MOODS: A prototype tutoring system that detects students’ motivation

Angel de Vicente and Helen Pain
{angelv.helen}@dai.ed.ac.uk

Artificial Intelligence; Division of Informatics
The University of Edinburgh

Abstract

Previous work concerned with motivation in Intelligent Tutoring Systems (ITS) has mainly focused on the strategies that an ITS could use to motivate the student. In this paper we focus on the prior task of detecting the student’s motivational state, on which such strategies could be used.

Detecting the student’s motivational state is not straightforward, since a lot of information about the student is available to a tutoring system during an instructional interaction. In this paper we discuss the advantages and disadvantages of different methods of motivation detection. We also present a couple of empirical studies that we performed in order to inform the motivation detection strategies of MOODS, a simple ITS prototype, which we also present in this paper.

The approach taken for the development of MOODS is based on three main aspects:

1. The separation of the student model into two parts: one updated by the student and one updated internally by the system.
2. The separation of the motivation model into:
   (a) A ‘motivation state’ part,
   (b) A ‘motivation trait’ part.
3. The use of language as a communicator of affect.

While a formal evaluation of MOODS is yet to be done, we believe that the approach presented in this paper for motivation detection allows for the development of systems which interfere very little with student’s interaction, but that at the same time are able to detect a number of situations in which motivational problems arise.

1 Introduction

There has recently been a growing concern about why the ITSs developed to date are still not very ‘successful’ and about what should be investigated in order to increase their efficiency. Perhaps one of the most interesting suggestions in this direction refers to the importance of being able to incorporate into our tutoring systems models of how affective issues influence teaching and learning.

What affects students’ motivation to learn and how to influence the latter is a basic concern in classroom practice. Taking these issues into account in our Instructional Systems could bring many benefits. For instance, an Instructional System could focus on different aspects of the instruction depending on personality characteristics of the student; it could treat mistakes as less important if the student is going through a particularly bad time; or it could attempt to put the student in a particular mood in which he could be more receptive to the material to be taught.

But developing Instructional Systems that take into account issues of motivation in instruction (‘Affective Tutors’, for brevity) has a number of important hurdles:
1. Theories of motivation in Education are normally too vague and/or contradictory, and their implications for actual teaching practice are not always straightforward.

2. The components required for creating an ‘Affective Tutor’ cannot be developed in isolation. They require the development of an actual tutoring system, but one whose functioning has been radically changed by introducing components that deal with the affective aspects of instruction.

3. How can we know when we have successfully created an ‘Affective Tutor’? Measuring exclusively performance or enjoyment would certainly be inadequate. Ideally, we would like a composite measure of satisfaction and performance, but how should we weight each of them? And how could we compare this tutor with previously developed systems?

Despite this somewhat bleak picture, it is on the development of Affective Tutors that we focus in this paper. We build on the work of del Soldato (1994), who developed a system that uses a combination of traditional domain-based techniques and motivation-based techniques to plan the instruction. The domain-based planner suggests actions aiming to advance across the domain, while the motivational planner suggests tactics to increase or maintain the student’s motivation to work. A negotiation planner attempts to keep an adequate balance between the tactics (sometimes contradictory) suggested by the domain-based and the motivational planner. But her work focused mainly on the motivational planning issues, while not paying too much attention to issues of motivation detection. In this paper we explore some of these.

In section 2 we look in more detail at some of the background work on which we base a first approximation to a design of an ‘Affective Tutor’, which we present in section 3. Some of the issues of this first approximation needed empirical validation, and therefore a couple of empirical studies were performed. The first one, which is reported in section 4, had as its goal to explore the validity and the possibilities of self-report as a motivation diagnosis technique. The results seem to indicate that self-report is a valid approach to motivation diagnosis, but that it shouldn’t be used as the sole technique. In order to find appropriate and empirically valid rules that could guide an ITS in detecting a student’s motivational state, we performed a second empirical study, in which we elicited a number of rules to detect motivation, thanks to the expert knowledge of a number of human tutors. The role that language plays in our prototype is discussed in section 6.2, and a description of how all these pieces fit together in the actual prototype developed is left for section 6. Finally, we present some conclusions and possible further work in section 7.

2 Background Work

There has been recently some interest in issues relating to affective issues in tutoring systems, which comes in a context of renewed interest in the role that emotions play in human intelligence and how this affects Artificial Intelligence (AI) in general. This is reflected in an increasing number of conferences and publications covering these topics (e.g. Affect in Interactions, 1999; de Rosis, 2000; Emotion-based Agent Architectures, 1999; Emotion in HCI, 1999; Emotional and Intelligent, 1998; Picard, 1997; Practical Affective Computing, 1999).

This interest in the role of emotions in AI is itself motivated by research in the cognitive sciences arguing that emotions are not detrimental, but rather indispensable, for human intelligence. In one of the most influential recent books covering these issues, Damasio (1996) argues that reason may not be as pure as is commonly held, and that emotions and feelings play an important role in it for better or for worse, suggesting “that certain aspects of the process of emotion and feeling are indispensable for rationality.” (Damasio, 1996, pp. xiv–xv). And although questions about the validity of this argument have also been raised (e.g. Sloman (1999); Sloman and Croucher (1981)), this debate is helping to
increase the long overdue attention to affective issues in all areas of computing, including the field of Artificial Intelligence and Education (AI-ED).

But the concept of affect concerns a very large range of human phenomena, many of which have very little relevance for the creation of Affective Tutors. As Parkinson and Colman (1995) remind us, one of the most widely accepted ways of classifying human mental functions “defines three separate areas of cognition (thinking), affect (feeling), and conation (willing). Emotion is one of the most important and thoroughly explored forms of affect, and motivation is essentially just a new name for conation [...]” (Parkinson and Colman, 1995, pp. xi).

In this paper we treat the term ‘affective’ as a broad term referring to anything pertaining to emotions or not cognitive, following previous work on ‘Affective Computing’ (e.g. Issroff (1996); Picard (1997)). But the terms ‘emotion’ and ‘motivation’ need a more careful definition. There are obvious links between both, as emotions are often a precursor of motivational phenomena, and they influence the way we act towards our environment, but they refer to different concepts. As Parkinson and Colman put it, “Emotion and motivation both depend on the relationship between the organism and its environment. In the case of emotion, the emphasis is on the evaluative aspect of this relationship: how the situation makes the person feel; in the case of motivation, it is how the individual acts with respect to the situation that is of interest [...]” (Parkinson and Colman, 1995, pp. xi).

In the case of ‘Affective Tutors’, our emphasis is on motivation rather than on emotion. We are interested in how the student reacts to the situation, in order to create instructional systems that engage the student, that motivate him. As seen above, how the individual feels about the environment obviously influences how he reacts to it, and thus, research on detecting emotions (e.g. (Essa and Pentland, 1997; Lisetti and Rumelhart, 1998; Ortony et al., 1988; Roy and Pentland, 1996; Scherer, 1981; Tosa and Nakatsu, 1996; Vyzas and Picard, 1998; Yacoob and Davis, 1996)) is relevant to the development of Affective Tutors. But its relevance is not primary, as theories and models of motivation in education do not deal necessarily with the same constructs as emotion theories.

One such theory is that by Keller (1979), whose purpose is to identify major categories of variables of individual behaviour and of instructional design that are related to individual effort and performance. Keller sees behaviour as a function of the person and the environment, and the theory describes the influence of these two factors on three categories of responses: effort, performance, and consequences.

Effort (a direct indicator of motivation) refers to whether the individual is engaged in actions aimed at accomplishing the task, while performance means actual accomplishment. Performance is only indirectly related to motivation, and is affected by other factors, such as individual abilities or learning design of the instruction. According to this model, the effort that an individual puts into a task is influenced by three broad variables:

1. Motives (values), which refers to how individual needs and beliefs relate to choices of action.
2. Expectancy, which is concerned with how personal expectancies for success or failure affect behaviour.
3. The motivational design and management of instruction.

On the other hand, performance is influenced by:
1. Individual abilities, skills, and knowledge.
2. Learning design and management.

