'Intelligent' Tutoring Systems as a driving force behind, and a useful tool for good Educational Research[1]

Dick J. Bierman & Paul A. Kamsteeg
University of Amsterdam

Abstract
It will be argued that ITS's for complex domains are not primarily of practical use but that they rather are instrumental in bringing about advances in the understanding of learning and hence in the design of optimal teaching strategies. The basic assumption underlying the idea that ITS might be a better vehicle for instruction than traditional teaching is, that for each individual there exists an optimal teaching 'remedy' once a very specific cognitive 'diagnosis' has been made. Thus cognitive diagnosis is a crucial problem. In order to reduce the search space of potential 'buggy' knowledge that could explain the student's cognitive behavior, heuristics based upon a.o. research in mental models is necessary. During the development of an ITS for thermodynamics at the University of Amsterdam research had to be done simultaneously on the prevailing mental models and misconceptions in this domain. Results are briefly discussed. Not only the actual cognitive state of the student and the recent intervention history are thought to be important but also certain personality traits might relate to an optimal (form of) teaching interaction. Traditional Aptitude-Treatment Interaction research is also based upon this notion. ITS's, once in their operational phase, can be used as instruments in ATI- research. In contrast with the traditional approach, the research can be done with subjects who, apart from their specific personality structure and/or cognitive style, have specific cognitive states (eg. misconceptions). Together with the fact that the interaction using an ITS is better controlled this might result in effect sizes that are considerably larger and findings that are better interpretable in terms of theories of individual learning. Current research at the University of Amsterdam using an 'intelligent' simulation environment or microworld, focuses on the interaction of the amount of structure in the educational environment with some personality traits in pupils with well-established misconceptions.

Introduction

The development of an 'intelligent' Computer coach for thermodynamics at the UvA might be typical for the proposition that this kind of projects stimulates relevant educational research. The project started about 4 years ago and was concluded in 1986 (Bierman & Kamsteeg, 1987). It was a part of a large endeavour to study the acquisition of knowledge in formal domains like Physics. The focus of this larger global project was on how people solve problems and the main research method was the analysis of thinking aloud protocols produced by subjects while they are solving problems. To aid the researcher in the analysis of these protocols a program was written that was able to solve thermodynamics problems like

[1] Paper presented at the 2nd EARLI conference, Tubingen, F.R. of Germany, sep. 19-22 1987.

the ones presented to the subjects. This program could be seen as a model of an obedient problem solver and the thinking aloud protocols of an obedient subject should highly correlate with the trace produced by this program (Jansweijer et al, 1982).

The architecture of the computercoach was build around this program as can be seen in figure 1.

