Socially Intelligent Learner-Agent Interaction Tactics

W. Lewis Johnson, Sander Kole, Erin Shaw[1], Helen Pain[2]
[1] *Center for Advanced Research in Technology for Education (CARTE)*
USC / Information Sciences Institute
4676 Admiralty Way, Marina del Rey, CA 90292 USA
{johnson, skole, shaw}@isi.edu

[2] *School of Informatics, University of Edinburgh*
80 South Bridge, Edinburgh EH1 1HN, UK
H.Pain@ed.ac.uk

Abstract. The overall goal of this work is to provide pedagogical agents with social intelligence, so that they can judge when is an appropriate time to interact with the learner, be sensitive to the cognitive and emotional state of the learner, and try to develop a positive social relationship with the learner. This paper reports on a learner-agent interaction framework designed for use by socially intelligent agents. Interactions between a tutor and students working on computer-based exercises were videotaped and analysed, with particular emphasis on how the tutor sought to motivate and engage the learner. Interactions were analysed as a series of *interaction tactics,* where the speaker seeks to address one or more informational, motivational, or social goals, and monitors the listener's response to ensure that these goals are achieved. Parsing and generation frameworks implementing interaction tactics were developed, based upon the videotape transcripts, and evaluated by other tutors.

Topic: intelligent tutoring and scaffolding. Subtopic: agent-based interaction

Introduction

Animated pedagogical agents can promote effective learning in computer-based learning environments [13]. Learning materials incorporating interactive agents engender a higher degree of interest than similar materials that lack animated agents [17]. Animated agents can produce a positive affective response in the viewer, sometimes referred to as the persona effect [15] . They exploit the human tendency to treat computer systems as if they are social actors [19], capable of expressing emotions and attitudes.

Researchers have also come to recognize the importance in tutoring of recognizing the learner's affective and motivational states [14]. Computational techniques are being developed for inferring and tracking such states [22][3], and for influencing them, e.g., to try to promote a positive attitude about learning [21]. If such techniques were combined with animated agent technologies, it might then be possible to create an agent that can display emotions and attitudes as appropriate to convey empathy and solidarity with the learner, and thus further promote learner motivation.

The Social Intelligence (SI) Project is developing animated pedagogical agents with well-developed social skills, which they can employ to promote learning. The goal is to create agents that exhibit expressiveness (the ability to convey emotions and attitudes), empathy (sensitivity to learner motivational and emotional states), and politeness (an understanding of when and how to interact in socially appropriate ways). Such agents would then be able to tailor and adapt their style of interaction according to the needs, preferences, and affective and motivational states of each individual learner.

The current test bed for this work is the Virtual Factory Teaching System (VFTS), an on-line factory simulation used to teach industrial engineering concepts and skills [5][20]. An intelligent coach for scientific experimentation, ALI, has already been integrated with VFTS [6]. Nevertheless, learners who lack extensive industrial engineering background find the VFTS difficult to understand and use. We anticipate that by integrating a socially intelligent agent with the VFTS, particularly one that supports novice learners, a wider range of learners will benefit from working wih the VFTS.

This paper is concerned with developing an appropriate set of utterances that a socially intelligent agent can use to communicate with learners. An agent needs to know what to say and how to say it in order to convey and exploit its social intelligence. It also needs the ability to express its comments with an appropriate tone of voice, accompanied with appropriate gestures; these are also areas of active research at CARTE (Marsella and Gratch)[1](Johnson et al. 2002)[12], but are beyond the scope of this paper. Human tutorial dialogs were analyzed for this purpose, and some of the results of this analysis are presented here. The resulting dialog exchanges were implemented in an agent interface that is being used to emulate socially intelligent tutorial interaction.

1. Learner-Tutor Interaction Study

To investigate the role that social intelligence plays in learner-tutor interaction, we videotaped interactions between learners and tutor while the students were using the VFTS. Students read through an on-line tutorial in a Web browser, and carried out actions on the VFTS simulation as indicated by the tutorial. Learners were supposed to analyse the history of previous factory orders in order to forecast future demand, develop a production plan, and then schedule the processing of jobs within the factory in order to meet the demand. The tutor sat next to the students as they worked, and could interact with them as the student or the tutor felt appropriate. Completing the entire scenario required approximately two hours of work, divided into two sessions of approximately one hour. Three video cameras were used: one focused on the learner's face, one focused on the computer screen, and one providing a view of the learner and tutor together. This made it possible to track the learner's actions and focus of attention, as well as verbal and nonverbal interactions between the learner and the tutor.

