DOCUMENT RESUME

ED 314 008 IR 013 922

AUTHOR	Kamsteeg, Paul A.; Bierman, Dick J.
TITLE	Cognitive ATI Research: A Simulated Laboratory Environment in (PCE-)Prolog.
PUB DATE	Mar 89
NOTE	19p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 25-30, 1989).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Aptitude Treatment Interaction; Authoring Aids (Programing); *Cognitive Processes; *Computer Assisted Instruction; *Computer Simulation; *Courseware; Foreign Countries; Higher Education; Instructional Design; *Intermode Differences; Physical Sciences; Undergraduate Students
IDENTIFIERS	*Intelligent Tutoring Systems; Learner Control; Netherlands; PROLOG Programing Language

ABSTRACT

A study of 228 undergraduate psychology students examined the effectiveness of a prototype of "heatlab," a laboratory simulation written in PCE-PROLOG, intended for remedying misconceptions of the concepts "heat" and "temperature." The effect of varying the amount of structure on students' understanding and the possibility of an interaction between amount of structure and the students' negative fear of failure were also examined. Quantitative data (using pre-, post-, and retention tests) indicated that understanding of the concepts had been increased by "heatlab." The amount of structuring was shown not to make a significant difference, neither as a main effect nor in interaction with fear of failure. In a socratic remedy of a misconception, the structure imposed upon the discovery process is aimed at inducing a paradox which forces the student to re-evaluate his/her beliefs. Students' reactions to paradoxes were examined, with a focus on the circumstances under which surprise reactions occur and the effect of varying the amount of structure on these reactions. Qualitative data (analysis of a think-aloud protocol) indicated an interaction between fear of failure and structuring in that there is no socratic learning for high fear of failure subjects in the unstructured condition. (GL)

Cognitive ATI research: A simulated Laboratory Environment in (PCE-)Prolog[1]

paper to be presented at
AERA conference, San Francisco, march 25-30, 1989.

Paul A. Kamsteeg (1) & Dick J. Bierman (2)[2]
(1) Centre for Educational Research
(2) Psychology Faculty, Psychonomics Dept.
University of Amsterdam

Abstract

In this paper we describe a microworld environment simulating a laboratory in which a pupil can perform experiments relating to the concepts of 'heat' and 'temperature'. We discuss its appearance to the pupil, and the intended use of a range of similar simulations both in the educational context of a computer coach for thermodynamics and in a series of ATI-type experiments in which the quantitative ATI method is complemented by a qualitative cognitive method. We describe a first experiment, using two versions of the implemented simulation environment, in which the quantitative data did not indicate an ATI-effect but the qualitative data supported our (ATI) expectation. We discuss these results and their possible consequences for tutoring.

Introduction

One of the research clusters at the psychonomics department of our psychology faculty is called 'Knowledge acquisition in formal domains'. Research in this cluster is aimed at how people (learn to) solve problems in domains like arithmetic, physics, etc. Often think-aloud protocol analysis is used as a research method.

---

[1] The research described in this paper is partly funded by the dutch Foundation for Educational Research SVO.

[2] Authors address: Psychologisch Laboratorium UvA, vakgroep Psychonomie, Weesperplein 8, 1018 XA Amsterdam, The Netherlands.

For one of the knowledge domains studied in this research cluster, simple gas thermodynamics, a computerized semi-automatic protocol diagnosis tool (called PDP) was devised some ten years ago. Part of this tool was an expert system which could solve simple thermodynamics problems, another part was a tracer mechanism by which it could express its reasoning steps. These parts were further developed into a model of expert problem solving in thermodynamics during the following years (Jansweijer et al., 1982; Jansweijer, 1988).

In 1984, the project "A computer coach for thermodynamics' was started. Its goal was to build a prototype ITS for the procedural aspects of solving thermodynamics problems, using the expert system PDP. One of the difficulties arising during this project was that the literature on educational research lacked any theories detailed enough to be used to devise a set of tutorial strategies for the computer coach. Therefore, a technique was devised to study the strategies that actual teachers used in one-to-one tutoring, without disturbing the tutoring dialogue. This technique made extensive use of the different development stages of the computer coach, as well as contributing to the knowledge to be integrated into the next development stage. This technique, known to us as MUSPA (for MUltiple Source Protocol Analysis), is described elsewhere (e.g. Bierman & Kamsteeg, 1987).

All in all, however, the amount of knowledge (tutorial strategies, diagnostic techniques) we elicited from the teachers was disappointing. It seemed clear that even experienced teachers had too little insight into how pupils actually learned. Thus, we decided to 'go back to the roots', as it were, and focus on pupils learning instead of teachers teaching. But our ultimate goal in this has stayed the same: gathering knowledge about the teaching/learning process at a level detailed enough to be used in an ITS system.