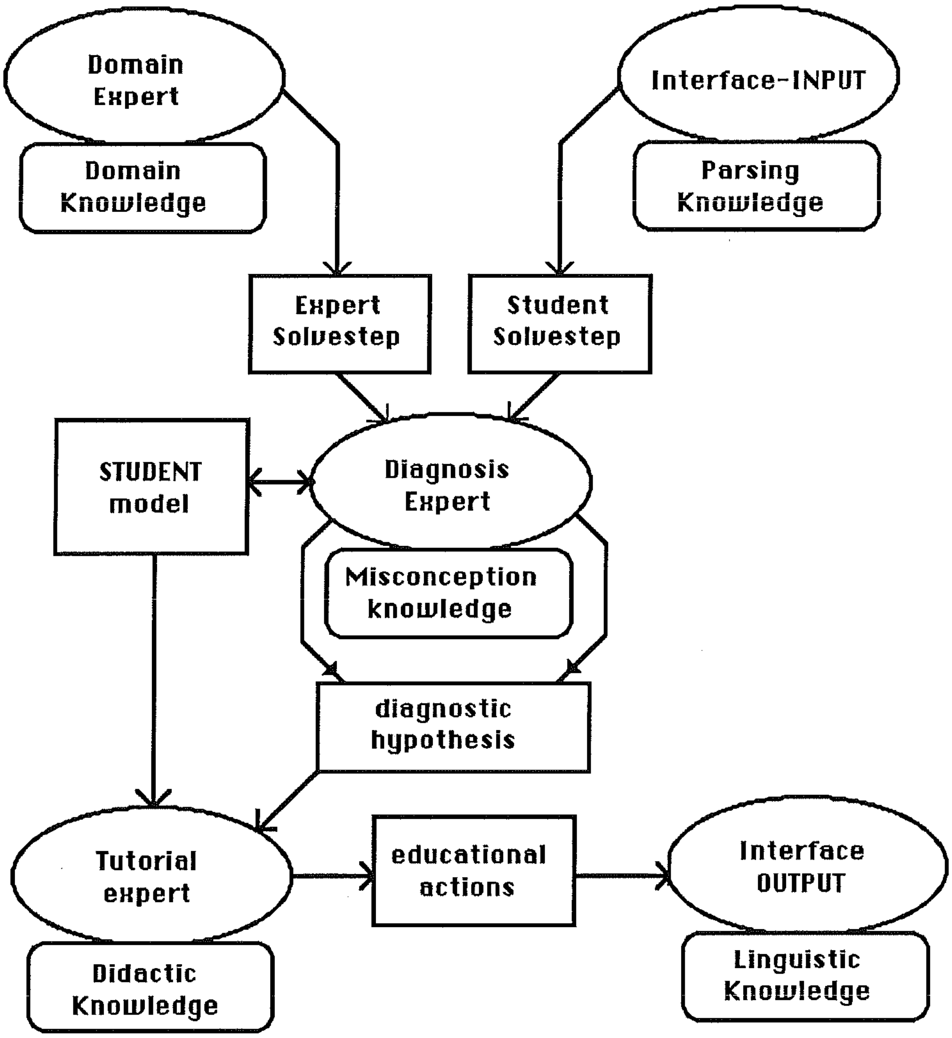


fig.1: the architecture of the computer coach.

On top of this architecture there is a component which either selects the next problem or decides that the pupil has learned enough.

The system consists of four components. Each of these components performs a separate function.

The domain expert

The domain expert component is the original 'obedient problem solver'. This component accepts a problem text as input; consecutively it solves the problem, producing a 'mental trace'. The trace is also called 'norm' trace because of the underlying educational goal that the pupil should eventually confirm to this solving norm[2]. From the beginning of the project on, the intention has been to localize all knowledge pertaining to the domain of thermodynamics within this component. The reason being that a similar computer coach for a different domain could then be constructed by replacing nothing but the domain expert[3].

The diagnose expert

The function of the diagnose expert is to generate hypotheses about the reason for deviating cognitive behavior on the part of the pupil. A further important task is to maintain and update a model of the pupil. One way of doing this is estimating, for each knowledge item that the domain expert has access to (factual knowledge, procedural knowledge as well as knowledge about solve strategies), the mastery level of the pupil. In this approach the current knowledge of the pupil is seen as a subset of the eventual knowledge he should come to possess. The pupil model is therefore an incomplete, but apart from that identical, copy of the expert model. This kind of pupil model is called 'overlay model', because (apart from the missing elements) it exactly covers, as it were, the expert model (eg. viz. Goldstein, 1982).

In domains, such as physics, about which one may assume that a pupil already has some (possibly incorrect) knowledge to begin with, it is furthermore necessary to keep track of the incorrect knowledge (misconceptions, incorrect mental models) the pupil has. Since this is knowledge an expert generally does **not** have, it does not fit in an 'overlay model' but must be represented separately, for instance in a list of misconceptions ('bug catalog') or by extending the overlay model with pointers to incorrect knowledge (eg. Brown & Burton, 1978).

The aforementioned function of generating hypotheses about deviating pupil behavior is performed in the context of such a pupil model. That is, the likelihood of a certain hypothesis is deduced on the basis of information in the pupil model.