Prior to the first session learners were given a brief questionnaire to assess their familiarity with manufacturing and business management techniques, and with computers. The Myers-Briggs Type Inventory (MBTI) was administered to each learner in order to assess personality characteristics. After the learners completed the tutorial and all the exercises the tutor had the learners complete a post-test relating to factory management. Finally, the student and tutor were each interviewed separately. The learners were questioned regarding subjective motivational and affective factors, e.g., how difficult or easy they found the material, how confident they felt, and whether they felt the work to be enjoyable or frustrating. They were also asked whether these factors changed over the course of the sessions. The tutor was asked to make similar assessments of the student, e.g., whether the student felt confident, frustrated, etc. The tutor was also asked to explain what led him to draw these conclusions.

The tutor in this preliminary study was an industrial engineering professor who had won awards for excellence in teaching, and who uses the VFTS in his courses. We plan follow-up studies with other tutors with varying skills, as well as with pedagogical agents, in order to assess the generality of the results of this study. Two students participated in the study. One was an electrical engineering student who was familiar with working computers and solving engineering problems, but did not have expertise in industrial engineering per se. The other student was a business major who had little experience with engineering problems, but was ultimately able to apply her knowledge of business to the industrial engineering problem.

The interactions in this experiment differ in a number of respects from those observed in most previous studies of tutorial dialog. First, the tutor was there to coach the learner as needed, rather than act as tutor per se. The tutor was aware of the instructional objectives of the exercise, and might interrupt to assist the learner with a particular topic if the tutor deemed it necessary, but most of the time the tutor would respond to the learner's questions, or offer advice and hints about the learner's problem solving activities. This contrasts with tutorial dialog studies such as those of Graesser, Person, et al. (e.g, [9]) and Chi et al. [4] in which the tutor leads the dialog via a series of questions. Furthermore, the learner was interacting with both the VFTS and the tutor. The VFTS does not critique the learner's actions, but it does respond to them immediately, which can help the learner to tell whether he or she is making progress. Other researchers such as Merrill et al. [16] have studied tutorial dialog in the context of problem solving, but there the tutor is the sole source of feedback for the student, e.g., because the student works out the problem on paper. In the VFTS context the tutor must decide not just how to guide the learner, but also when. This is similar to the problem that a pedagogical agent faces in deciding when to interrupt a learner with advice or feedback.

To analyse the interactions, and use them in designing learner-agent dialog, we transcribed them and annotated them using the DISCOUNT scheme [18]. DISCOUNT represents the structure of educational dialogs as a series of *episodes*, each pertaining to a particular topic. Episodes are divided into *exchanges* between the parties in the dialog, which are composed of a series of *turns* (e.g., initiate, respond, reinitiate). Each turn consists of one or more dialog moves, classified according to speech act (hint, support, contradict, etc.) and marked with predicate labels that indicate the function of the move in the dialog.

DISCOUNT was intended to model continuous dialog. The dialogs in this study, however, are not continuous, but are interspersed with periods where the student is working without talking. Nevertheless, DISCOUNT proved useful in making explicit the structure of the dialog and the relationships between dialog moves, and made it relatively easy to compare similar dialog moves, to see the different ways in which moves are realized. Since most exchanges were followed by a pause while the learner continued her work, these pauses were used to structure the dialog into episodes; this contrasts with uses of DISCOUNT in continuous dialog where the episode boundaries are signalled by topic shifts.

2. Overall observations

Although the tutor did not explicitly ask the learners questions about their motivational state, e.g., their level of confidence, he reported that he was able for the most part to assess this, and the tutor's assessments agreed with the learners' self-reports during the post-session interviews. One learner started out with little confidence in her ability to solve the problem, and grew in confidence over time; the other had a high degree of confidence throughout. The tutor was able to recognize when the students were growing frustrated, and would intervene to offer help. He could tell that one student was interested in the material, but had more difficulty assessing the other student's level of interest, commenting that she was "just the type of student that you wouldn't be able to identify" in this regard.

The tutor also noted different working styles between the students, which were consistent with their self-reports. One student preferred to work out problems on her own, whereas the other preferred to work collaboratively with the tutor. The tutor took the individual differences into account in deciding when and how to interact with the learners. With the student who preferred to work on her own, the tutor would refrain from intervening unless the student was clearly stuck. Ironically, this student did not perform as well on the post-test, so perhaps the tutor should have intervened more to probe her understanding. The learners differed significantly in their scores on the Myers-Briggs Type Inventory, raising the possibility that it might be useful for predicting these differences.

