1989

Onderzoek van het onderwijs: leren en onderwijzen

Jules M. Pieters (red.)

Onderzoeksthemagroep
Onderwijsleerprocessen

toegepaste
ONDERWIJSKUNDE

Universiteit Twente

Cognitive ATI research using a simulated Laboratory Environment[1]

Paul A. Kamsteeg & Dick J. Bierman
Vakgroep Psychonomie
Universiteit van 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

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 (Jansweijer, 1988). 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

Our prototype laboratory environment, called 'heatlab', is intended for remediating the confusion between the concepts of 'heat' and 'temperature', by doing experiments concerning these concepts.

Experiments consist of two stages. First, an experimental set-up is built by connecting objects: blocks to each other and/or to a bunsen burner (possibly after having given them an initial temperature in a thermostat room), thermometers to blocks. Also part of the set-up is setting manipulation controls 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.

CONTEXTS FOR THE SIMULATION

EDUCATIONAL CONTEXT

Simulation environments have so far been used mainly in the LOGO approach to education (Papert, 1980), which is rooted strongly in the 'discovery learning' philosophy. 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. One form of *strictly guided* discovery, namely guided self-remedy of misconceptions, is the Socratic Dialogue technique (e.g. Collins & Stevens, 1980).This 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. We will use our laboratory simulation, and ones similar to it, to try to remedy misconceptions in such a 'see for yourself' way. In the long run we intend to integrate these simulations into an Intelligent Tutoring System for thermodynamics. How strongly guided such a discovery-like misconception treatment should be, is a question which shall be discussed in the next section.

EXPERIMENTAL CONTEXT

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 disclose 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.

EXPERIMENTAL USE OF THE 'HEATLAB'

THE QUESTIONS

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: Can we find utterances of surprise and disbelief in the think aloud protocols (as indications of unexpected experimental results), 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 RESULTS

QUANTITATIVE DATA

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

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) 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'.

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. 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. However 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 guaranteed 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.

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.

A subject-by-subject re-analysis of the protocols, with an emphasis on why learning did or did not occur yielded the following (preliminary and 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.

NOTES

[1] This article is a condensed version of a paper presented at the AERA conference, march 25-30, 1989. It describes SVO-funded research project 7015: 'De relatie tussen vormen van computergestuurde practica, leerlingkarakteristieken en cognitief leereffect'.

REFERENCES

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.

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 UvA.

Kamsteeg, P.A. and D.J. Bierman, (1989). Cognitive ATI research: A simulated laboratory environment in (PCE)-Prolog. Paper presented at AERA conference, San Francisco, march 25-30.

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.