The work described in this paper is aimed at getting insight into the way pupils learn to overcome incorrect (pre)conceptions about a knowledge domain by doing experiments in a laboratory, i.e. by seeing how things really are as opposed to how the pupil thinks they are.

A simulated laboratory

The pupil's viewpoint

A simulation environment is not the real world. But, at least in an educational context, it is intended to teach about certain aspects (concepts and relations) of the real world. To this end, as well as for practical reasons, a simulation environment is limited in scope, in force and in complexity.

As for scope, a simulation environment only covers a small part of the world in space, time and types of objects. Only that much of the world as is sufficient to teach a certain domain of interest is portrayed, moreover a pupil can only perform actions which are relevant for the domain: it is in most cases completely useless to place at a pupil's disposal a simulated sledge-hammer to simulate smashing up the simulated environment.

The above example also relates to force. The force of a simulation environment is limited in that a pupil can not perform really destructive actions, be it intentionally or by mistake (e.g. shorting an amplifier). Even if the simulation reacts by 'breaking down' (which a good simulation should almost never do) the program can always be restarted, and nothing has happened. But more importantly, a simulation environment is equally unable to physically harm a pupil. Even blowing a simulated nuclear power plant leaves a pupil with nothing damaged but his/her trust in nuclear energy. A final matter pertaining to force is the physical strength a pupil needs, to perform certain certain actions which in reality would require considerable power. In a simulation environment these actions can be performed virtually without effort. In short, a simulation is watered-down in relation to reality.

Pertaining to complexity, a simulation environment is more simple than a real one. This is not only because of the aforementioned limitation in scope, but also a simulation is an abstraction. Relations are straightforward and consistent, irrelevant complicating aspects and exceptions are ignored, hidden variables may be exposed, measurements are easy. This is of course in line with the use of simulation environments as a teaching aid.

However, in constructing a simulation environment, one must be careful not to limit and abstract too much. Through interaction with the simulation, a pupil should gain insight into the relevant aspects of the target domain in reality, not only in simulation. Therefore, a pupil should be led to view the simulation as a metaphor· he/she should be aware of its relevancy to reality as well as its incomplete reproduction thereof. This is a task for the teacher using the simulation as well as (or maybe even more than) for the simulation itself.

What the type of simulation we have in mind looks like from the viewpoint of the user (pupil) is best described by using as an example the existing prototype used in the experiments reported further on: the 'heatlab' (see fig. 1).

(insert fig. 1 approx. here)

fig. 1: *user interface of the 'heatlab'*

The computer screen contains three windows. The biggest one, the simulated laboratory proper, is only visible when the pupil is actually required to perform experiments. If not, only a text window is visible. In this text window, questions are asked (and answered by the pupil), experiments are prompted and possibly described, etc. In short, a tutorial dialogue is conducted in this window. The third window (overlapping the text window) is, like the laboratory window, initially invisible. It can be made visible by the pupil using a button in the laboratory window (i.e. only when experimenting). It itself contains a button to make it invisible again, thus permitting the text window to show again. In this third window (the log book) the pupil can order measurements to be automatically recorded.

In the actual laboratory window, there are four types of objects.
First, there is a series of manipulatible objects on which to perform the experiments. In the case of the 'heatlab', these are blocks of different materials having different weights. They can be moved, stacked and unstacked, and their relevant properties (in the 'heatlab': their temperature) can be measured by attaching measuring devices.
Second, there are manipulating agencies which may be used to 'do things' to the manipulatible objects. In the 'heatlab' there is a bunsen burner by which heat can be added, and a thermostat room in which a temperature can be preset.
Third, we have the controls for the manipulating agencies. E.g. temperature control for the thermostat room, timer and flame-height for the bunsen burner. Also in this category are buttons for stopping the experimentation, starting afresh, and automatically taking measurements.
Fourth, measuring devices. In the 'heatlab' there are two types. Attached to the bunsen burner is a 'heat meter' measuring the amount of energy given off by the burner. Furthermore, the pupil can create thermometers which may be attached to a block of material, measuring its temperature.

Experiments are done in two stages. First, an experimental set-up is built by connecting objects: blocks to each other and/or to the bunsen burner (possibly after having given them an initial temperature in the thermostat room), thermometers to blocks. Also part of the set-up is setting the manipulating agencies at their acquired values. The second stage consists of performing the actual manipulation (adding heat, connecting -stacks of- objects) and taking measurements.

Thus, in short, experiments are performed by connecting and disconnecting, activating and de-activating objects.