Analysis of videotapes revealed a number of cues that the tutor used to assess learner motivation and decide when and how to intervene. The types of questions posed by the learners were important, as well as their frequency; one student asked many more questions about terminology, and repeatedly asked for confirmation that she was doing the right thing. The tutor noted when the learners pressed the wrong button on the VFTS, or pressed the same buttons repeatedly. Conversely, the *failure* to press a button was indicative; if the tutorial instructions called for a particular action to be performed on the VFTS, and the student failed to take any action, the tutor might intervene. This required the tutor to track the learner's focus of attention, and infer effort and intention from it, e.g., to assess how much effort the learner was employing to read and understand the written materials. The speed at which the learner carried out tasks was also important, but not uniformly so; the tutor wanted to see the learners read the written materials carefully, and then work quickly with the VFTS. Many of these cues are amenable to automated analysis, and we have taken what appear to be the most significant ones into account in the design of the new agent.

3. Interaction Tactics

The tutorial exchanges can be characterized as a series of *interaction tactics*, where each tactic is intended to communicate particular information or have a particular effect on the listener. When a dialog exchange is initiated the speaker looks to see that the intended effect has been achieved, and if not rephrases the comment accordingly. The following example illustrates this, where the tutor is trying to get the student to choose which type of analysis to perform.

Tutor: Wanna use regression? You've been using regression, why not stay with regression?

Student: Huh?

Tutor: Which are you going to use? The three techniques are expansion smoothing, regression, and moving average. Which one?

Student: I'm supposed to use regression, right?

Tutor: Might as well. You've been using it.

Student: OK.

Since the tutor in particular frequently needed to rephrase comments when the learner did not understand them, we find in the transcripts multiple utterance variants all intended to achieve the same communicative goals. Further examples of these can be found in [11].

The tutor was often trying to achieve a combination of informational and motivational goals simultaneously. Comments and suggestions were phrased so as to engage the learner, reinforce the learner's sense of control, and encourage metacognition. For example, hints and suggestions were sometimes phrased as questions about what the learner wants to do, e.g.:

Tutor: Want to look at your capacity?

Direct instructions were quite rare, and mainly concerned with operating the user interface. Instead the tutor spoke of what the learner could do, or what the tutor would do if he were making the decision. Other comments implied joint decision making, e.g.,

Tutor: So why don't we go back to the tutorial factory…

Likewise direct critiques of the student performance were uncommon. Instead, the tutor tried to encourage the learner to critique her own work, e.g.,

Tutor: So you're happy with that?

Although many tutor comments had motivational aspects, there were few direct expressions of praise and encouragement, such as "You did good" or "There you go." These were addressed only to the learner with low confidence, and expressions of praise occurred only at the end of each session. Equally rare were confirmations that the learner was on track, like "right" and "OK". This differs from the tutors studied by Fox [8], who provided continual confirmatory feedback. Our previous pedagogical agents Steve and Adele [13] also gave continual confirmatory feedback: we must re-evaluate the desirability of such feedback.

4. Implementation of Interaction Tactics

Based upon these analyses, an implementation of interaction tactics was incorporated into the user interfaces that the SI Project is currently developing. The SI learner interface is designed to facilitate two-way communication between the learner and the agent. The learner's screen includes a hypertext window for viewing tutorial materials, a problem solving window (i.e., the VFTS simulation control panel), the agent persona, and a text type-in window. A camera placed on the top of the learner's display is focused on the learner's face. The agent has access to multiple sources of information for monitoring the learner. The camera is used to track the position of facial features, particularly the eyes. The interface sends a notification when a paragraph enters or leaves view, or when an object is moused on the screen. We are currently integrating these sources of information to estimate the learner's focus of attention. The learner can also send comments and questions to the agent. If the learner mouses on a key word or phrase in the tutorial text, she can select from a menu of questions relating to that topic. Alternatively, the learner can type a question or comment into a text window.