Many theories of motivation exist, but they do not necessarily apply well in the field of Education. Issues of sexual and aggressive instincts, reduction of biological needs, etc. certainly seem very unlikely to play an important role in classroom behaviour.
3. The effort the individual puts into the task.

The third category of responses in this theory is ‘Consequences’, which refer to the intrinsic and extrinsic outcomes that accrue to an individual (e.g. emotional responses, social rewards, material objects). These consequences are related to the performance and to the (environmental variable) “Contingency Design and Management”. These consequences play an important role in motivation as they feed back (through cognitive evaluation) into the motives and values of the individual. For example, a positive performance followed by an external reward (such as cash) may, as a consequence, influence the value that the individual places on the activity.

From this theory and the work of a number of other authors concerned with motivation in Education (Ball, 1977; Briggs et al., 1996; del Soldato, 1992; Gentner, 1992; Issroff, 1996; Keller, 1983; Lepper et al., 1993; Lepper and Chabay, 1988; Malone and Lepper, 1987), we present in table 1 a summary of the most important motivational goals of instruction. From their work, we also summarise the factors that influence these goals and some suggestions as to how to achieve the desired motivational goals. Table 1 offers a taxonomy from which guidelines for instructional design can be obtained and it could also be used to evaluate ‘Affective Tutors’. But as Malone and Lepper (1987) remind us (regarding their proposed guidelines), these suggestions should not be treated as a prescription, but rather as “a way of guiding and sharpening intuitions and aesthetic sensitivity, not a way of replacing them” (Malone and Lepper, 1987, pp. 249).

Have these suggestions been followed in order to create ‘Affective Tutors’? In many instances, it is argued that computers are intrinsically motivating or that their use may foster self-esteem. This is particularly so when new or uncommon techniques are used (for example, companion agents, learning by disturbing, etc.). These can be amusing and engaging (probably due to their novelty), but to attach to them an enduring intrinsically motivating value does not, we believe, approach the issue of creating motivating instructional tutors in the best possible way. Ideally we should be able to make use of theories and/or empirical studies that can tell us under which circumstances these techniques can have the best motivational influence for a particular student.

To date, the most relevant work in this direction is that by del Soldato (1994), which we take as a basis for our research. She developed two systems: a generic instructional planner called MORE (“MOtivational REactive plan”), and a Prolog-debugging tutor application as a ‘vehicle’ for MORE. In these systems, the model of the student’s motivation is characterised as a set of three numerical variables: effort, confidence and independence. The system also incorporates a domain-based planner and a motivational-based planner. As an example, in the case where the student succeeds in performing a task, the typical domain-based planner would suggest a harder problem. On the other hand, a motivational-based planner would take into account other variables, such as effort and confidence. Sometimes the domain-based planner will suggest a different action from that suggested by the motivational-based planner. To resolve these conflicts, she also implemented a negotiation planner, that is responsible for negotiating “between traversing the domain or increasing the student’s motivation”, pp. 45.

In the following section we explain in detail how the work reviewed here has shaped our project, and which avenues we set to explore.

3 Outlining the design of an ‘Affective Tutor’

Self (1988) reviewed 18 well-known Instructional Systems and identified 20 different uses that have been given to student models. By reviewing this list of student model uses, we see that all of them refer to modelling of cognitive characteristics of the student. In this section we focus on modelling affective characteristics. Ideally, a tutor should use a combination of these, and base the adaptation of the instruction both on student’s intelligence and motivation. But which motivational factors should we
<table>
<thead>
<tr>
<th>Goals</th>
<th>Influences</th>
<th>Suggestions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Challenge</strong></td>
<td>Level of difficulty</td>
<td>Provide some intermediate level of difficulty and challenge</td>
</tr>
<tr>
<td></td>
<td>Goals of task</td>
<td>Explicit, proximal and personally meaningful goals</td>
</tr>
<tr>
<td></td>
<td>Performance feedback</td>
<td>Goal attainment uncertain (probability of success one half)</td>
</tr>
<tr>
<td></td>
<td>Learner’s sense of self-esteem</td>
<td>Performance feedback that will engage and enhance the self-esteem (frequent, clear, constructive, encouraging)</td>
</tr>
<tr>
<td><strong>Curiosity</strong></td>
<td>Informational complexity</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Discrepancy or incongruity from present expectations and knowledge</td>
<td></td>
</tr>
<tr>
<td><strong>Sensory curiosity</strong></td>
<td>Light, sound, or other sensory stimuli of an environment</td>
<td></td>
</tr>
<tr>
<td><strong>Cognitive curiosity</strong></td>
<td>Higher-level cognitive structures</td>
<td>Making people believe that their existing knowledge structures are not well-formed (ie. they lack completeness, consistency or parsimony)</td>
</tr>
<tr>
<td><strong>Sense of control</strong></td>
<td>Range of outcomes provided by environment</td>
<td>Provide contingency (individualisation)</td>
</tr>
<tr>
<td></td>
<td>Probability of the person influencing those outcomes</td>
<td>Provide choices (to provide apparent and salient responsiveness)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Students’ actions should have ”powerful effects”</td>
</tr>
<tr>
<td><strong>Fantasy</strong></td>
<td>The emotional needs they help to satisfy (power, success, fame, fortune, etc.)</td>
<td>Create different fantasy contexts which users can choose from</td>
</tr>
<tr>
<td></td>
<td>Identification with imaginary characters (based on a) perceived similarity, b) admiration, and c) salience of that character’s perspective</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: A summary of important factors influencing student’s motivation.
model? And how could we model them? It is these two aspects that we pay attention to in the following sections.

3.1 Which motivational factors to model?

As mentioned in section 2, there are many factors that influence student’s motivation. Some of these are environmental while others are personal, but theorists do not always seem to agree on which these variables are.

Based on the factors that influence student’s motivation (a summary of which can be seen in table 1), we constructed a model of student’s motivation with a number of variables for which these theories offer suggestions of how they affect the student’s motivation and how they should be treated. This model can be seen in figure 1, and differentiates the motivational factors in two classes: trait and state variables.

As we mentioned, the motivation model is divided in two main categories: trait variables, or ‘permanent’ characteristics of the student; and state variables, or more ‘transient’ characteristics. This distinction between traits and states is common in the relevant literature, and is useful for motivational planning. The information about trait characteristics allows the system to individualise the instruction based on student types. The state variables allow the system a more detailed individualisation based on changes during the interaction with the system. Based on the mentioned literature, we give a definition for each of these variables in table 2.

The trait variables aim at giving the system a general picture of the goals it should pursue with a particular student. But to represent these personality characteristics as simple variables is, no doubt, a tradeoff between rigour and pragmatism. For example, a measure of how much fantasy a student likes during an instructional interaction in general, is an oversimplification of all the complex aspects affecting this particular construct.

On the other hand, even a general and simple approach like this can help to create a better tutoring system, better ‘tuned’ to a particular student. Suggestions on how to deal with each of these variables have been sometimes obtained by performing studies with tutors teaching a particular subject to a particular group of students (e.g. mathematics to elementary-school children (Lepper et al., 1993)). Whether
<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>Refers to the degree of control that the student likes having on the learning situation (i.e. does he like to select which exercises to do, in which order, etc. rather than let the instructor take these decisions?).</td>
</tr>
<tr>
<td>Challenge</td>
<td>Refers to the degree that the student enjoys having challenging situations during the instruction (i.e. does he like to try difficult exercises that represent a challenge for him?).</td>
</tr>
<tr>
<td>Independence</td>
<td>Refers to the degree that the student prefers to work independently, without asking others for help (i.e. does he prefer to work on his own, even if he finds some difficulties, and try to solve them by himself rather than asking collaboration or help from others?).</td>
</tr>
<tr>
<td>Fantasy</td>
<td>Refers to the degree that the student appreciates environments that evoke mental images of physical or social situations not actually present (i.e. does he like the learning materials being embedded in an imaginary context or does he prefer just “the facts”?)).</td>
</tr>
<tr>
<td>Relevance</td>
<td>Refers to whether important personal needs are being met by the learning situation (i.e. does he think that the instruction materials are personally relevant to him?).</td>
</tr>
<tr>
<td>Confidence</td>
<td>Refers to the student’s belief in being able to perform the task at hand correctly.</td>
</tr>
<tr>
<td>Sensory interest</td>
<td>Refers to the amount of curiosity aroused through the interface presentation (i.e. appeal of graphics, sounds, etc.).</td>
</tr>
<tr>
<td>Cognitive interest</td>
<td>Refers to curiosity aroused through the cognitive or epistemic characteristics of the task (i.e. regardless of the presentation issues, does the student find the task at hand cognitively appealing?).</td>
</tr>
<tr>
<td>Effort</td>
<td>Refers to the degree that the student is exerting himself in order to perform the learning activities.</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>Refers to the overall feeling of goal accomplishment (i.e. does the student think that the instruction is satisfying and that is getting him closer to his goals?).</td>
</tr>
</tbody>
</table>