The tutoring expert

The task of the tutoring expert is to decide upon an optimal instructional intervention with respect to the diagnosis. In this task also, the pupil model plays an important role. Apart from the cognitive aspects of the pupil model, personality features of the pupil can be taken into

---

[2] Of course, the idea that there is one single norm is nonsense. For instance, the 'mental trace' of the computer contains sequences which can be permuted. More generally put: a norm trace represents only one program of actions, whereas there may be several correct alternatives

[3]From the remainder of this paper, especially from the neccessity to implement knowledge about 'screwy' mental models in the diagnostic component, it follows that this goal was rather 'idealistic'.

consideration in deciding upon an optimal instructional action. E.g. the personality trait 'negative fear of failure' can play a part in deciding to give either much or little procedural hints.

The interface

An (user) interface is that part of a computer program which takes care of the interaction between the user and the rest of the program. In our case this means that the actions the pupil performs at the computer terminal have to be translated into a form that can be used by the diagnosis expert. That is, motor actions have to be interpreted as reflections of cognitive actions. Reversely, instructional interventions of the tutoring expert have to be presented to the pupil in the form of one or more comprehensible sentences.

Implicit vs. Explicit Knowledge

A major difference between this kind of CAI and the more traditional frame-based CAI is the fact that the didactic knowledge in this latter type of CAI is implicit in the program (for instance in the branching decisions or in the remedial texts) whereas in ITS's the didactic decisions are based on explicit didactic knowledge.

In the early stages of the Computer coach project it became clear that there was no formal source of information with regard to this explicit knowledge, the process of individual cognitive diagnosis nor of the relation between diagnoses and optimal tutorial interactions. The majority of educational research concerns the classroom situation, i.e. its results are rather unspecific in as far as individual students are concerned. For instance the traditional Aptitude-Treatment Interaction research generally uses large groups with students selected on the basis of personality traits but not on the basis of specific cognitive deficiencies. Also theories of learning are generally too unspecific to derive a coaching strategy from (Kamsteeg, 1984). According to Anderson (Anderson, 1985) his ACT-theory is an exception and one might infer from his theory that student should be kept as close as possible to the correct solution path.

The lack of information regarding didactic knowledge stimulated us to start a number of small projects each of them potentially having some practical use in the more traditional environment of the classroom (Looy,1987 ; Malhas,1986 ; Stehouwer, 1986).

Enhanced Band-width of communication

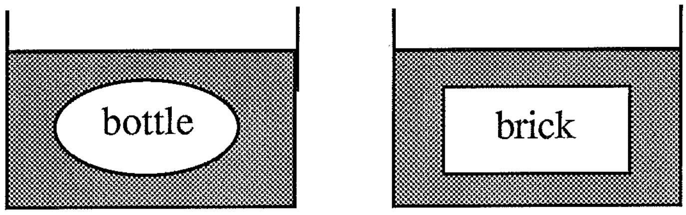
With regard to diagnosis it was necessary that the motor actions of the student at a terminal should be parsable in terms of cognitive actions. To that purpose we developed a graphic user interface called the 'scratchpad' using PCE-Prolog, an object oriented extension to the Prolog-language (Bierman & Anjewierden, 1986). The student is forced to make all his notes on the machine itself. The most interesting feature of this interface is the ease with which the contents of the scratchpad can be parsed symbolically. For instance it takes only a few lines of code to find out whether the student did recognize that a given problem was a dynamic one. Of course the 'scratchpad' is not a fundamental educational finding but this largely technical spin-off of our project is directly useable in traditional CAI-programs.

Mental Model research

A more fundamental aspect with diagnosis is caused by the fact that the malrule approach fails in the domain of thermodynamics. The malrule approach to diagnosis is essential the search for a piece of false or lacking (procedural) knowledge that could explain deviating behaviour of the student. This approach has originally been explored by Brown and Burton (1978) for simple arithmetics (subtraction). They showed that for a majority of the incorrect behavior it was possible to single out (combinations of) malrule(s) that could simulate this particular behavior.

