Emotional intelligence for computer tutors

Beverly Woolf,[1] Winslow Burelson,[2] Ivon Arroyo[1]
[1] Department of Computer Science, University of Massachusetts Amherst
[2] Department of Computer Science and Engineering/Arts, Media and Engineering
Arizona State University

Abstract

We are developing intelligent tutors that elicit, sense, communicate, measure and respond to students' emotion. The vision is to use affect in conjunction with cognition to customize instruction and exploit the potential impact of responding to a student's affective state. This paper describes software and hardware solutions to redress the cognitive vs. affective imbalance to teaching. We investigate problems of sensing and responding to affective experiences, such as frustration, motivation, and self-confidence. We describe theories and technologies to appropriately integrate affect into intelligent tutors and are investigating how to respond to emotions.

1. Connecting emotion to learning

The obvious next frontier in computational instruction is to systematically examine the relationship(s) between *affective* components in feedback and learning outcome (performance) (Shute 2006). If computer tutors are to interact naturally with humans, they need to recognize emotion and express social competencies. Emotion is completely intertwined with cognition in guiding rational behavior, including memory and decision-making and the human brain is described as a system in which emotion and cognitive functions are inextricably integrated (Cytowic 1989). Emotion has been shown to be more influential than cognitive abilities for personal, career and scholastic success (Goleman, 1996) and understanding human emotion is one of the greatest challenges of our time (Norman, 1981). For instance, impulsivity was twice as powerful a predictor as verbal IQ in future delinquent behavior (Block, 1995). Recent findings suggest that too little emotion is not desirable. When basic mechanisms of emotion are missing in the brain, intelligent functioning is hindered.

Teachers recognize the central role of emotion, devoting as much time to the achievement of students' motivational goals as to their cognitive and informational goals (Lepper and Hodell, 1989). Students who feel anxious or depressed do not assimilate information properly (Goleman, 1996). These emotions paralyze the so-called 'active' or 'working memory,' which sustains the ability to keep information about a task (Baddeley, 1986); in fact, some studies found that the level of worry of failure is directly proportional to achievement level (Ortony et al., 1988). Students with high intrinsic motivation often outperform students with low intrinsic motivation. Caring relationships have also been shown to be related to academic performance. The encoding of affect within human-human interactions is very powerful. In their research on "thin slices," Ambady and Rosenthal demonstrate that when participants in their studies are shown a short segment of video, as little as six seconds of a teacher's first interactions with their student, they can predict teacher effectiveness and student end of term grades (Ambady and Rosenthal, 1993).

Curiously, however, theory and classroom practice have tended to privilege the 'cognitive' over the 'affective,' where affect is often ignored or marginalized in theories of learning that often view thinking and learning as information processing (Picard et al., 2004). The extension of cognitive theory to explain and exploit the role of affect in learning is at best in its infancy. This research aims to redress the imbalance between cognition and affect by developing theories and technologies to appropriately integrate both in human-computer tutors. We exploit the potential on learning outcome of responding to a student's affective state and investigate the problems of sensing and responding to affective experiences, such as frustration, motivation, and self-confidence. Research questions include:

- How do emotions predict learning?
- How is emotion evidenced in student behavior? Are these dependencies consistent across teaching environments, e.g., open-ended critical thinking vs. problem solving systems?
- How accurate are different models at predicting emotions from student behaviors? How do latent Bayesian or hidden Markov models compare to other models (such as plain regression)?

- How effective are interventions at changing specific emotional states? Can machine learning learn optimal policies for improved long-term student attitude and learning?

This paper describes hardware and software solutions to address these research challenges. Section 2 describes technologies to recognize emotion, first hardware solution and then software solutions. Section 3 suggests ways to respond to students once a particular level of emotion is recognized. Section 4 reviews affect in education and Section 5 outlines prior research to detect emotion in intelligent tutors. Section 6 proposes a framework for using affective feedback in two existing intelligent tutors.

2. Recognizing student emotion

We have developed several methods to recognize student emotion, including a hardware research platform comprised of validated instruments and hardware technologies (e.g., sensors) to model emotions and software platform that uses machine learning techniques to reason about student affect.

Hardware solutions to recognize student emotion. Our research platform of hardware technologies includes four sensors (camera, posture sensing devices, skin conductance wristband, and pressure sensitive mouse) and this information is analyzed along with the learning task and participants' interactions to train classifier algorithms, Figure 1. Interventions are provided to access their impact on providing affective support and helping learners.

Facial Expression Camera: We use an in-house camera and software system based on strategies leaned from working with the IBM Blue Eyes Camera that tracks pupils unobtrusively using structured lighting which exploits the red-eye effect to track pupils (Haro et al., 2000). Pupil positions are passed to a method that detects head nods and shakes based on Hidden Markov Models, Figure 1a (Kapoor & Picard, 2001). We train an HMM that uses the radii of the visible pupil as inputs to produce the likelihoods of blinks.