The programmer's viewpoint

The 'heatlab' is, and intended future laboratory simulations will be, written in PCE-Prolog. This is an object-oriented graphical extension to the logical programming language Prolog (Anjewierden, 1986; Anjewierden & Wielemaker, 1988). Actually, it runs as two separate processes within a Unix operating system on a Sun work station. The two processes, the PCE process and the Prolog process, exchange messages through a 'pipeline' but are, apart from that, completely independent. In fact, PCE can work in conjunction with programs in any language.

For the Prolog process, the PCE process and all of the graphical manipulations and administration are hidden except for three added predicates: *new* (which has as a side-effect a message to PCE to create an object), *send* (side-effect: a message to an object within PCE), and *get* (which instantiates variable arguments with acquired aspect-values of an object within PCE). One of the PCE-objects at which a *get* may be directed is the 'queue', a list of messages representing user actions. PCE updates this list as the user manipulates objects on the screen. By regularly polling the queue, a Prolog program may be kept informed about user actions, but the program may as well decide to ignore the queue temporarily or even continually (although the latter is not so smart), flush the old queue, etc.

We will not get into more detail about PCE-Prolog and the actual implementation of the 'heatlab' simulation here. These are described more extensively in Kamsteeg & Bierman (1989).

Intended use of (the family of) this simulation

Intelligent coaching and LOGO-type discovery

Simulation environments have so far been used mainly in the LOGO approach to education. LOGO, apart from being a simple and child-oriented programming language, has from its inception also been intended as an educational method (Papert, 1980). This method is rooted strongly in the 'discovery learning' philosophy which was put forward from the early sixties by educational researchers like Bruner (1961). The goal of a simulation environment in this tradition is to provide a pseudo-world which a pupil can freely explore for consequences of various actions, thereby gaining insight into the laws which govern this pseudo-world (and, presumably, the corresponding part of reality).

In our view, this approach to simulations poses a couple of problems. First, to induce a pupil to meaningful exploration, a simulation environment must be inherently motivating (DiSessa, 1986). This seems to be difficult to achieve for every pupil, especially in certain less spectacular domains. Second, apart from being motivated, a pupil must also use a method of systematically varying all relevant aspects in order to gain any real insight into the domain. But not every pupil will spontaneously use such a method. Third, it is not always that easy for a novice to discern the relevant aspects within a domain. Some *a priori* knowledge of the domain (by prior instruction, experience, or possibly intuitively) seems often to be needed.

Therefore, we think that a more guided form of discovery learning will yield better results of using a simulation environment. In practice, even in the LOGO approach, guidance is usually provided in the form of explanation, suggestions etc., either by a textbook (e.g. Abelson & DiSessa, 1980), a teacher, or both.

The Intelligent Tutoring Systems research community has so far had little interaction with the LOGO community. In existing ITS's, little if any simulation is incorporated. More importantly, the 'free discovery' philosophy is diametrically opposed to the viewpoint underlying ITS's, which calls for fairly strict monitoring and guidance of a pupil solving problems in a certain domain. But a form of *strictly guided* discovery, namely guided self-remedy of misconceptions in the form of a Socratic Dialogue, does appear in the ITS literature (e.g. Collins & Stevens, 1980). The Socratic Dialogue technique normally uses thought experiments to falsify logically derived consequences of a pupils misconception, and thereby (hopefully) the misconception itself. In domains like physics, however, it seems probable that actually performing a (real or simulated) experiment is of more value.

Educational use of this simulation

As the previous section already suggested, we will use our laboratory simulation, and ones similar to it, to remedy misconceptions in what one might call a 'see for yourself' way. How strongly guided such a discovery-like misconception treatment should be, is a question which shall be discussed further on.

In the long run, we intend to integrate these laboratory simulations into the ITS for coaching simple thermodynamics problem solving that was mentioned in the introduction (the 'computer coach'). Pupils who are diagnosed by the ITS to have a certain misconception may be directed to the simulated laboratory, in which they are required to answer a series of questions by performing experiments. The ITS should monitor the pupils behaviour and decide which questions to ask and which feedback to give. Another use of the laboratory simulation in this context is, to let the pupil check his/her solution of a problem by actually carrying out the problem as an experiment. Here also, the ITS should direct and monitor the pupils actions.

Experimental use of this simulation

Apart from (and before) being employed as an educational tool, e.g. in the context of an ITS, a prototype laboratory simulation can be used in experiments to get insight into different aspects of discovery learning and misconception treatment. By performing analyses of think-aloud protocols from pupils working with the laboratory simulation, we try to find out more about the process underlying the formation and alteration of mental models about a domain, i.e. what exactly happens as a pupil is exploring a domain or is confronted with events that do not fit in with his/her conceptions of the domain.