The learner interface currently communicates with an experimenter interface running on another computer, designed for Wizard-of-Oz testing of learner-agent interaction. The experimenter can view the learner's screen activities, using NetMeeting, and send and receive messages to and from the learner. The agent speaks the comments sent by the experimenter, using test-to-speech synthesis software.

```
<move>
      <!-- 7S P1 47:11  T  -->
<!-- So number 2, the number of seasons may not be 2 then.-->
      <predicate role="initiating" move="all" name="action1"/>
      <predicate role="initiating" move="all" name="noun1"/>
      <predicate role="initiating" move="hint" name="suggest"/>
      <predicate role="initiating" move="inform" name="identify"/>
      <predicate role="initiating" move="reason" name="explain"/>
      <template>
            So <nounphrase case="object" type="parameter"
                         name="noun1.nounphrase1"/>
            may not
            <verbphrase type="parameter" name="action1.action1"
                         form="infinite"/>
            .
      </template>
</move>
```

Figure 1. An example dialog move template

The underlying generation scheme utilizes DISCOUNT as a way of classifying dialog moves. It operates on a set of move templates, each of which includes a set of DISCOUNT predicates and a template for expressing the move in natural language. The templates can specify language elements, which are filled in during the generation process with values selected by the experimenter. The move templates and language elements are specified using an XML syntax and all defined in one language definition file. **Error! Reference source not found.** shows an example move from the currently used language definition file. The moves are based upon utterances found in the transcripts; the comments at the top of the move template show the original utterance and the transcript and time code where it was found. The move template may classify the move in multiple ways, reflecting the fact that the same utterance may have multiple communicative roles, and different coders may code the same utterance differently.

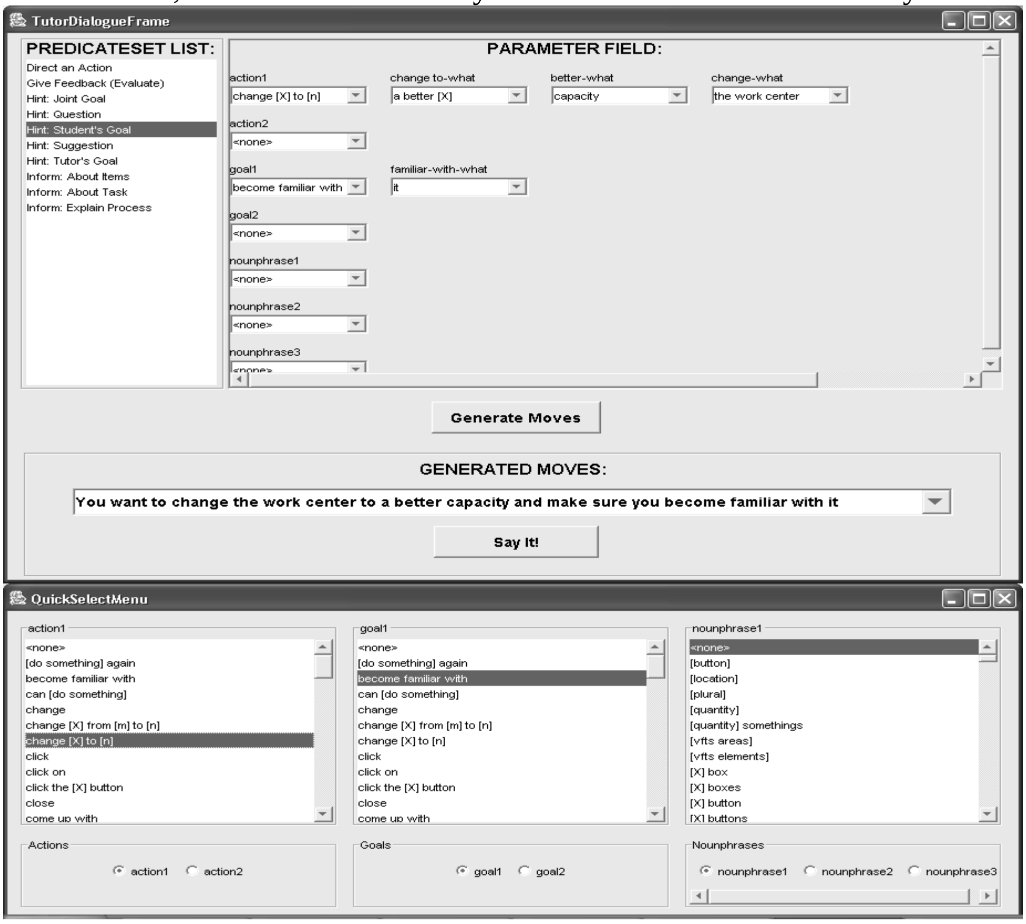

Figure 2. Experimenter interaction tactic interface

Our goal is to have the agent select utterances based upon its communicative intent, expressed in DISCOUNT terms. As DISCOUNT is not a very simple or efficient formalism for a human tutor to use in describing an utterance to be generated, we have developed a set of dialog move pattern templates for constructing the desired move, as shown on the left column in **Error! Reference source not found.**. These combine primary communicative intent with secondary motivational intent; for example, "Hint Tutor's Goal" means give a hint of what the tutor would do in the current situation. This approach has the benefit of making explicit the sophisticated interaction tactics that our expert tutor employed, thus codifying tutorial expertise.