Table 2: Definitions of motivation model variables
these findings can be appropriate in other settings is an issue that the authors themselves raise: “Plainly, [], the specific strategies that prove most effective with elementary-school children may not be the most effective for more advanced students. Young children, for example, may be especially susceptible to the use of fantasy or playful competition, whereas older students may find those techniques silly or dishonest” (Lepper et al., 1993, pp. 100). The provision of trait variables in our motivation model offers a simple solution to this problem, as students can let the system know about their major preferences for an instructional interaction.

The trait variables in our motivation model are: control, challenge, independence and fantasy. Theorists seem to agree on the importance of control and challenge for student’s motivation. Fantasy, although not many times included in theories of motivation in Education, seems to be a factor that can play an important role in engaging the student (Malone and Lepper, 1987). Independence, as defined in table 2, is related to challenge, but also to the interpersonal motivations in Malone and Lepper (1987)’s taxonomy: cooperation, competition and recognition. This concept was used by del Soldato (1994) in her motivation model, but here we consider it a trait characteristic that can guide a system using this motivation model to decide whether and how often it should offer help to the student.

The state variables all come from important factors recognised by theories of motivation, and they represent transient characteristics of the student that relate to the material being learnt. In figure 1 the state variables are presented in a more or less ‘chronological’ order. Thus, considerations of how relevant the task is to the student will likely happen before engaging in the task. This, together with how confident he feels about succeeding in the task, and the interest (both sensory and cognitive) that the lesson arouses in him, will influence the effort that he will put into the task. Satisfaction, as defined in table 2, represents the overall feeling of goal accomplishment, and will be influenced by all the variables above, plus by the outcomes (in the sense of the Keller (1983)’s theory of motivation, summarised in section 2) of the task.

3.2 How to diagnose those motivational factors?

How can we get the information necessary to give values to the variables in our motivation model? The approach used by del Soldato (1994) was to use knowledge-based rules which determine how the variables in her motivation model have to be updated depending on a number of aspects of instruction. But our model requires a deviation from this type of approach for two main reasons:

1. The model requires information about variables which are unlikely to change during the interaction, and for which other methods of elicitation should be used.

2. More importantly, given the intrinsic difficulty of detecting the student’s motivation, and given that there is not much theoretical grounds on which to base our knowledge-based rules, these should be used with caution. In the work of del Soldato (1994), the rules used seem to have been elicited through common sense, and although the rules seem more or less consistent with theoretical knowledge, their accuracy was not empirically tested. Knowledge-based rules offer the advantage that they can be used unobtrusively and that they can be implemented easily, but we believe that, if used, some other means of verification of their accuracy in real time should be used.

Our approach to motivation modelling requires a combination of four different approaches:

1. Questionnaire, to gather information about student’s trait characteristics.

2. Self-report and

3. Verbal communication, to gather information about student’s state variables throughout the interaction.

See (de Vicente and Pain, 1998) for a detailed review of possible approaches to motivation diagnosis.
4. Knowledge-based rules engine (based on empirical study) to update and verify the student’s motivation model.

3.2.1 Questionnaire and self-report

Information about trait characteristics can be gathered easily through a simple questionnaire at the beginning of the interaction. During the interaction, the approach of asking the student whether he wants help, or whether he is bored, etc. seems very attractive. It is probably one of the easiest methods to implement, and at the same time it is sometimes claimed that users can give subjective ratings easier than other explanations (Briggs et al., 1996). On the other hand, social studies research warns us of the difficulties of obtaining valid data regarding peoples’ attitudes (see, for example, Oppenheim, 1992). And even if we were able to obtain valid and reliable data through self-report, the question remains of how much such a method would interfere with instruction, and what would be the users’ reaction to it. In order to study students’ reaction to and usefulness of the self-report approach, we performed an empirical study in which a simple questionnaire and self-report was used as the way of communicating with the computer about students’ motivation. This study is discussed in detail in section 4.

3.2.2 Verbal communication

Language is a powerful communicator of affect, and it can be used in tutoring systems to empathise with students and to detect their emotional state during the interaction. Related work exists (e.g. (Allport, 1992; de Rosis and Grasso, 1999; Horvitz and Paek, 1999; Person et al., 1999)), but a computational model of affective educational dialogue is missing. We present the foundations of such a model in section 6.2.

The purpose of this model is to being able to generate a simple educational dialogue with a student, focusing primarily on its affective impact. Thus, the model has two main goals:

1. Given a particular student and an instructional interaction, the model should generate utterances of appropriate affective characteristics.
2. Given a reply by the student, the model should infer what this particular reply means in terms of the student’s affective state.

3.2.3 Empirically based motivation diagnosis knowledge rules

Following the self-report study presented in section 4, we performed another study in which participants with tutoring experience were able to see replayed previous student interactions with MOODS and were asked to predict students’ updates of the motivational model, and this, in turn was translated into formalised knowledge rules to diagnose the student’s motivational state. This study is discussed in detail in section 5.

3.3 Putting it all together

Although motivation modelling is the core aspect of this paper, we need to evaluate the approach taken, and for this we developed a prototype ITS. In order to develop it, we need to incorporate motivational planning rules, which are based on the work of del Soldato (1994), but with some changes to take into account our different approach to motivation modelling. Details of how all these pieces fall together into our prototype are given in section 6.
4 Self-report study

The motivation for the study presented in this section was to study students’ reaction to and usefulness of the self-report approach to motivation diagnosis. With this study we wanted to find out whether having the option of reporting the self-perceived motivational state is intrusive for the students and whether this technique could be useful as a way to detect student’s affective state in ITSs. Amongst the issues that we wanted to investigate were:

- Was self-report a method accepted by students?
- Were there any usage patterns of the different self-report sliders that we could use to refine the motivation model presented in section 3.1?
- Is the information obtained through the motivational factors sliders trustworthy?

4.1 Materials

For this study we used an early version of the MOODS system. The content taught with this early version of MOODS is Japanese numbers, and the main interface of this MOODS system is presented in figure 2. In this study we were not aiming for creating an engaging instruction. Rather, we wanted to know if self-report would be a viable option for motivation diagnosis and we aimed at offering to participants a variety of instructional situations that would hopefully result in clearly differentiable motivational states.

With this in mind we organised the instructional units in MOODS into six instructional paths. The first instructional path presents all the theory, then difficult exercises and then the easier exercises. We expected that this instructional path would create boredom, lack of confidence, etc. in the participants. On the other side of the continuum, instructional path number 6 deals only with Japanese numbers up to 20, with a pedagogically more appropriate approach: first presenting theory, then an easy exercise, then a game to consolidate the knowledge learnt and lastly the difficult exercise. We expected that this instructional path would create interest, pose a challenge, and through the game create a ‘safe’ environment (Spitzer, 1996) in which the students could practice their knowledge and which hopefully would increase students’ overall motivation.