Furthermore, we intend to study what structure of a laboratory simulation (e.g. how much guidance) works best, and how this interacts with characteristics of pupils. This, of course, is a type of Aptitude-Treatment Interaction research. Out hope is, that by automating (uniforming) the treatment to a great extent and thereby lessening error variance, we will be able to show ATI effects, which are known to be small, better than with traditional methods.

A first experiment using the 'heatlab'

The questions

Our first experimental question arises both from our interest in more or less guided forms of LOGO-type environments, and from the ATI-literature (e.g. Cronbach & Snow, 1977; Entwistle, 1981). Much of the aptitudes actually interacting with treatments in ATI-research seem to be related to the personality construct 'negative fear of failure' (Hermans et al., 1972). The treatment effects these aptitudes are interacting with usually have to do with guidance, structuring or security in the learning task. Moreover, the interaction between negative fear of failure and task structuring can be interpreted in theoretical terms, a requirement which a.o. Simons (1980) imposes upon useful ATI-research. This theoretical interpretation would be that pupils with high (negative) fear of failure tend to perform better in situations where they can proceed step by step, always knowing what to do next, performing, as it were, a series of small tasks in which they can not easily fail.

So, the first research question for this experiment is: Given a number of pupils who have shown misapprehension of the concepts of 'heat' and 'temperature', does performing experiments in the 'heatlab' result in better understanding of these concepts, does the amount of structure (guidance) provided during work in the 'heatlab' differentiate in this understanding, and is there an interaction between amount of structure and the pupils negative fear of failure in the effect on this understanding?

The other research question is a qualitative one, intended to be answered by think-aloud protocol analysis. In a socratic remedy of a misconception, the structure or guidance imposed upon the discovery process is aimed at inducing a paradox which forces the pupil to re-evaluate his/her beliefs. Such a paradox should cause surprise and disbelief on the part of the pupil. But also in free exploration, we expect a pupil only to alter his/her conceptions after an unforeseen and surprising event. In the latter case, however, these surprising events will take place less often since there is no structuring specifically aimed at them and the pupil will encounter them only 'by accident'.

Our second research question then is: Can we find utterances of surprise and disbelief in the think aloud protocols, if so, at what points and in what circumstances do they appear, and are they more frequent when more structure (guidance) is provided during work in the 'heatlab'?

The design

There are two experimental and one control conditions. Subjects in the experimental conditions follow a structured and an unstructured version, respectively, of a lesson using 'heatlab'. Subjects in the control condition spend an equal amount of time doing a computer-game. Directly afterwards, all subjects fill in a post test intended to measure insight in the concepts of heat and temperature. A similar retention test is filled in three weeks later.

Subjects are selected for the experiment on the basis of performing poorly on a pre test some months before the experiment. They are tested and matched for intelligence and negative fear of failure, then each matched group is distributed randomly over the three conditions.

For five (randomly chosen) subjects in each experimental condition, think aloud protocols are recorded.

Quantitative data analysis is performed by multiple regression analysis of post test and retention test scores against condition and fear-of-failure scores (Kerlinger & Pedhazur, 1973; Cohen, 1983). Protocol data are qualitatively analysed.

The experimental procedure

228 Unselected freshman full-time psychology students were given a test consisting of 27 correct or incorrect statements about heat and temperature (e.g. "temperature is a measure of heat") to be labeled correct or incorrect, as well as 2 descriptions of experiments asking for a qualitative prediction of the outcome. This test was administered along with a test for negative fear of failure and a series of Guilford intelligence tests.

From these subjects, 48 rating in the lowest 50% on the 'heat-test' were included in the experiment, which took place about half a year later. They were matched for mean score on the Guilford tests and for fear-of-failure score, then randomly assigned to conditions (16 subjects in each condition).

Subjects in the 'structured' experimental condition were given a socratic-type question sequence about 6 different aspects of the heat/temperature relation: they were asked to predict the outcome of various experiments, then to perform these experiments, which were described in detail. In the 'unstructured' condition, pupils were merely asked, for each aspect of the heat/temperature relation, to think of a way to explore this aspect and carry it out; this condition was intended to reflect the LOGO-type free discovery approach. In the control condition, subjects played an adventure-type computer game which was far too difficult to be completed in a matter of hours. They were told beforehand that this game might contain 'things having to do with heat and temperature' (in fact it did not). Total time on task was about 90 minutes in all conditions; when necessary[3] the experimentator broke off exploration of an aspect and urged the subject to continue with the next aspect. The game in the control condition was simply stopped after 90 minutes.

Five randomly selected subjects in each of the experimental conditions were asked to think aloud while working with the

---

[3] There were six aspects to be explored, so each aspect should take about 15 minutes. Exploration of an aspect was broken off when a subject was getting more than 25% behind this 'schedule'.