For complex domains, like physics, this approach is not feasible since the search space is too large. The number of cognitive procedures which are activated before a correct solution is found is an order of a magnitude larger than for subtraction. Even when diagnosis is done on intermediate steps there are too many potential malrules that could explain an incorrect step. Therefore we need some heuristics (rules of thumb, eg. telling which cluster of malrules is most probable) to reduce the search space[4]. This need triggered a research project into different mental models which are present in the domain of thermodynamics and their respective probability of occurrence (Stehouwer, 1986). A number of qualitative questions were asked of about 200 students. A typical question is given in figure 2:

[4] Also it is necessary to be able to make an abstraction of a specific error to an error-class because tutorial actions should be based upon type of error instead of on a specific instance of the error.



A brick and a bottle with water of the same weight have both a temp of 50 C. The bottle is heated for 15' until 100 C. The brick is heated for 30' also until 100 C. They are plunged into identical containers with water. If they have reached a T=50 the T of the containers is measured.

What container has the highest T?
a. with bottle, b. with brick, c. identical

WHY?

Fig.2: example of mental model question

The subjects were not only requested to mark the proper alternative but they were also asked to justify their answer. From these justifications we were able to derive a taxonomy of mental models concerning the concepts of heat and temperature. At the core of the incorrect models is the general idea of heat as a particle (figure).

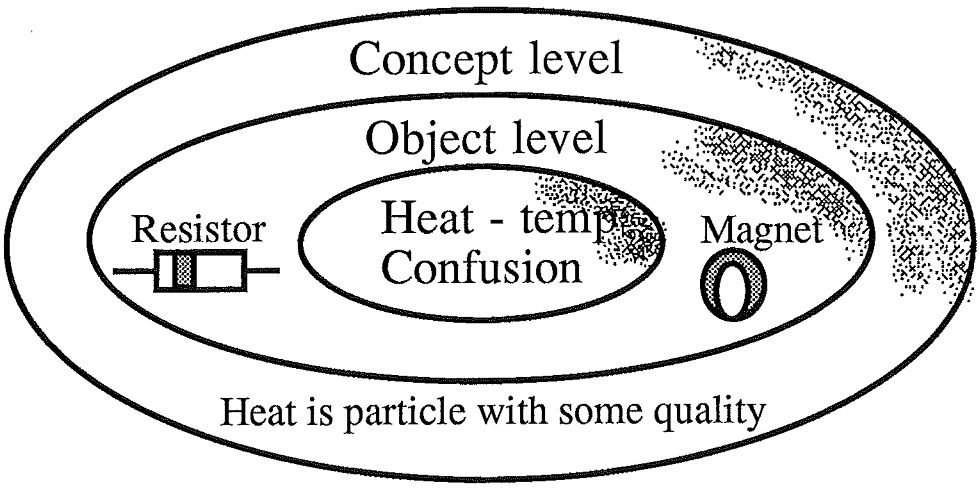


Fig.3: The postulated Mental-Models underlying the Heat-temperature confusion.

Around this fundamental metaphor we found that object property could be labeled as

Magnetic-like or Resistance-like with regard to heat-transfer. 'Magnetic' thinkers would say that objects that are heated easily, will loose their heat difficultly while 'Resistant' thinkers would say that objects that are heated easily will loose their heat easily. Finally we found protocols that indicated some model describing the relation between an object and its surroundings. This relation was described in terms of quality of heat transfer. Generally this quality would reflect itself in the **time** needed to heat an object embedded in a heat bath. Some people had the idea that slow heating implied a low quality of the heat (i.e. the heat-'particles' to have a low temperature) while others had the reverse idea. We called them 'Battery' thinkers and 'Simmer' thinkers.

These hypothetical models were tested by presenting a number of simple qualitative questions directly related to these models using the same subjects.For instance the following question was asked:

If an object makes a container with water increase its temperature by 10 °C*, the object will have lost 10* °C.