The natural language generator uses the move template predicates and natural language elements (noun phrases and verb phrases), and generates a set of possible utterances to say. In the event that there is no utterance template in the language definition file that matches the selection criteria exactly, those that partially match the selection criteria are considered,. Four matching criteria are used for ranking these: (1) closeness of match of move predicates, (2) degree of correspondence between natural language elements chosen by the experimenter and natural language elements used in the utterance, (3) recency criteria, to avoid repeating the same utterance, and (4) a random factor to add variety to the agent's language and keep the agent from repeating himself. The five most highly ranked options are presented to the experimenter as options. The experimenter then selects an option and edits it if necessary.

A natural language parser has also been developed, utilizing a technique adapted from eDrama FrontDesk [7]. When the user types a sentence, the sentence is canonicalized, matched against move templates, and the closest matches are offered to the user as options. We have incorporated this into the experimenter's interface, and we plan to add it to the learner interface as well, so that the learner can converse with the agent in natural language.

5. Validating the Dialog Generation Model

Do the generated utterances allow a tutor, or agent, to say everything it would like to say in the context of the VFTS? The goal of the next experiment was to compare interaction tactics across different tutors and either confirm that there are no gaps in the dialogue generation model, or add new speech acts to the model if there are gaps.

This experiment was a more focused version of the preliminary study. The tutor was a developer of the VFTS and also a female (versus a male in the initial study), and the student was a liberal arts major who had little experience with engineering problems. A second tutor, a graduate student and expert user of VFTS, will conduct a second trial. As in the initial study, the tutor sat next to the student and provided help in a way that was most natural to her. The dialogue was recorded and transcribed, and a screen recording was made using Camtasia Studio to provide context when there were questions about the dialogue.

Utterances that the tutor initiated, and responses that might be initiated, were classified into the ten high level categories, listed in Figure 2, "Direct an action", "Give feedback", etc.,. We then attempted to generate the utterances using the tool. Of the twenty two utterances analyzed, 3 were almost exact matches; 3 were good matches; 7 were matched, but in an awkward way, or way that slightly altered the meaning of the utterance, and 9 were not matched, or matched so poorly as to be unacceptable. The awkward matches typically used the same words but in a different way, e.g. "generate the plan" was matched by "do generate plan" and "run the planning". The utterances that did not match were typically missing a key word that the preliminary study tutor didn't use, and that couldn't be easily substituted, e.g. "compare" and "paragraph". When these situations arose, we searched the raw templates to verify that the key word was not available to the interface. This is not a serious limitation since new terms can be added on the fly via the experimenter interface. When the interface is configured for use by an agent, care will be taken to ensure that the interface has sufficient vocabulary to cover the domain of discourse.

6. Related Work

Several researchers are looking at the how richer models of learner-agent interaction can address social aspects of communication, in part to promote learner motivation [1][2]. However, these other efforts draw a distinction between task-oriented dialog and social- or motivation-oriented dialog moves, and suggest a model where different types of utterances are interspersed. Our data suggests instead that social and motivational factors have a pervasive effect on tutorial interaction, and a pedagogical agent must continually consider social and motivational factors when interacting with learners.

7. Conclusions and Next Steps

This paper describes initial efforts toward the development of pedagogical agents capable of socially intelligent communication, based upon empirical studies of learner-tutor interaction. These studies revealed a sophisticated use of interaction tactics to achieve multiple communicative and motivational goals. A natural language generator was developed that generate dialog moves to implement these interaction tactics. As next steps, we must test the repertoire of verbal dialog moves by applying them to other learning tasks, and extend the interaction model to include nonverbal gestures. We plan to run Wizard-of-Oz studies to compare learner-agent interaction and human-agent interaction in similar learning contexts. Based on these tests we can then proceed to develop a pedagogical agent that can select appropriate interaction tactics, observe the learner's response, and adapt tactics as needed.