'heatlab'. This thinking aloud was taped for later transcription on paper.

Directly following the lesson (or the game) subjects gave their opinion on the 27 statements of the original (selection) test. Three weeks later they came back to do a retention test consisting of 27 very similar statements (much of them being the opposite, or a rephrased version, of a statement in the original test).[4]

The results

Quantitative data

An overview of variable means and variances is given in fig. 2. Also the correlation between pre test and post test and between post test and retention test are portrayed there.

	control	structured	unstructured
Fear of Failure			
mean	11 44	11.50	11.63
variance	20.53	22.93	20.38
Intelligence			
mean	16.75	17.34	16.46
variance	06.15	12.87	05.93
Pre test			
mean	35.88	36.06	36.63
variance	20.25	14.20	22.92
Post test			
mean	59.19	69.69	69.31
variance	45.76	64.64	11 17
Retention test			
mean	60.60	66.13	67.94
variance	21.97	74.65	36.33
Correlations			
pre - post	.233	.051	.005
pre - retention	-.004	.049	.152
post-retention	-.403	.322	.071

fig. 2: *overview of means, variances and correlations*

---

[4] One subject in the control condition failed to show up for the retention test and could not be reached, bringing the number of retention test scores in the control condition to 15.

Multiple regression analysis was performed separately for post test scores and for retention test scores as dependent variables. In both analyses the predictors were the fear-of-failure (FoF) scores, two orthogonal dummy variables representing ***A*** the two experimental conditions vs. the control condition, and ***B*** the 'structured' condition vs. the 'unstructured' condition with the weight of the control condition nullified; further predictors were two multiplication factors FoF x A and FoF x B, representing interaction effects. This technique is described in Kerlinger & Pedhazur (1973).

For post test scores as dependent variable, there appeared to be virtually no interaction effect, as measured by the gain in explained variance of the dependent variable when adding the interaction factors (F=.103, p>>.1). This permitted us to analyse the main effects in isolation. The factor fear-of-failure did not contribute at all to the variance of the dependent variable (F=0!). The factor B ('structured' vs. 'unstructured') also had practically no effect (F=.018; p>.25). But the factor A (experimental vs. control conditions) was very significant (F=28.7; p<.0001) in the direction of better post test performance in the experimental conditions.

For retention test scores as dependent, the results were similar, be it that the independent factors together explained less variance of the dependent, i.e. there is more error variance here. In short: no interaction effect (F=.411; p>>.1), no effect of fear-of-failure (F=0!) or of factor N (F=.494; p>.25), and very significant effect of the combined experimental treatments (F=9.51; .0001<p<.0005) be it less strong than on the post test scores.

Qualitative data

From 5 subjects in each of the experimental conditions (structured and unstructured) think aloud protocols were obtained. Following our qualitative research questions, a scoring scheme was constructed in which each experiment the subject did was divided into 5 phases. These were: designing the experiment, predicting its result, conducting it, checking its result and learning from it. For each phase, relevant categories were made concerning the amount of initiative, correctness, specificity and certainty (overview). Further categories pertained to the reaction to unforeseen or conflicting results, a special case being an '*Aha-erlebnis* ' as prototypical for the kind of learning we expected (especially when following a period of surprise and/or confusion).

Analysing each experiment for each subject seperately, yielded 35 subject/experiment instances in the unstructured condition, and 40 instances in the structured condition. In these 75 instances, 16 different 'scenarios' were discernable as to how the experiment was performed and what overt learning effect it had. A summary is given in fig. 3.

A) Two of the sixteen scenarios were characterized as "having a more specific (detailed) grasp of the relevant aspect as a result of the experiment". This happened in 3 instances in the structured condition and in 3 instances in the unstructured condition.

B) One of them was characterized as "having learnt the irrelevancy of an aspect as a result of the experiment". This happened in 2 instances in the unstructured condition only.

C) Two were characterized as "having learnt about an aspect only after explanatation of the experiment (not just by the experiment itself)". This happened in 1 instance in the structured condition and in 2 instances in the unstructured condition (one of them accompanied by an 'Aha-erlebnis').

D) One was characterized as "having acquired a misconception as a result of the experiment (because of incorrect execution)". This happened in one instance in both the structured and the unstructured condition.