4.2 Methodology

18 first-year university students volunteered to participate in our study. Conveniently, having developed six different instructional paths, we allocated 3 students to each of the instructional paths. After the participants had read the study instructions, they were asked to fill in a small on-line questionnaire regarding the trait characteristics of our motivation model. Once this questionnaire was filled in, the actual interaction with the system started. Student interaction with MOODS lasted for a varying amount of time for each student, ranging from only 8 mins 27 sec to 27 minutes (on average, interactions lasted about 14.5 minutes). After the interaction finished, the students were asked to fill in a questionnaire with 13 questions, regarding mainly their opinions on the system and the usage of the self-report motivational model. For each question they could answer within a scale from 1 to 5 (1 meaning ‘strongly agree’ and 5 ‘strongly disagree’), and they were also encouraged to give extra comments.

As part of the instructions for the study, it was explained to the participants that the system would have a set of sliders representing various motivational factors, and they were encouraged to “use these sliders as often as possible whenever you think there is a change in any of these factors, since it is necessary for the computer to understand your current situation in order to modify the instruction accordingly.”

---

3Originally published in part as (de Vicente and Pain, 1999b)
Figure 2: Interface of modified MOODS system

(emphasis in original instructions). This sentence was ambiguous on purpose in order to make the students believe that the computer would react according to their motivational state, as reflected in the settings of the motivational sliders. It was only after they had filled in the post-questionnaire and had offered any extra comments that they were debriefed on the true purpose of the study.

In total, the collaboration for each participant took about 30-40 minutes, and allowed us to gather a considerable amount of data, which was recorded thanks to the software TkReplay (Crowley, 1996). The results of the analysis of these data are presented in the following section.

4.3 Results

On acceptance of the self-report method From the post-questionnaire given to the students, two questions relate directly to students’ willingness to use the self-report method. Question 9 refers to the trait questionnaire and reads “I would prefer not to answer the trait questionnaire, even if it makes the instruction more efficient and personalised.”

The average answer to this question was 4.21 (on a scale from 1 to 5, where 5 meant ‘strongly disagree’). A similar result was obtained for the question regarding the motivational traits. The acceptance of the use of sliders as a self-report method seems high given our initial expectations, but apparently students were not discouraged by having to update these sliders regularly.
Is the scale used to represent the different motivational factors appropriate? We can see that:

1. virtually all the possible values in the sliders’ scale were used;

2. more importantly, that the distribution of all of them (except for Relevance) resembles an inverted ‘U’ shape. If the extreme values of the slider’s scale were chosen very often by the students it could indicate that the scale would require more points in that direction. But in our case we can see that the peak of the scale value distribution is in all figures around the central or one of the adjacent points in the scale.

Thus, these two points seem to indicate that the scale used for this study is appropriate for our purposes, and that we should not subtract or add any more points to it.

How often were the sliders updated? There was a considerable variation in the number of times that the different motivational factors sliders were updated. The confidence motivational factor is the one that was updated more often (amounting to nearly a 25% of all updates). This was expected, and it is in accordance with the suggestion made by Briggs et al. (1996) that self-confidence is a factor easy to report. On the other side of the scale we see that relevance was updated very rarely (on average only 1.33 updates per student). The updates for the rest of the motivational factors were between these two extremes.

Although most of the participants would be willing to use self-report facilities if this would make the instruction more efficient (as seen above), it remained to be seen whether students’ acceptance of the self-report method had any influence on the number of updates to the motivational sliders that they made.

By analysing the data recorded we can readily appreciate that the variety of number of updates is very great, and there is no clear relation with the level of self-report acceptance. Actually, if we consider the average number of updates for each of the possible levels of acceptance, there seems to be, somehow surprisingly, a tendency to update less often the motivational slider as the self-report acceptance increases. Nevertheless, the diversity of the amount of updates is too great and it seems clear that the level of self-report acceptance alone can not explain this diversity.

When were the sliders updated? In the version of MOODS used for this study the motivational sliders were available at all times during the interaction. But when were they used by the participants?

The biggest part of motivational sliders updates happened during the last part of the lessons. In fact, a 49.4% of all updates were made during the 70-100% duration of lessons. This would point towards a self-report interface that is only made available towards the end of each lessons, although the analysis of the recorded data suggests a more elaborate picture.

- **Effort.** The updates to this slider clearly happened mostly at the end of the lessons. A detailed look at the data shows that a 47.3% of the updates were made during the 80-100% duration of lessons.

- **Confidence.** In the case of the ‘Confidence’ slider, the largest frequency of updates is also towards the end of the lesson (58.7% of updates in the 65-100% duration window), although there is also a substantial amount of updates at the beginning of the lessons (27% of updates in the 0-30% duration window).

- **Satisfaction.** The satisfaction slider was updated mainly around the middle and the end of the lessons (30.4% of updates in 30-55% duration window and 47.8% of updates in 70-100% window, respectively), but the difference in frequency is not as high as in the data for ‘Confidence’ and ‘Effort’. Actually, we can interpret that the satisfaction slider was mainly updated during the second half of the lessons (73.9% of updates in 45-100% duration window).
• **Sensory Interest, Cognitive Interest, and Relevance.** These three factors have fewer updates than the other motivational sliders and therefore it is more difficult to draw generalisations on their data. Nevertheless the three of them seem to have a similar pattern of usage, in the sense that most of the updates happen around the beginning and end of the lessons. (21.1% of updates in the 0-20% window and 42% of updates in the 70-100% window for ‘Sensory Interest’; 33.3% of updates in the 5-30% window and 24.2% of updates in the 80-100% window for ‘Cognitive Interest’; 16.7% of updates in the 0-10% window and 50% of updates in the 70-100% window for ‘Relevance’).

**Can we trust the information obtained through the motivational factors sliders?** Do the data collected really represent the participants’ motivational state? It is impossible to prove this, but we can approach this issue in the following way. The participants were presented with certain lessons organised in a number of instructional paths. By applying our pedagogical knowledge as teachers and the motivational theories presented in chapter 2, we can predict what the motivational state of the participants ‘should’ be at certain points during the interaction with MOODS. If none of these predictions match the data obtained through the self-report interface, we can assume that either our predictions or the data given by the students is not accurate. If some of the predictions are proved right by the data obtained, then we will have some assurances that the self-report interface provides, at least sometimes, useful information about the student motivational state that can be used by an ‘affective tutor’ to plan the next instructional step.

First we looked at the overall values of the motivational sliders according to the different instructional paths. As explained in section 4.1, the six instructional paths were designed attempting to create a ‘motivational continuum’, starting with a very demotivating one (number 1) and ending with a much more appealing one (number 6). From the four factors analysed4, our expectation seems to be confirmed mainly for cognitive interest and sensory interest. The other two factors (satisfaction and confidence) show a timid increase up to path 5 and a sharp decrease in path 6. The pattern for satisfaction is slightly puzzling, since the participants commented that they quite enjoyed the game, although we should note that except for the value of path 5, path 6 shows the highest ‘Satisfaction’ average of all the paths. Perhaps these data were influenced by the fact that the game in path 6 was actually quite engaging and participants did not concern themselves much with updating the slider for these factors.

More interesting still is to check if there are variations in the motivational state of the participants for individual lessons according to the path that they belong to. For example, the lesson “Writing20”5 appears in Instructional paths 2 and 5 (among others). In the path 2 this lesson comes immediately after the first theory lesson on Japanese numbers. On the contrary, in path 5 the student is given two easier practice lessons before coming to lesson “Writing20”. Therefore, we would expect the confidence level in the path 5 to be higher than in path 2.

In figure 3 we summarise graphically the data corresponding to this and two other predictions about participants’ motivational states:

(a) The lesson Theory100 (Th100), in which theory about the Japanese numbers up to 100 was explained, appears in all but the last instructional path. In paths 1 to 3 this lesson is presented immediately after the theory for small numbers. In contrast, in path 4 the participant performs two exercises on small numbers before moving to this lesson. And in path 5 he has three exercise lessons prior to this one. Therefore we would predict that his confidence for this lesson would be higher in paths 4 and 5, which is clearly the case, as can be seen in figure 3(a).

---

4*Relevance* applies more to individual lessons, and it is difficult to predict which values effort ‘should’ take, so these were not analysed.

