlin: Springer Verlag.
- Min, F.B.M. (1987). *Computersimulaties als leermiddel: Een inleiding in methoden en technieken*. Schoonhoven: Academic Service.
- 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.
- Mynatt, C.R., Doherty, M.E., & Tweney R.D. (1978). Consequences of confirmation and disconfirmation in a simulated research environment. *Quarterly Journal of experimental Psychology*, 30, 395-406.
- Njoo, M., & Jong de, T. (1989). *Learning processes of students working with a computer simulation in mechanical engineering*. Paper presented at the Third European Conference for Research on Learning and Instruction, EARLI, Madrid, Spain.
- Plötzner, R., Spada, H., Stumpf, M., & Opwis, K. (1990). *Learning qualitative reasoning in a micro-world for elastic impacts*. Research group on cognitive systems, Forschungsbericht nr. 59. University of Freiburg: Psychological Institute.
- Reigeluth, C.M., & Schwartz, E. (1989). An instructional theory for the design of computer-based simulations. *Journal of Computer-Based Instruction*, 16(1), 1-10.
- Shrager, J., & Klahr, D. (1986). Instructionless learning about a complex device. *International Journal of Man-Machine Studies*, 25, 153-189.
- Shute, V., & Bonar, J. (1986). Intelligent tutoring systems for scientific inquiry skills. Unpublished manuscript.
- Shute, V., & Glaser, R. (1990). A large scale Evaluation of an Intelligent Discovery World: Smithtown. *Interactive Learning Environments 1*, 59-77.
- Shute, V., & Glaser, R. (in press) An intelligent tutoring system for exploring principles of economics. In: R. Snow & D. Wiley (Eds.) *Improving inquiry in social science: A volume in honor of Lee J. Cronbach*. Hillsdale, NJ: Erlbaum.



- Shute, V., Glaser, R., & 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., & 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., On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of experimental Psychology*, 12, 129-140.
- Zeigler, B.P. (1976). *Theory of Modelling and Simulation*. New York: John Wiley and Sons.

**Notes:**

1. Due to the small number of students it will be clear that none of the results is expected to be significant. Therefore no statistical tests were applied.
2. Of course, this question was stated in Dutch, this is a literary translation.

## APPENDIX Analysis scheme for hypotheses forms

### 1 Well-formedness

#### 1.1 *Variables selected*

Are the variables that are mentioned well defined? Do they occur in the Cognitive or runnable model?

#### 1.2 *Relation selected*

Is the relation well described?

#### 1.3 *Relation applicable*

Is the relation that is specified applicable to the variables? E.g. a monotonic relation is not applicable to a not-ordered set of values.

### 2 Description of hypothesized relation using C.M. and relation typology

A description of the relation that is hypothesized in terms of the variables and relation frames that are defined in the cognitive model.

### 3 Relevance of relation

A relation can be more or less relevant in the cognitive model. Generally speaking are relations that are represented directly the most relevant.

#### 3.1 *Is there a relation between the selected variables in the C.M.?*

If there is no relation present in the C.M., neither direct, nor indirect then the relevancy is at the lowest

### 4 Complexity of relation and position in the C.M.

The complexity of a relation is dependent of two things: The structure of the cognitive model: is there a direct relationship and the number of variables involved.

#### 4.1 *Direct relationship in C.M.*

#### 4.2 *Indirect relationship (no. of levels)*

#### 4.3 *Number of variables*

### 5 Preciseness

#### 5.1 *Very general*

E.g. A and B are related.

#### 5.2 *Qualitative, general descriptive*

If A changes, B changes

#### 5.3 *Qualitative, specific*

E.g. If A increases B also increases

#### 5.4 *Quantitative, descriptive*

E.g. There is a linear relation between A and B

#### 5.5 *Quantitative, equation*

E.g.  $A = 2 \cdot B$

### 6 Generality

Roughly speaking, the generality can be measured by the number of 'instances' to which the relation is applied in the hypothesis. Three levels are distinguished.  $\infty$  for a very general statement, S for a subset of all possible cases and 1 for a single instance.

#### 6.1 *Applied to all instances*

#### 6.2 *Applied to a subset*

#### 6.3 *Applied to a single instance*

## 7 Matching with the C.M.

### 7.1 *Completely correct*

The hypothesis is completely correct which means that it is true in all cases in the cognitive model. This does not mean that the relation that is stated occurs as such in the cognitive model, the C.M. may contain a more precise relation. Mark with C.

### 7.2 *Too general*

The relation stated is true in some cases but not always. Causes: A condition has been omitted (G(C)), a variable type in the relation should be replaced with a subtype (G(V)).

### 7.3 *Too specific*

A relation has been selected which is too precise. E.g. in the C.M. a relation is only specified as monotonic while the hypothesis specifies a precise mathematical relationship. Mark with S.

### 7.4 *General case applied to wrong instance*

A relation which is true in general has been applied to an exceptional case. Mark with W

### 7.5 *Not correct*

All other cases

## 8 Relation with previous hypotheses

If a previous hypothesis is about the same or related variables then the two hypotheses can be compared

### 8.1 *Matching: compatible or incompatible*

(M)+ or -

### 8.2 *More/Less complex*

(C)+, = or -

### 8.3 *More/Less general*

(G)+, = or -

### 8.4 *More/Less precise*

(P)+, = or -

## 9 Relation with previous experiments

### 9.1 *Compatibility (the hypothesis explains an experiment)*

(C) Mark with C if the hypothesis is compatible with a previous experiment, with N if it is incompatible, with - if there is no relation.

### 9.2 *Hypothesis changed on basis of experiment*

(H) If an experiment has been performed to test a previous hypothesis and this experiment resulted in a contradiction of the hypothesis then mark with '+' if the hypothesis changed, '-' if it didn't change. leave a space if the condition was not met.

## Analysis of experiments

- A Well-formedness
  - A.a *Input Variables*
  - A.b *Output variables*
  - A.c *Manipulation*
- B Input variables
  - name(s) of input variables
- C Output variables
  - name(s) of output variables
- D Control
  - D.a *Direct control*
  - D.b *Indirect control*
- E Manipulation
  - E.a *Preciseness*
    - 1 : very general; 2: descriptive; 3 quantitative
  - E.b *Manipulation type*
    - description, e.g. inc(rease), dec(rease), c(hange), \*2, etc
- F Relation with hypothesis
  - F.a *Capable for confirming evidence*
  - F.b *Capable for disconfirming evidence*
- G Well-formed prediction
  - G.a *Variable*
  - G.b *Change*
- H Change
  - H.a *Preciseness*
    - See E.a
  - H.b *Change type*
    - See E.b
- I Relation with hypothesis
  - I.a *Compatible*