E) Two were characterized as "having learnt about an aspect as a result of the experiment". This is the (socratic) scenario we were after. There were 2 uncertain instances (no prediction given) in the structured and in the unstructured condition each. One of the two in the unstructured condition was followed by an 'Aha-erlebnis' during explanation. There were 5 certain instances in the structured condition only, every time in the same experiment (i.e. for **all** subjects), two of them accompanied by an 'Aha-erlebnis'.

	structured		unstructured		TOTALS			
	low FoF	high FoF	low FoF	high FoF	str.	unstr.	FoF-	FoF+
more specific	2	1	0	3	3	3	2	4
unlearning	0	0	0	2	0	2	0	2
after explan.	0	1	1	1	1	2	1	2
neg.learning	0	1	1	0	1	1	1	1
learning ?	1	1	2	0	2	2	3	1
learning !	2	3	0	0	5	0	2	3
learning ?+!	3	4	2	0	7	2	5	4
no learning	11	17	17	8	28	25	28	25
TOT	14	23	21	11	37	32	35	34

fig. 3: *# instances for each learning scenario category, split by 2 x 2 levels.*

F) The rest of the scenarios (9 of them) had to be characterized as "no overt learning", either because the subject gave a satisfactory prediction and argumentation before the experiment, or because the subject did not overtly show sufficient grasp of the relevant aspect after the experiment. This happened in the majority of instances (28 in the structured, 25 in the unstructured condition).

Discussion

The quantitative data analysis shows rather clearly that understanding of the topics 'heat' and 'temperature' has been increased by the 'heatlab', both on short and somewhat longer term. But the amount of structuring in the 'heatlab' has not been shown to make any difference, neither as a main effect nor in interaction with fear-of-failure.

Unless the amount of structuring really does not matter at all, which seems doubtful, this invalidates our claim that automation of ATI-research procedures will yield stronger effects. Or at least, the results show that strong effects are not guarantied by automation.

We think the most likely cause for the lack of effect in this study is the influence of the experimentator. That is, the experimental procedure was still not enough automated. The experimentator was present while students worked with the 'heatlab' and occasionally interfered, be it to prompt the student to think aloud, to help out when the interface mechanism was not understood, or to break off exploration that took too long. The think aloud protocols show that, in these cases, involuntary hints were given which may have tended to lessen the difference between the experimental conditions.

Obviously, the way to proceed now is to explore the data further for anomalies (the think aloud protocols can be very helpful to this end), and perform some more experiments in which the experimental procedure is made tighter yet (i.e. still more automation and less experimentator interference).

Apart from this, the table in fig. 2 contains some suspicion-arousing data.

First, the matching procedure seems to have been successful in the sense that the average scores for fear-of-failure, intelligence and pre test scores are fairly equal. But in the 'structured' condition, the variances of both pre test and intelligence scores deviate considerably from those in the other conditions.

Second, on both post test and retention test, scores of pupils in the unstructured condition are markedly more homogeneous (have less variance) than those in the structured condition. This means that weak pupils profited more from this condition than good ones. If anything, we would expect the opposite result, since the structured condition is more uniform, and since the variance of the pre test scores was lower in the structured condition to begin with! It could be that the experimentator's interference, which tends to be more frequent in unstructured circumstances, has been instrumental in bringing about this effect.

Third, correlations among pre test, post test and retention test are fairly low. There are two noticeable exceptions. Post test and retention test correlate reasonably, but only in the 'structured' condition. Even more strangely, in the control condition, the correlation between pre test and retention test is markedly negative, although in that condition, the correlation between pre test and post test is higher than in the other conditions.

Currently we are not able to definitely explain these findings.

As for the qualitative data, analysis as performed indicates that socratic learning does take place, be it not often. It happens more often in the structured than in the unstructured condition, but there is no difference on the fear-of-failure factor per se. There is some indication of an interaction between fear-of-failure and structuring in that there is no socratic learning for high fear-of-failure subjects in the unstructured condition. This is exactly the interaction we expected, but which did not show in the quantitative data!

From this analysis it would seem that little learning took place at all, yet the quantitative data show a very significant amount of learning. Note, however, that this analysis is a very conservative one in that only overt indications of learning were taken into account. I.e. if the subject gave no prediction, learning could usually not be ascertained.

However, this kind of conglomerate analysis of protocol fragments still tells us little (if anything) about the actual learning **process**. For instance, both occasions of a self-induced Aha-erlebnis (i.e. not caused by explanation) arise in the context of a socratic-type learning event and are preceded by utterances of surprise, moreover the clearest indications for socratic learning are given by subjects doing the same experiment in the same condition and are given by all 5 of them. As another instance, some subjects use the same scenario fairly consistently over the series of experiments they do.