5In this lesson the student is asked to write the Japanese number in text, given the number in figures.
(b) Similarly, the lesson Practic100 (Pr100) in path 5 is performed after the participant had the chance to practice with smaller numbers more than in the other paths and in a more coherent pedagogical style, so we would expect both his confidence and satisfaction to be higher, which is the case, as seen in figure 3(b). We should also note, though, that we get an unexpected result, as the confidence for this lesson in path 4 seems to be lower than for paths 1 and 3.

(c) The third prediction involves the lesson Writing20 (Wr20), which as for the other two examples, is presented in path 1 right after the theory lessons. In contrast, the student has more practice with easier exercises as the instructional path number increases. We can see the result of this in figure 3(c), where it is clear that the overall satisfaction was higher for the high number instructional paths.

4.4 Discussion

This study offered an important insight into the issue of the usefulness of the self-report method for motivation diagnosis. The results obtained from this study give useful clues as to how this method could be used in a real ITS, and we discuss them in this section.

One of the main doubts about the self-report method before conducting this study, was that of its acceptance. How would students react to it? Would it be used? Would it provide useful information about students’ motivational state? Given the results presented in section 4.3, we do believe that the method of self-report could be used satisfactorily for motivation diagnosis in ITSs. We have seen that participants of this study seemed to accept the use of the self-report interface and that its use provided (though not always) useful information about their motivational state.

Participants seemed to think that self-report could be a good method for communicating with the computer about motivational issues, and a method not intrusive on their learning. Although some of the students also commented that a longer period of use of the system may make them loose interest in using the self-report facilities.

The reaction of the system to the values of the motivational sliders may have a great impact on the willingness to continue using the self-report facilities. Since this version of the system didn’t react to the
motivational sliders values and the interaction was quite short, we can only speculate that although the method seems acceptable, care should be paid to making the reaction of the system to users updates of the sliders very obvious, in order to encourage the use of the self-report facilities.

Students’ perception of how appropriately the system reacted to her motivation model could be useful in a formal evaluation of a ‘motivating’ ITS. After all, as it is noted in the motivation literature, it is the feeling of control, rather than control itself, what seems to be important in motivating students. The perception that the system is reacting appropriately to one’s self-reported motivational state may be regarded as a kind of indirect control.

In this sense, it is interesting to note that some students commented that they changed some sliders in an exaggerated way to try to make the computer react to their inputs. In these cases, a speedy and obvious reaction would be very important, as it would give the students the feeling that the system reacts appropriately to their actions.

At the same time, the results obtained in section 4.3 suggest some changes in the self-report interface, which would make it more efficient and therefore more likely to be used by students. Several students commented how they mainly updated the sliders at the end of each exercise. Others commented how it was good to have them available at all times. For instance, the confidence could be high at the beginning of one exercise, but lower later during the same exercise, when realizing that the exercise was actually more difficult than expected. A compromise between these different approaches is outlined below.

1. We should try to avoid those factors that were poorly understood or barely used. In our study we have seen that “Relevance” was very seldom updated. Although providing relevant material is a major factor for creating motivating instruction, it seems clear from this study that to question students about each particular lesson’s relevance does not seem to be appropriate and/or useful. With a limited curriculum like the one presented in our MOODS prototype, all the lessons are relevant to the task of Japanese number learning, so the relevant motivational slider can be considered superfluous and it is probably best to consider the creation of relevant material simply as a curriculum design issue.

2. We could use the infrequently used factors for occasional problem detection rather than for continuous update by the students. In this sense, ‘sensory interest’ and ‘cognitive interest’ should be available to the student in case he wants to update them, but not as an integral part of the self-report interface. They should be hidden from the main self-report interface, but accessible to the student via a menu in case he wants to update them. This would make the interface much simpler, but offering similar informative capabilities.

3. Therefore, the main self-report interface could be reduced to the three factors that are easier to update and more regularly updated in our study: ‘confidence, ‘effort’ and ‘satisfaction’. But at the same time, we should make these available at the times when they are most likely to be updated, and when their interpretation is clearer. Thus, ‘satisfaction’ should be available at all times, being the overall index for student’s motivation. ‘Effort’ should be available at the end of each lesson, and ‘Confidence’ at the beginning and end of each lesson, in order to be able to appreciate discrepancies between the initial expectations of the students towards the lesson, and the confidence in his actual performance at the end of the lesson. Providing these factors at these particular times limits the expressiveness of the students, but gives a clearer interpretation of the meaning of each motivational factor slider.
5 Motivation Diagnosis study

Although self-report seems to be an adequate technique for motivation detection, it is also interesting to be able to understand how we could detect the motivational state of the student, given his interaction with an Instructional System. The issue of how human teachers detect their students’ motivation seems to be taken for granted, and has been virtually unexplored in AI-ED research. Introspection and observational studies could throw some light on this issue, but they may be of limited usefulness for motivation diagnosis in ITSs. In a ‘traditional’ instructional setting or any other social interaction there is an ‘overflow’ of information. An incredible amount of information is available through various communication channels, such as facial cues, intonation, posture, etc. Davis (1976). Many ‘cues’ that help us detect other people’s emotions are perceived unconsciously via these channels, which makes it difficult to elicit emotion detection knowledge.

In order to limit the number of sources of information available for knowledge elicitation, we designed a study in which a tutor will see exclusively the screen interaction of a student with an instructional system. That is, the tutor will be able to see in a computer screen only the interface of the instructional system which the student is manipulating.

We wanted to be able to extract and formalise tutors’ knowledge about motivation detection, and we expected that it would be easier for tutors to rationalise their motivation diagnosis knowledge in this setting than if they were presented, say, with video-recordings of tutoring interactions. At the same time, we believed that the knowledge thus inferred would also be easier to formalise in terms of the information available to the instructional system (such as time of interaction with system, mouse movements, etc.). We present the details of this study in the following sections.

5.1 Materials

The goal of this study was to understand and infer knowledge about the detection of students’ motivation during instructional interactions. In order to do this, we designed the study so that the participants would watch a number of recorded interactions of a student with MOODS, and they would be asked to infer and comment on the motivational state of the student during the interaction. For this, we developed A_MOODS, which can be used to replay the actions of a previous student interaction with MOODS and to predict his motivational state. The A_MOODS interface can be seen in figure 4.

The window to the left (with title MOODS v.1.0) is where the actions made by the student are replayed. Due to technical difficulties, the mouse movements are not replayed. In its place, and to facilitate the viewing of the student interaction, an arrow (around the centre of the window in figure 4) indicates the mouse movements.

The window to the right (titled Motivation model) consists of three frames.

1. The top frame is a representation of the student traits of the motivation model, which provides information about some general learning characteristics of the student. The values for these were obtained through the trait questionnaire of the self-report study (see section 4).

2. The middle frame contains three buttons in order to control how the student interaction is replayed.

   • The replay of the student interaction will starts by pressing “Play” and stops by pressing “Stop”.
   • The button “ Initialise Lesson” allows to ‘rewind’ the interaction to the beginning of the current lesson.

---

6 Originally published in part as (de Vicente and Pain, 1999a)
The replay of the interaction is done in real time, except for the replay of the theory lessons. In this type of lesson there is very little student activity, and therefore the system simply shows a message informing of how long the student took to learn the lesson (see figure 5), and afterwards the interaction will continue.

Figure 5: Sample message to inform of time spent in theory lesson.

3. The bottom frame is a representation of the student’s motivational state, as presented in section 3. The task of the participants for this study was to predict the likely values of these motivational variables as the instruction replay took place.

In order to let the participant update the student’s motivational model, the interaction stops by default in the following three cases:

1. When the student presses any of the buttons (Done, Give Up, or Help), but before any feedback is given by the system.
2. When feedback is presented to the student.
3. When a new lesson is presented to the student.

In these cases a small message similar to that in figure 6 will be shown to remind the participant to update the motivational model. But if the participant did want to update the motivational model or comment on any aspect of the instruction at any other time, he could stop the interaction by pressing the “Stop” button.
5.2 Methodology

For this study 10 post-graduate students with previous teaching and/or tutoring experience volunteered to participate. After reading the study instructions, the actual interaction with A_MOODS started, which can be summarised as follows:

1. The participant was presented with information about certain trait characteristics of a student (who had no knowledge of Japanese before the interaction with MOODS took place).