Acknowledgments

Thanks to Wauter Bosma, Cynthia Chan, Maged Dessouky, Jason Finley, Lei Qu, and other members of the SI Project for their help with the work presented here. This work was supported in part by the National Science Foundation under Grant No. 0121330, and in part by a grant from Microsoft Research. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

References

[1] Aist, G., Kort, B., Reilly, R., Mostow, J., Picard, R.W. (2002). Adding human-provided emotional scaffolding to an automated reading tutor that listens increases student persistence. Intelligent Tutoring Systems 2002: 992.

[2] Cassell, J. and Bickmore, T. (2001). Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. *User Modeling and Adaptive User Interfaces*, in press.

[3] Conati, C. & Zhou, X. (2003). A probabilistic model to assess user affect from goals, personality, and context. *Proc. of the Intl. Conf. On Intelligent User Interfaces*. New York: ACM Press.

[4] Chi, M.T.H., Siler, S.A., Jeong, H., Yamauchi, T., & Hausmann, R.G. (2001). Learning from human tutoring. *Cognitive Science*, *25*, 471-533.

[5] Dessouky, M.M., Verma, S., Bailey, D., & Rickel, J. (2001). A methodology for developing a Web-based factory simulator for manufacturing education. *IEE Transactions* 33, 167-180.

[6] D'Souza, A., Rickel, J., Herreros, B., & Johnson, W.L. (2001). An automated lab instructor for simulated science experiments. In J.D. Moore, C.L. Redfield, and W.L. Johnson (Eds.), *Artificial Intelligence in Education*, 65-76. IOS Press, Amsterdam.

[7] EDrama Learning™ (2002). White paper, Natural language in simulated conversation.

[8] Fox, B.A. (1991). Cogntiive and interactional aspects of correction in tutoring. In P. Goodyear (Ed.), Teaching knowledge and intelligent tutoring, 149-172. Norwood, NJ: Ablex.

[9] Graesser, A.C., Person, N., & Magliano, J. (1995). Collborative dialog patterns in naturalistic one-on-one tutoring. *Applied Cognitive Psychology*, *9*, 359-387.

[10] Gratch, J. & Marsella, S. (2001). Tears and fears: Modeling emotions and emotional behaviors in synthetic agents. In *Proc. of the 5th Intl. Conf. on Autonomous Agents*, 278-285. New York: ACM Press.

[11] Johnson, W.L. (2003). Interaction tactics for socially intelligent pedagogical agents. In. *Proc. of the Intl. Conf. on Intelligent User Interfaces*. New York: ACM Press.

[12] Johnson, W.L., Narayanan, S., Whitney, R., Das, R., Bulut, M., & LaBore, C. (2002). Limited domain synthesis of expressive military speech for animated characters. In *Proceedings of ICSLP Workshop on Text-to-Speech Synthesis*.

[13] Johnson, W.L., Rickel, J.W., and Lester, J.C. (2000). Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments. *International Journal of Artificial Intelligence in Education* 11, 47-78.

[14] Lepper, M.R., Woolverton, M., Mumme, D., & Gurtner, J. (1993). Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In S.P. Lajoie and S.J. Derry (Eds.), *Computers as cognitive tools*, 75-105. Hillsdale, NJ: Lawrence Erlbaum Associates.

[15] Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., and Bhogal, R. S. (1997). The persona effect: Affective impact of animated pedagogical agents. In *Proceedings of CHI '97*, 359-366.

[16] Merrill, D.C., Reiser, B.J., Ranney, M., & Trafton, J.G. (1992). Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems. *J. of the Learning Sciences 2*(3), 277-305.

[17] Moreno, R., Mayer, R.E., Spires, H.A., & Lester, J.C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction* 19(2), 177-213.

[18] Pilkington, R.M. (1999). Analysing educational discourse: The DISCOUNT scheme. Technical report 99/2, Computer-Based Learning Unit, University of Leeds.

[19] Reeves, B. and Nass, C. (1996). The media equation. New York: Cambridge University Press.

[20] Rickel, J., Johnson, W.L., ... (2003). AIED paper, submitted.

[21] Del Soldato, T. and du Boulay, B. (1995). Implementation of motivational tactics in tutoring systems. *J. of Artificial Intelligence in Education* 6(4), 337-378.

[22] De Vicente, A. and Pain, H. (2002). Informing the detection of the students' motivational state: An empirical study. In S.A. Cerri, G. Gouardères, & F. Paraguaçu (Eds.), Intelligent Tutoring Systems: 6th International Conference, ITS 2002, 933-943. Berlin: Springer-Verlag.