What we need to do, then, is a holistic subject-by-subject re-analysis of the protocols, with an emphasis on why learning did or did not occur. This re-analysis is not yet completed at this moment, but we can state some preliminary tentative findings:

1. For real socratic learning to occur, it seems to be necessary that the pupil states, or at least is explicitly aware of, a prediction about the experiment. But more than this, the pupil has to have made some emotional investment in this prediction (really believe it or being curious about it).
2. Most misconceptions pupils have (at least on heat and temperature) are not solid models. They are quite volatile and context-dependent and therefore do not permit predictions with much emotional investment.
3. In 'experimental socratic learning', therefore, it would be beneficial not to try to disconfirm a pupil's model immediately, but first to strengthen it by a series of congruent experiments and only then giving the disconfirming experiment, which should be as blatantly incongruent as possible.
4. Pupils seem to have a individually differing attitude to experiments (e.g. whether to explore, how quickly to believe surprising evidence, etc.) which would have to be taken into account during the teaching process. This would seem to call for a high level of structure and strict monitoring, using intelligent COO techniques.

These findings examplify the type of results we strive for in our qualitative analyses. It must be made clear that the findings are **tentative** until educational models based upon them are implemented, used in future experiments, and indeed show superior learning.

Overall Discussion

We have described a type of laboratory simulation intended for remedy of misconceptions, and an experiment performed with a prototype of such a simulation ('heatlab'). Although the results from the experiment are inconclusive in many respects, we feel entitled to state the following conclusions.

The 'heatlab' simulation did cure misconceptions to a very significant amount. This means that our intended educational use of this type of laboratory simulation, as a tool to be integrated in an ITS, seems promising indeed. Our idea that the ITS should structure and monitor exploration of the simulation environment may need to be revised, since unstructured exploration seems to give equally good results. But, as said in the previous section, we need to do more experiments, possibly with other laboratory simulations.

Coupled with the object-oriented graphical system PCE, the programming language Prolog can be used to write real-time simulation environments. However, the speed of the PCE-Prolog system we used is limited. This poses a limit to the possible complexity of a simulation. More recent versions of PCE-Prolog are considerably quicker, however. Still greater speed, and therefore more complex simulation, are expected to be possible in the future.

Writing applications in PCE-Prolog appeared to be quite straightforward. The source code does not have to be concerned with low-level graphical routines, since PCE takes care of these itself. Moreover, PCE stimulates writing fairly independent program blocks, which makes testing much easier. PCE-Prolog therefore seems very suitable for quick prototyping. Especially now that we have the 'heatlab' program, we expect other laboratory simulations to take relatively little time and effort.

References

Abelson, H. and A.A. DiSessa, (1980). *Turtle geometry, the computer as a medium for exploring mathematics*. Cambridge (Mass): MIT Press.

Anjewierden, A.A., (1986). How about a Prolog object? Proceedings of 3rd Workshop on Object Oriented Languages, Paris.

Anjewierden, A.A. and J. Wielemaker (1988). *PCE-3 Programmers Manual*. Technical Report C1.3, ESPRIT project 1098. Amsterdam: SWI, Un. of Amsterdam.

Bierman, D.J. and P.A. Kamsteeg (1987). A methodology for the development of 'intelligent' educational systems. In: C. Schwarzer and B. Seipp (ed.), *Trends in european educational research*. Braunschweig: Technische Universität Braunschweig.

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

Cohen, A., (1983). Comparing regression coefficients across subsamples. *Sociological methods and research, 12*, 77-94.

Collins, A. and A.L. Stevens, (1980). *Goals and strategies of interactive teachers*. Report 4345. Cambridge (Mass): Bolt, Beranek & Newman.

Cronbach, L.J. and R.E. Snow, (1977). *Aptitudes and instructional methods*. New York: Irvington.

DiSessa, A.A., (1986). Artificial worlds and real experience. *Instructional Science, 14*, 207-227.

Entwistle, N.J., (1981). *Styles of learning and teaching*. Chichester: Wiley.

Hermans, H.J.M., J.J.F. ter Laak and P.C.J.M. Maas, (1972). Achievement motivation and fear of failure in family and school. *Developmental Psychology, 6*, 520-528.

Jansweijer, W.N.H., (1988). *PDP*. Doctoral dissertation Un. of Amsterdam.