2. Then he was shown a replay of the student’s interaction with the MOODS instructional system.

3. Throughout the interaction, and particularly at any stop points, the participant was encouraged to give verbal comments (which were recorded for analysis) on the student’s motivational state and the possible factors affecting it.

4. Whenever the interaction was paused, the participant was asked to update the motivational state variables if he thought he had enough information to make an inference. At the same time, the participant was asked to verbalise the reasoning behind her inferences in terms as concrete as possible.

5. When the student pressed any of the three buttons available to him (Done, Give up, or Help) the participant was also encouraged to comment on the type of feedback that she thought would be the most appropriate to give to the student at that particular moment.

5.3 Results

It was a common comment among the participants before they started the study that the task was very difficult, and that it would not be possible to make any inferences based on the information provided. As we said earlier, in face-to-face teaching the tutor has an incredible amount of information at her disposal to infer about student’s affective state (his facial expressions, posture, etc.). But contrary to their own expectations, most of the participants made a considerable number of reasoned inferences about student’s affective state, and at the end of the study they commented that the task was actually not so difficult and that there was quite a lot of information available to them in order to perform these inferences.

The time devoted to each participant was 40 minutes, but some of them finished interacting with the system somehow sooner, while it took much longer to some others. On average, the time participants spent with the system was of around 36 minutes. In this time, an average of 4 lessons were covered, and 11 inferences were made per participant.

Although the number of inferences per participant is not very high, we can see that in total we collected a total of 109 inferences, which is a very valuable information to incorporate into our prototype Affective Tutor.

Let’s see first an example of a dialogue with one of the participants, in order to understand the type of information collected thanks to this study and the type of rules that were inferred from it.
**Excerpt**

**Participant:**

OK, well that was interesting cause he seemed to show hesitation at the start and the end, [...] But the middle ones he completed very quickly in general. I would say that this reinforces my belief that he is interested an confident, and he put effort into it. I would be inclined to say that he is satisfied, I still don’t think that you can say very much about the sensory interest.

**Interviewer:**

And [...] why do you think he is satisfied at this point?

**Participant:**

Well, because from the movements of the mouse, he is hovering the mouse over the answers each time, he wasn’t randomly moving the mouse, he is looking for the answer, and obviously thinking about what the right answer was, [...] and that he didn’t take a long time to answer the questions. To me that would suggest that the task is interesting enough to complete with some attention and to do it properly, if you like. [...] So, I would increase the satisfaction here, just for the fact that he did it with confidence,

**Inferred rule from Excerpt**

This excerpt illustrates a motivation diagnosis rule which infers about the satisfaction of the student based mainly on the interaction with the interface. Because the mouse movement through the interface is not at random, the participant could infer that the student was paying attention to the task. Because it was quickly performed, he inferred that he was confident. And given that:

1. He was interested in the task
2. He was confident
3. He performed the task well

the participant could infer that he would be highly satisfied. In figure 7 we present this graphically. We also have to note the dashed arrow and box on figure 7. This rule comes from another excerpt, but serves to illustrate the complex nature of the knowledge inferred through this study. We see that performing quickly a task can also mean lack of interest, but it is the combination of other evidence that can lead us to believe that in this case a quick performance was due to confidence. Similar conflicts will arise with other rules when trying to apply them in a real situation, and a way to resolve this type of conflicts has to be devised.
5.3.1 Ellicited motivation diagnosis rules

By analysing the rest of the data gathered from this study we ellicited one hundred and nine rules that could be incorporated in our Affective Tutor. After discarding repetitions, variations or subparts of other rules already present and those that were based on information that would be too difficult to implement into a computerised tutoring system, we ended up with a more manageable set of rules. Due to lack of space, we cannot present all the inferred rules in here, but we present as examples some of them in table 4. In this table we have attempted to be true to participants’ comments and therefore the rules are represented midway between the verbose explanations of participants and the necessary formalisation for their incorporation in an ITS. In order to make table 4 easier to understand, we have followed the graphical conventions in table 3 in order to represent the different parts that make each rule.

<table>
<thead>
<tr>
<th>Node</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Image" /></td>
<td>Steps where the student’s motivational models are involved</td>
</tr>
<tr>
<td><img src="image" alt="Image" /></td>
<td>Steps where interface issues are involved (e.g. moving mouse a lot)</td>
</tr>
<tr>
<td><img src="image" alt="Image" /></td>
<td>Performance issues (e.g. time required to perform exercise)</td>
</tr>
<tr>
<td><img src="image" alt="Image" /></td>
<td>Other intermediate steps (e.g. student finds exercise harder)</td>
</tr>
<tr>
<td><img src="image" alt="Image" /></td>
<td>Steps involving feedback to be given to student</td>
</tr>
</tbody>
</table>

Table 3: Graphical conventions for table 4

In table 4 we try to present a variety of rules, and each of them refer to different motivational factors. Thus, the first two rules deal with diagnosis of the ‘Satisfaction’ construct. Rule 1 refers to a situation where we could infer that student’s satisfaction is high, while in rule 2 the situation would indicate that student’s satisfaction is low. The rest of the rules refer to the factors of ‘confidence’, ‘effort’ and ‘interest’, respectively.
Table 4: Some motivation diagnosis rules

<table>
<thead>
<tr>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quickly performed</td>
</tr>
<tr>
<td>Well performed</td>
</tr>
<tr>
<td>Previous lesson wrongly performed</td>
</tr>
</tbody>
</table>

(1) 

<table>
<thead>
<tr>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very little filled in</td>
</tr>
<tr>
<td>Give up quickly</td>
</tr>
</tbody>
</table>

(2) 

<table>
<thead>
<tr>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficult exercise</td>
</tr>
<tr>
<td>Well performed</td>
</tr>
<tr>
<td>Spent time to answer (No simple guesses)</td>
</tr>
</tbody>
</table>

(3) 

<table>
<thead>
<tr>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long hesitation between answers</td>
</tr>
</tbody>
</table>

(4) 

<table>
<thead>
<tr>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well performed so far</td>
</tr>
<tr>
<td>He gave up halfway</td>
</tr>
</tbody>
</table>

(5) 

<table>
<thead>
<tr>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficult exercise</td>
</tr>
<tr>
<td>He spent some time</td>
</tr>
<tr>
<td>He left one undone and later went back to it</td>
</tr>
</tbody>
</table>

(6) 

<table>
<thead>
<tr>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Confidence</td>
</tr>
<tr>
<td>Low Satisfaction</td>
</tr>
<tr>
<td>Offered an easier exercise</td>
</tr>
</tbody>
</table>

(7)
Table 4: Some motivation diagnosis rules (continued)

<table>
<thead>
<tr>
<th>N.</th>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>(8)</td>
<td><img src="image" alt="Diagram" /></td>
</tr>
</tbody>
</table>

5.4 Discussion

Participants in this study were initially quite convinced that the task would prove extremely difficult and that it would be virtually impossible for them to extract any useful information about student’s motivational state without being able to see him. But, as we have mentioned earlier, this was not the case. The information available by looking exclusively at the computer interaction is enough to make a number of valid inferences about student’s motivation.

Furthermore, had we attempted to eliciting motivation diagnosis knowledge via, say, video-recordings of tutoring interactions, the amount of information available for participants would have increased substantially. They would be able to draw inferences from facial cues, gestures, intonation, etc. This might have opened the door to unconscious diagnosis which would have been more difficult to rationalise. By eliciting diagnosis knowledge only through the replayed interactions with MOODS we made sure that most inferences were properly rationalized and that they were based mainly on information which is readily available to an ITS (such as the order in which participants perform different exercises, time spent in doing them, mouse movements, etc.)

In conclusion, this study has provided a considerable amount of motivation diagnosis knowledge which can be incorporated into our Affective Tutor, and which is empirically based. Most of this knowledge is in accordance with the literature on motivation in education, but the subject is too complex to expect a set of rules to cover all situations. Thus, the rules inferred have obviously many gaps and in some cases they are contradictory, but we believe that by using a combination of self-report and these rules, we can provide a more robust approach to motivation diagnosis than by using only either of both approaches. How we combined these two approaches to create a prototype Affective Tutor is presented in more detail in section 6.