It turned out that 'thinkers' who combined the Resistant and the Battery model were very consistent in their use of this model while other thinkers were much less so. Although the Resistant-Battery model is physically incorrect it very often yields the correct prediction in daily life. For the other, less stable models it appears that they are highly context dependent and probably it is not justified to label those as mental models which suggest some form of stability (van Looy, 1987)[5]. Although we embarked on this research to find heuristics that could help us in the diagnosis of errors for our ITS it must be obvious that these findings can and must be of value in the classroom too!

Field Research

With regard to the question of optimal tutorial interventions we started field research asking teachers for common cognitive problems that they encountered in the classroom and also which action they undertook to relieve such a problem (Malhas, 1986). This field research yielded lists of problems ranging from inadequate support knowledge, like insufficient knowledge of algebra, to more complex problems like the aforementioned heat-temperature confusion. However the list of tutorial actions that was offered was rather short. The most popular action was 'explain it again'. During this research it became rather apparent that most teachers were primarely interested in getting the pupils ready for the exams instead of having them understand physics in a more fundamental way. This was even true to such a degree that we became so depressed as to propose to remove the subject of Physics from the pre-university curriculae! However we also received a few nice examples of Socratic dialogues

[5] Although it can not be ruled out that we did not postulate the right mental models to begin with. That too can explain the apparent instabiltity.

which would be valuable in an ITS but also might be communicated to other teachers of thermodynamics. On the whole we were forced to conclude that we had to elicit didactic knowledge from teachers with experience in individual coaching rather than from classroom teachers. This was done using a methodology called MUSPA.

The MUSPA, individual knowledge elicitiation research

This methodology has been extensively described elsewhere (Bierman & Kamsteeg, 1987). The essence of the method is to create a hybrid system of the unfinished Computer coach and the human coach working together while coaching the students problem solving. The set-up is given in figure 4.