We recover shape information of eyes and eyebrows[1] (Kapoor & Picard, 2002). Given pupil positions and facial features we localize the image around the mouth and extract two real numbers corresponding to two kinds of mouth activities, smiles and fidgets. A large difference in images is treated as mouth fidgets. We look at the sum of the absolute difference of pixels of the extracted mouth image with the images in the last 10 frames and use a support vector machine (SVM) to compute the probability of smiles using natural examples of mouth images. The resulting output is passed through a sigmoid to compute smile probability. The system extracts features in real time at 27-29 frames per second on a 1.8 GhZ Pentium 4 machine and tracks well as long as the student is in the reasonable range of the camera. As children move a great deal, it is important to have a system that is robust to movement.

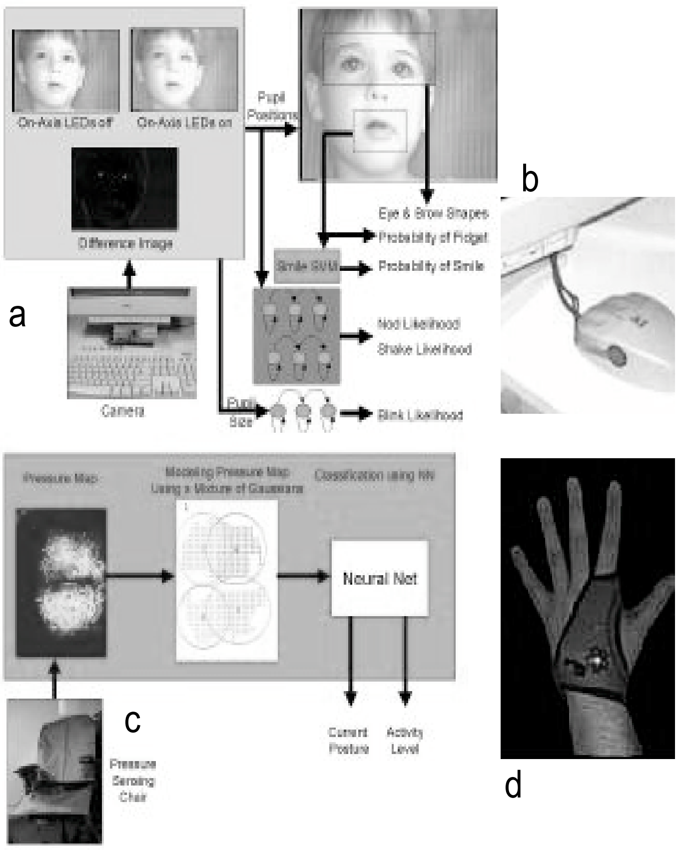

Figure 1. Sensors that collect physiological and motion data. Blue eyes camera (a), pressure mouse (b), posture sensing device (c), skin conductance glove (d).

Posture Sensing Devices: We detect student postures using matrices of pressure sensors made by Tekscan. The sensors detect a static set of postures, e.g., sitting upright, leaning back, and activity level, e.g., low, medium and high. One matrix is positioned on the seat-pan of a chair; the other on the backrest, Figure 1c. Each matrix is 0.10 millimeters thick and consists of a 42-by-48 array of sensing pressure units distributed over an area of 41 x 47 centimeters. This variable resistance, in which the normal force applied to its superficial area determines resistance, is transformed to an 8-bit pressure reading, is interpreted as an 8-bit

[1] Sophisticated web-cams can detect many of these same facial features quite robustly and for lower cost.

grayscale value and visualized as a grayscale image. First, the pressure maps sensed by the chair are pre-processed to remove noise and the structure of the map was modeled with a mixture of Gaussians. The parameters of the Gaussian mixture (means and variances) are used to feed a 3-layer feed-forward neural network that classified the static set of postures in real time at 8 frames per second, which are then used as posture features by the multimodal affect classification module.

Pressure Mouse: We use a Pressure Mouse with eight force-sensitive-resisters that captured the amount of pressure placed on the mouse throughout the activity, Figure 1b (Reynolds, 2001). Users who find an online task frustrating often apply significantly more pressure than do those who do not find the same task frustrating (Dennerlein et al., 2003).

Wireless BlueTooth Skin Conductance: A wireless version of a glove that senses skin conductance is used, Figure 1d (Strauss et al., 2005). While the skin conductance signal does not explain anything about valence - how positive or negative the affective state is - it does tend to correlate with arousal, or how activated the person is. A certain amount of arousal is a motivator toward learning and tends to accompany significant, new, or attention-getting events.

Completed studies. We conducted experiments to recognize and respond to frustration. Non-invasive multimodal real-time devices were used to sense a student's affective state and were coupled with an agent capable of supporting learners by engaging in real-time responsive expressivity, Figure 1 (Burleson, 2006; Kapoor et al., to appear). The agent was coordinated with the hardware to mirror non-verbal social behaviors that influence persuasion, liking, and social rapport and responded to frustration with empathetic or task-support dialogue.

Students were engaged in the Towers of Hanoi activity and two non-verbal interactions conditions were developed: sensor driven 'mirroring' interactions and pre-recorded interactions. Classifier algorithms predicted frustration with 79% accuracy. This research developed a theory for using affective sensing and appropriate relational agent interactions to support perseverance through failure. It focused on meta-cognitive awareness and personal strategies. For example, it sought to provide students with awareness of their affective state and to help them understand failure and to develop the motivation to move onward.