6 MOODS. A prototype Affective Tutor

In section 3 we presented an outline of the design for an ‘Affective Tutor’, but it is in this section that we give a detailed account of the design of MOODS, our prototype Affective Tutor. MOODS is a very simple tutoring system, but it incorporates all the knowledge gathered through the empirical studies presented in sections 4 and 5.

6.1 Structure Overview

MOODS is based on a basic cycle which can be seen at the bottom of figure 8. The basic steps in this cycle are the selection and presentation of a learning unit, the interaction of the student with MOODS (Performance) and the feedback given by the system and the student himself. These steps are represented by the thick outlined boxes, and the dotted arrows represent the chronological order of these steps during the interaction with MOODS. Once feedback for a particular unit is provided, MOODS goes back again
to the selection and presentation of a new learning unit, repeating this cycle until the interaction is finished either by MOODS or by the student.

In order to decide which lessons and which feedback to provide, MOODS makes use of a number of student models and a number of sets of inference rules. The student models available to MOODS are given in figure 8 in oval-shaped dotted boxes. These are:

- **Reported state model**: which is obtained via the self-report interface, which was discussed in section 4.
- **Inferred state model**: which is obtained via the motivation diagnosis rules discussed in section 5.
- **Combined state model**: since the reported state model and the inferred state model do not always coincide, a combined one has to be created, which is considered by MOODS as the most accurate model of the student.
- **Traits model**: this is a simple representation of student’s trait characteristics, which is obtained by a trait questionnaire given to the student at the beginning of the interaction.

The inference rules available to MOODS are given in figure 8 in trapezoidal grey boxes. These are:

- **Combination rules**: these are the rules to combine the reported state model and the inferred state model into a single one.
Motivation diagnosis rules: these are the rules to update the inferred state model, based on student’s performance and his feedback. Some of these rules were presented in section 5.

Motivational planning rules: these are the rules that will inform the system about which lesson to select at any given time, given the student’s traits model and his motivation state model.

Affective Dialogue rules: these are the rules to decide on the type and content of the feedback given to the student according to his performance, his motivation state model and his traits model. These rules play an important role on the overall structure of MOODS, and there are discussed in more detail in section 6.2.

At the same time all these inference rules make use of the interaction history information, where information about past lessons, performances, trends, etc. are kept.

6.2 The role of language in ‘Affective Tutors’

An “ideal” ITS should use language to communicate with its student. Although particular instructional units can be ingeniously designed to avoid language, a complete instructional interaction requires a number of tasks for which it would be very difficult not to use language. Among these tasks we find: presentation of new learning materials, performance feedback, help, etc. In general, it could be argued that language would be the most appropriate means of communicating about social and another aspects of the instruction. Administrative issues (when, where, with whom, ... to perform a particular task), affective issues (giving encouragement, empathising, ...), and many other aspects of instruction would seem very strained if non-verbal communication was to be used. This is so not only on the arts and humanities subjects. As du Boulay and Luckin (1999) comment, “[...] even in mathematics and science, despite all the ‘apparatus’ of representations and formal manipulations on those representations, an essential part of the ‘glue’ which binds what otherwise might be a fragmented understanding are conversations both with others and with ourselves.”

In what concerns the issue of motivation diagnosis, verbal communication offers some benefits over other types of motivation diagnosis:

Once a verbal interaction or dialogue is started, the student is somehow ‘forced’ to answer the questions posed by the system. This avoids a common problem of the self-report method (see chapter 4), namely that lack of student interaction with the self-report facilities can be misinterpreted as a constant value for the variables represented in the self-report interface.

Through verbal interaction it is possible to infer certain aspects of student’s motivational state, while it might appear to him that the goal of the dialogue is somehow different. For example, asking the student if he would like to continue with a lesson of similar difficulty could inform us about his confidence, but without asking directly about it. People tend to give a social treatment to machines (Reeves and Nass, 1998), which can lead to the student ‘lying’ about his motivational state if asked directly, in order to ‘please’ the ITS. By obscuring the actual purpose of the verbal interaction we might be able to avoid this, and therefore be able to obtain more accurate information about student’s motivational state.

In order to create an actual Affective Tutor we need to provide, in the words of du Boulay and Luckin (1999), the ‘glue’ to bind all the pieces of the tutoring system together. In this section we focus on how this ‘glue’, the verbal communication with the student, can be approached from an affective perspective, and how it can help to further improve the diagnosis of student’s motivation. We do not intend to present a psychologically valid model of affective educational dialogues or even to develop a fully functioning
conversational agent. Any of these tasks is outside the scope of this paper. Instead, we want to simply present an approach by which verbal communication can be seen as the binding glue between different aspects of an Affective Tutor.

Thus, we developed a model\(^7\) of Affective Educational Dialogues (AFDI), which is too simple to create realistic Educational Dialogues, but it offers a vehicle to combine the different aspects of our prototype Affective Tutor\(^8\). The model makes use of a database of sentences that are classified according to their content and their affective characteristics. At the same time, the database of sentences is divided into ‘Tutor moves’ and ‘Student replies’. For each tutor move, a number of possible student replies are provided. The selection of which particular tutor move to generate, or which possible replies to present to the student, is determined by the particular characteristics of the student and the history of the interaction, as we will see below.

As we can see in figure 8, AFDI plays a role in the whole MOODS structure, and below we explain briefly each of its roles:

**Providing feedback.** The feedback given to the student should be tailored not only to student’s performance, but also to his motivational state. An example of an AFDI rule that deals with this is given in table 5. For clarity, in this and following rules we represent the possible values for all the variables in the student model as: -10, -5, 0, 5 and 10, corresponding respectively to: very low, low, average, high and very high. Also, the performance column refers to the mark obtained in a particular instructional unit (from 0 to 100).

<table>
<thead>
<tr>
<th>RULE</th>
<th>VARIABLES</th>
<th>OUTPUT</th>
</tr>
</thead>
<tbody>
<tr>
<td>decide_feedback</td>
<td>performance</td>
<td>effort</td>
</tr>
<tr>
<td>75 100</td>
<td>—</td>
<td>&gt;= 0</td>
</tr>
<tr>
<td>75 100</td>
<td>—</td>
<td>&lt; 0</td>
</tr>
<tr>
<td>50</td>
<td>&gt;= 0</td>
<td>&lt;= 0</td>
</tr>
<tr>
<td>50</td>
<td>&lt;= 0</td>
<td>&lt;= 0</td>
</tr>
<tr>
<td>50</td>
<td>&lt;= 0</td>
<td>&gt; 0</td>
</tr>
<tr>
<td>50</td>
<td>&gt; 0</td>
<td>&gt; 0</td>
</tr>
<tr>
<td>0 25</td>
<td>&lt; 0</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 5: Provide feedback

Thus, we see that if the performance is high (more than 75% correct) and his confidence level is ‘Average’ or higher, we would simply provide positive feedback. But if his confidence was ‘Low’ or ‘Very Low’, we would provide not only positive feedback but also feedback geared towards increasing student’s confidence.

Given the characteristics of the desired feedback, the actual feedback can be then sought in the sentences database. In the present implementation of MOODS, possible options for positive feedback include “Well done!” and “Congratulations, you are doing great.”

**Detection of student’s motivational state.** As explained earlier, every tutor sentence has attached to it a number of possible student replies. The student can choose any of the replies given, and according to his selection, we can sometimes infer certain information that can be used to update his motivational model (this is complementary to the motivation diagnosis rules described in section 5).