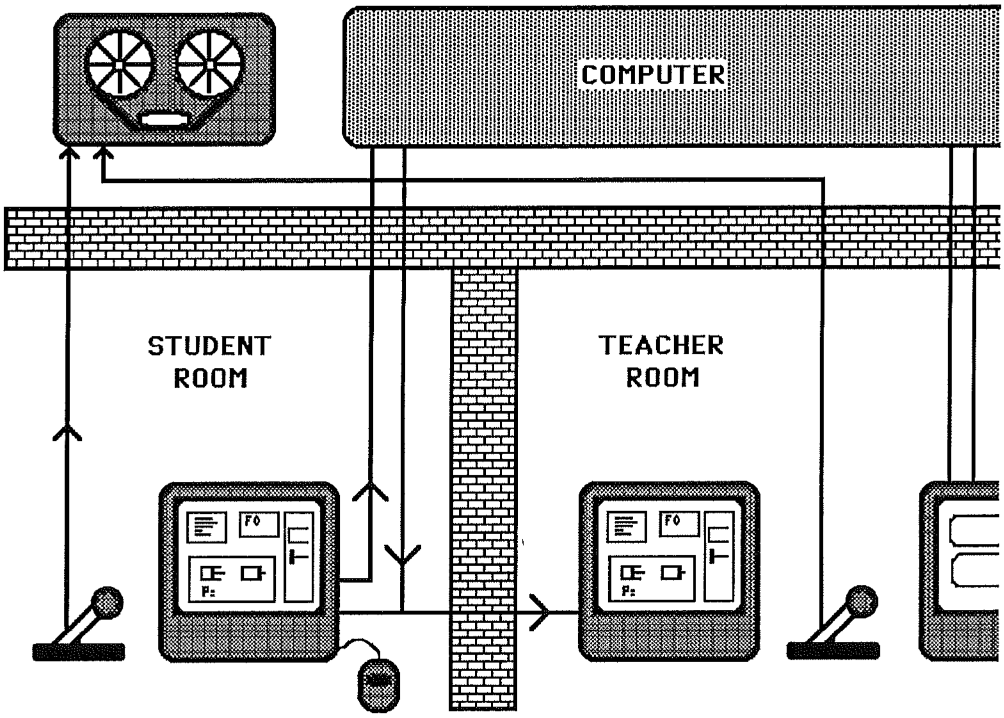


fig.4: The MUSPA experimental set-up

This approach yield three protocols, one from the student, one interaction-protocol and one from the teacher. The thinking aloud protocol produced by the human teacher is analyzed carefully in the search of im- or explicit didactic knowledge (eg. coaching strategies). This didactic knowledge is subsequently implemented in the Computer coach and a new testcycle begins. In the more final stages the human coach is able to follow the reasoning of the computercoach on his terminal and his thinking aloud task gradually shifts into a criticizing aloud task. In this way substantial didactic knowledge was elicited. For instance some teachers used a strategy of generating hints which became more and more specific (Bierman & Kamsteeg, 1987). This technique is useful for the development of any knowledge based system meant to run within the context of an ongoing dialogue since it also represents a crucial test on the design of the user-interface.

Student Models

In order to explore the knowledge to be represented in the student-model we reanalyzed the teacher protocols, specifically looking for im- or explicit references to the inferred student cognitive state. Alas, we only found rather superficial remarks dealing with global impressions of the human coach. Thus the question arises if it is necessary to have a very specific student model. Sleeman, for instance, presented a paper at the last AI & Education conference with a large number of educational heuristics each of which supposedly would improve the teaching. None of these heuristics were dependent on the specific type of student. Sleeman claims (informal communication) that student models are a temporary fad.

Comparison of teaching result of human and computer coaches

Although not very much comparative research has been done on the teaching results obtained through human versus computer coaching, the preliminary conclusion might be that classroom teaching yields the worst and individual human coaching the best results. Computerized teaching falls in between. So the conclusion must be that the individual human coaches do posses some knowledge that is lacking in the computer coaches. A possible candidate might be the specific knowledge concerning the cognitive state of the student, the so-called student model. After all the whole idea of individual coaching being better than collective teaching is based upon the idea of individual differences which might require different teaching actions. And although virtually each article on ITS stresses the need for a good student model, most computer coaches do only maintain a very superficial student model. However from out student-model research it appears that human teachers do not use very specific student-models.

The fact that CAI yields better results than the classroom teacher could then be explained by the general valid heuristic that (near) immediate feedback is better for every student independent of his/her cognitive state.

ATI-research

But then again how do we explain human **individual** coaches outperforming computer coaches? Is it the subtle interaction between personality trait and educational treatment? None of the CAI systems, that we know of, has knowledge implemented that takes personality traits into account. Generally Aptitude Treatment Interaction research does not show very strong effects (Snow et al, 1980). So at first sight this hypothesis might be far fetched. However it should be remarked that ATI research is generally evaluated by using rather global measures from pre and post treatment tests. This means that these finding tell us something about one specific treatment for students with one specific trait but irrespective of their particular cognitive deficiencies. This brings us to the current research at the University of Amsterdam. Here we might use ITS's to select students with specific cognitive deficiencies and then study the ATI for this specific cognitive state. In our research 'Resistor-Battery' thinkers are selected and placed in either a structured or an unstructured learning environment. This environment is a computer simulated lab., or it might be called a 'microworld', which is constructed using AI-techniques (see figure)

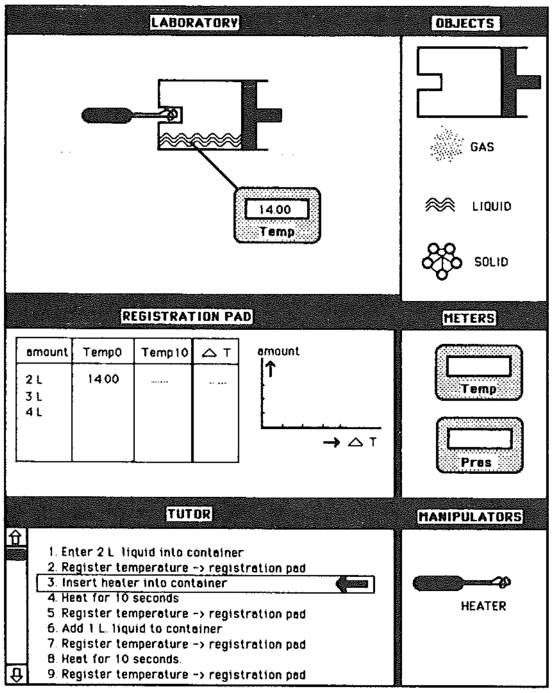


fig.5: Interface of Heat-lab.