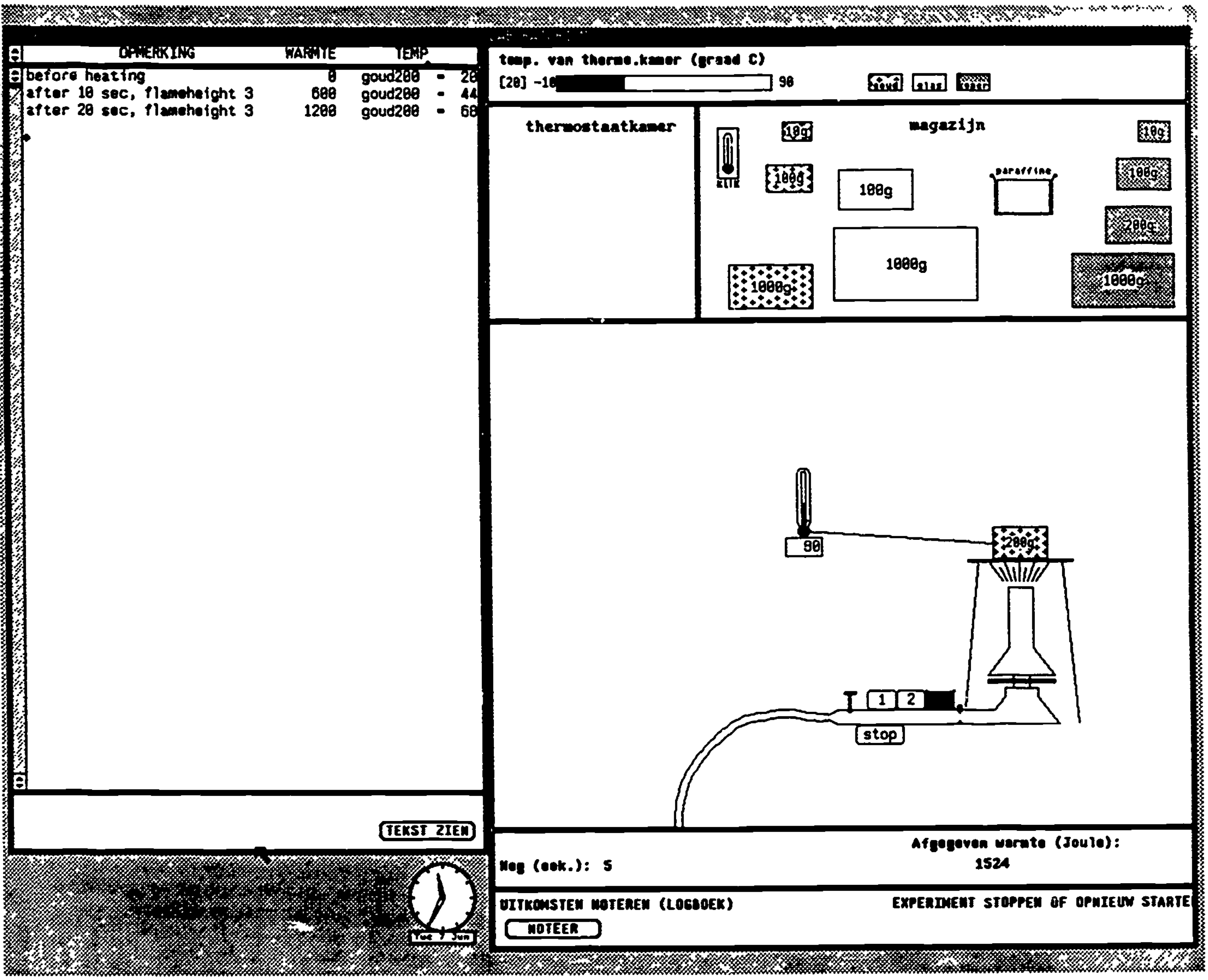
Jansweijer, W.N.H., L. Konst, J.J. Elshout and B.J. Wielinga (1982). PDP: A protocol diagnostic program for solving problems in physics. Proceedings of 5th European Conference on AI, Paris.

Kamsteeg, P.A. and D.J. Bierman, (1989). A laboratory simulation environment in PCE-Prolog. To appear in: L. de Leeuw (ed.), *Computers and educational research in the Netherlands*. Lisse, Swets & Zeitlinger.

Kerlinger, F.N. and E.J. Pedhazur, (1973). *Multiple regression in behavioral research*. New York: Holt, Rinehart & Winston.

Papert, S., (1980). *Mindstorms: Children, computers and powerful ideas*. New York: Basic Books.

Simons, P.R.J., (1980). *Vergelijkenderwijs: Onderzoek naar de invloed van metaforen op het leren*. Doctoral dissertation Un. of Tilburg.



OPMERKING	WARMTE	TEMP
before heating	0	goud200 - 20
after 10 sec, flameheight 3	600	goud200 - 44
after 20 sec, flameheight 3	1200	goud200 - 68

temp. van thermo.kamer (graad C)

[20] -10 90

Nog (sek.): 5

Afgegeven warmte (Joule): 1524

UITKOMSTEN NOTEREN (LOGBOEK)

EXPERIMENT STOPPEN OF OPNIEUW STARTE

fig. 1: *user interface of the 'heatlab'*