---

\(^7\)An early version of which was published as (de Vicente and Pain, 2000)

\(^8\)See (de Rosis and Grasso, 1999; de Rosis \textit{et al.}, 1999) for an approach in which natural language generation techniques are used for affective text generation, and (Porayska-Pomsta \textit{et al.}, 2000) for more linguistically motivated research on generating teachers’ language in educational dialogues.
In table 6 we present an example of a motivation diagnosis rule in AFDI. This rule deals with the student reply to a positive feedback. In the current MOODS version, the student is offered two choices to reply to a positive feedback:

1. Thanks, it was easy.
2. Thanks, but it was hard!

Depending on the reply chosen by the student, we can infer the amount of effort that the student put into the task. If he considered the task easy, we would decrease the student’s inferred effort variable by 5 points. On the other hand, if he considered the task very hard, we would increase his inferred effort variable.

**Dealing with exceptional circumstances.** It is important to note that the rules described above affect only the inferred motivational model, while the student is responsible for updating the self-report model. As part of the design of MOODS, these two models can sometimes have different values, and the ‘Combination rules’ will combine both models into the ‘Combined state model’.

But if the motivation diagnosis rules are not adequate for a particular student or if he is not making use of the self-report facilities, it can happen that the values of the variables in both models are very different. In this case, we would ask the student directly to update the self-report facilities.

### 6.3 Instruction materials

If we were to test the many possible rule combinations presented in the previous sections, we would require a tutoring system with a large number of different instruction materials and we would need to imitate the behaviour of a real student. For example, we would need to spend a lot of time in some of the lessons if we wanted to simulate student’s hesitation.

In order to avoid this, we developed MOODS as a simulated tutoring system. That is, we developed MOODS with 100 simulated lessons. These lessons do not have actual content, but MOODS knows about their details (difficulty level, estimated time, etc.). With simulated lessons, we can also simulate the performance of the student (time taken to finish the instructional unit, percentage attempted, mark obtained, etc.), so that we can evaluate the functioning of MOODS’ knowledge in a faster and more convenient way. To illustrate this, a sample interaction of MOODS with simulated lessons can be seen in section 6.4.

A description of the simulated lessons created for MOODS can be seen in table 7. There are one hundred lessons divided equally into ten difficulty levels. For each difficulty level there are three theory instructional units, four exercise units and three game units. The arrows above the unit numbers indicate prerequisite dependencies, so that unit number four is a prerequisite for unit number five.

### 6.4 Example interaction

In order to clarify the functioning of MOODS, we here present and comment on a number of snapshots of the current version of MOODS.
Table 7: Simulated lessons description

When MOODS is started, it presents a small questionnaire in order to update the trait model for the simulated student. This can be seen in figure 9.

After filling the on-line trait questionnaire, the main interface appears. This can be seen in figure 10. In it there are three frames:

1. The top frame is where the instructional unit is displayed. As explained earlier, the current instructional units are simulated, so no content is provided. Instead, a description of the lesson is given. For the example given in figure 10 we can see that it represents: the first lesson of the curriculum; with a difficulty level of 1; with an estimated necessary time of 100 seconds; with no prerequisites; and the type of the lesson is theory.

2. The middle frame shows the verbal communication between the tutor and the student (the simulated student in the current version of MOODS). An example of the verbal communication is given later in figure 13.

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Theory</th>
<th>Exercise</th>
<th>Game</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>91</td>
<td>94</td>
<td>98</td>
</tr>
</tbody>
</table>

Figure 9: Trait questionnaire

<table>
<thead>
<tr>
<th>Trait Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONTROL: What is the degree of control that you like having over the learning situation (to select which exercises to do, in which order, etc)?</td>
</tr>
<tr>
<td>very low</td>
</tr>
<tr>
<td>INDEPENDENCE: What is the degree of work that you like doing independently, without asking for help from others?</td>
</tr>
<tr>
<td>very low</td>
</tr>
<tr>
<td>PAPTIC: What is the degree of fantasy (i.e., mental images of situations not present; such as games; etc.) that you normally enjoy during your instruction?</td>
</tr>
<tr>
<td>very low</td>
</tr>
<tr>
<td>CHALLENGE: What is the degree to which you enjoy having challenging situations during the instruction?</td>
</tr>
<tr>
<td>very low</td>
</tr>
</tbody>
</table>

Done
3. The bottom frame is where the motivation model is displayed. The motivation model follows the recommendations drawn from the self-report study. Thus, in figure 10 only the ‘Satisfaction’ variable is visible at this time. At other times during the interaction the other variables are made visible, so that the student would be encouraged to update his self-report, as can be seen in figure 11.

![Figure 10: Main interface](image)

The simulation of the student is performed by choosing one of a number of options representing his performance. In the case of theory lessons, the main performance variable is the amount of time spent in the lesson, and we can simulate the amount of time spent by pressing the button “Done” and choosing one of the five options, as seen in figure 12.

When a lesson is finished, MOODS provides feedback in the Dialogue frame. An example of this can be seen in figure 13, where MOODS has given positive feedback and the student can choose amongst the possible replies.
When the simulated lesson being displayed is a practical lesson, we also have to simulate the percentage completed and the mark obtained. This can be seen in figure 14.

7 Conclusions

In this paper we have presented a complete framework and a prototype implementation of a tutoring system that explicitly models the motivational state of its students. We originally based our research on the work of del Soldato (1994), and set up to study some of the issues that were not greatly explored in her work.

By reviewing the relevant literature, we reached the conclusion that in order to provide a useful motivation model, we needed to focus on only a well chosen number of motivational factors, some of which were more or less permanent characteristics (traits), while others were of a transient nature. Both types of factors required different methods for diagnosing them. Information about a person’s traits could be easily gathered via a simple questionnaire at the beginning of an interaction, but the transient factors needed another approach.

Self-report was considered as a method to diagnose student’s motivational state, but we performed an empirical study to: investigate whether self-report could really be viable as a motivation diagnosis technique; and if this was the case, to find out which approach to self-report should be the most appropriate to present to a student. We found that self-report was in general well received by the students, but analysis of the data recorded suggested a number of modifications to the original self-report interface in order to make it more suitable for detection of student’s motivational state.
But we also found that to rely exclusively on self-report would not be appropriate, as there are a number of situations when the student might not update the self-report facilities as frequently as we would require in order to adapt the instruction appropriately to his motivational state. Thus we set to study the possibility of formalizing knowledge about motivation diagnosis, which could be then implemented in an actual tutoring system. Previous work in this area seemed to be based mainly on intuition, and the relevant theories of motivation in education seemed to offer suggestions which were most of the time vague, too general, or too complex to implement in a tutoring system.

Therefore we performed another empirical study in which we presented participants with a previously recorded instructional interaction, and we asked them to try an infer the motivational state of the student at certain times during the interaction. We also asked them to formalize their reasoning in order to create a set of motivation diagnosis rules that could be easily implemented in a tutoring system.

By combining these techniques and a modified version of the motivational planning rules developed by del Soldato (1994), we created a framework for a prototype Affective Tutor: MOODS. As a way
of combining all these pieces and to show the importance of language in relation to affective issues of instruction, we also incorporated into MOODS a simple model of Affective Educational dialogues. This model helped us to bring unity to the different parts of the prototype Affective Tutor and also to provide some extra ways of detecting the motivational state of the student.

As a way of exploring the possibilities of the developed framework, we built into MOODS a number of simulated lessons and facilities to simulate the interaction of a student. Thanks to this, the knowledge incorporated into MOODS can be easily queried upon.

Obviously, there are many questions regarding Affective Tutors in general and motivation detection in particular that this paper has not addressed. For example, the possibility of using other channels to communicate with the computer (such as sound and video) offers many new areas that would be very interesting to investigate in relation to motivation detection. Another pending issue is clearly a formal evaluation of the validity of the motivation diagnosis techniques introduced, although this presents a whole range of issues that should be tackled separately. We are sure that many things in our approach to Affective Tutors can still be improved, but in this paper we have presented what we believe is a promising step into creating tutoring systems that care!

References

Affect in Interactions (1999). International Workshop on Affect in Interactions. Towards a New Generation of Interfaces (Annual Conference of the EC I3 Programme). Available at the World Wide Web: http://galva.inesc.pt/i3ws/i3workshop.html (October 1, 2001). Revised full versions of several papers and various selectively solicited papers by key people in the field can be found in (Paiva, 2000).