Student might either explore or be well (in a Socratic way) guided in this lab. We will examine the interaction of these treatments with the personality trait 'fear of failure'. Students are requested to think aloud during their sessions. These protocols will enable us to trace down the moments of insight yielding the switch from the wrong mental model tot the correct mental model. We hope that this type of research will eventually yield theories of learning that are specific enough to be of value in individual teaching.

References

Anderson, J.R., (1987), Methodologies for studying human knowledge. In: *Behavioral and Brain Sciences* (yet to appear). Abstract in: UUCP newsgroup mod.ai, Message-ID:

<8701071550.AA01452....>.

Bierman, D.J. & A.A. Anjwierden, (1986), *The use of a graphic scratchpad for students in ICAI*. Proceedings of the 27th ADCIS Conference (pp.68-71), New Orleans.

Bierman, D.J. & P.A. Kamsteeg, (1986), Het monitoren van het Gewenst Handelingsverloop met behulp van een "intelligente" computercoach (monitoring the preferred program of actions and methods using an "intelligent" computer coach). In: P. Verhagen & B.J. Wielinga (Eds.), *Media in het onderwijs*. Lisse(Neth): Swets & Zeitlinger.

Bierman, D.J. & P.A. Kamsteeg, (1987). A methodology for the development of 'Intelligent' Educational Systems. in: E. Schwarzer (Ed.) *Current Trends in European Educational Research*. In Press.

Bierman, D.J. & P.A. Kamsteeg, (1987). *A computercoach for thermodynamics, Final Report*. PSYxx.87

Brown, J.S. & R.R. Burton, (1978), Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science, 2*, 155-192.

Goldstein, I.P., (1982), The genetic graph: a representation for the evolution of procedural knowledge. In: D. Sleeman & J.S. Brown (ed), *Intelligent tutoring systems*. London: Academic Press.

Jansweijer, W.N.H., L. Konst, J.J. Elshout & B.J. Wielinga, (1982), *PDP: A protocol diagnostic program for solving problems in physics*. Proceedings of the 5th European Conference on AI (pp.278-280), Paris.

Kamsteeg, P.A., (1984), *Kennis van docenten bij individuele coaching (knowledge of teachers in individual coaching)*. Internal paper 25.6.84.421, vakgroep psychonomie. Amsterdam(Neth): University of Amsterdam.

Kamsteeg, P.A. & D.J. Bierman, (1985), *A method for the study of teachers' thinking in coaching situations*. Paper presented at 2nd ISATT conference, Tilburg.

Looy, J. van, (1987), *Effekten van een, via de computer aangeboden, rekonstruktiedialoog op mentale modellen in de thermodynamica. (effects of a computer-presented reconstruction dialog on mental models in thermodynamics)*. Doctoraal werkstukverslag vakgroep psychonomie. Amsterdam: University of Amsterdam.

Malhas, M.S.M, (1986), *Inventarisatie en analyse van misconcepties en moeilijkheden met betrekking tot warmteleer. (inventarisation and analysis of misconceptions and problems regarding thermodynamics)*. Stageverslag vakgroep psychologische funktieleer RUL & vakgroep psychonomie UvA. Amsterdam: University of Amsterdam

Snow, R.E., Federico, P.A. & Montague, W.E. (1980). Aptitude, learning and Instruction: vol II. Hillsdale, Erlbaum.

Stehouwer, M., (1986), *Naieve ideeen en mentale modellen in de thermodynamica (naive* *ideas and misconceptions in thermodynamics).* Doctoraal werkstukverslag vakgroep psychonomie. Amsterdam: University of Amsterdam.