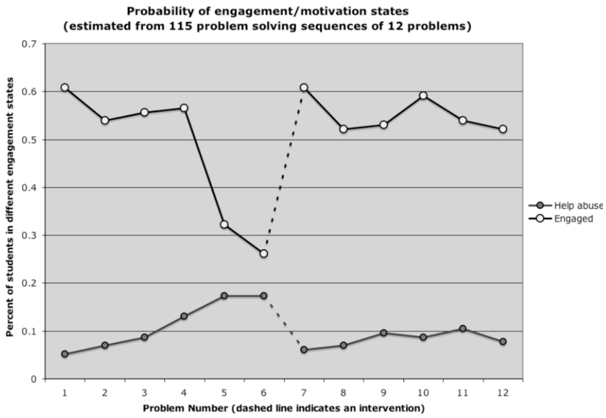

Figure 2. Students became motivated after feedback. Student engagement declined after the 6th problem (top). After an intervention was presented on-target student performance returned. Help abuse mirrors this behavior (bottom).

Software solutions to recognize emotion. In addition to the complex hardware research platform described above, we use software techniques (e.g., machine learning, Bayesian nets) to track emotion. Most prior work on emotion recognition has focused on deliberately expressed emotions within a laboratory setting and not in natural situations such as classroom learning. The studies described here were conducted in classrooms as part of a regularly used intelligent tutor.

Student models have benefited from both supervised and unsupervised machine learning methods (Johns & Woolf, 2006). We used hierarchical Bayesian networks to represent the structured nature of lessons and hidden variables were used to model relationships among skills that were inferred but not directly observable. These systems automatically created classification or prediction rules from a collection of data. The training set (collection of examples) typically consisted of labeled data, e.g., estimates of a student's abilities, the student's progress in solving the current problem.

In one study we used a Bayesian net to infer emotion, using a student's observed problem-solving behavior and based on estimated from surveys filled out by prior students (Arroyo & Woolf, 2005). Bayesian networks were used to discover links between observable behavior (time spent on hints, number of hints selected) (Arroyo et al., 2004; Arroyo & Woolf, 2005) and emotion. The probability of being correct about a student's affective state (e.g., predicting a student's response about motivation as shown in the post-survey) has been measured within a window of 80-90%. We correlated observable student activities and survey responses, converted this into a Bayesian network and then tested the predictions on the log data of new students. Affective hidden variables including emotion were integrated into the student model, enabling the tutor to refine its inference of student frustration, engagement and confidence. Links between students' behaviors, attitudes and perceptions exist and correlations between help requests and learning have been shown to be consistent with other authors' findings (Wood & Wood, 1999; Renkl, 2002).

In a second study, machine learning was used to show that disengagement negatively correlates with performance gain (Johns & Woolf, 2006). *Hidden Markov models* research was augmented in our Item Response Theory dynamic mixture model (IRT-HMR) to simultaneously estimated a student's changing motivation level and proficiency (Johns & Woolf, 2006; Johns et al., 2006). This tutor predicted the probability of a correct student response, up to 75% accuracy. It was tested dynamically with high school students using the Wayang tutor. By accounting for a student's motivation, the dynamic mixture model accurately estimated proficiency and the probability of a correct response. Motivation was modeled as a dynamic, discrete variable and proficiency as a static, continuous variable. These assumptions are based on a student's tendency to exhibit different behavioral patterns over the course of a tutoring session.

The model was a combination of a hidden Markov and an IRT model. Based on Item Response Theory, a well tested and validated model, we generated a probabilistic model of student proficiency consisting of four variables: student proficiency, motivation, evidence of motivation, and a student's response to a problem (Johns & Woolf, 2006). We suggest that motivation is a dynamic variable as it changes durinf a session as a student becomes more or less engaged with the material. The latent variables in the student model correspond to proficiency and motivation. Proficiency is a static variable that does not change over time and we investigated three types of motivation: motivated, unmotivated -- abusing hints and unmotivated -- quickly guessing.

In this software study of motivation we measured student reaction to interventions, Figure 2. Evidence cleaned from 115 problem-solving sequences shows that students do change their behavior based on our tutor feedback. The tutor intervened when unmotivated behavior was recognized after the 6th problem (top graph) and then on-target engaged behavior returned (top line) and hint abuse (quickly asking for hints) subsided.

Dependent variables ⇒⇒⇒	**Frustration**	**Motivation /Flow**	**Confidence**	**Boredom**	**Fatigue**
Instruments to measure dependent variables (to be predicted)	Frustration Button (Burleson, 2006); AMAS, reduced math anxiety scale)	Harter's Scale)	Fenemma-Sherman Scale; Eccles scale	Boredom Proneness Scale; Are you bored?	Mental fatigue Scale Are you tired?
Behavioral Variables	Sensing data (camera, pressure sensitive chair, skin conductance glove, sensitive mouse)				
Behaviors that help predict the dependent variables	High state of arousal, high gaming; high effort; Gaussian Classification;	Record student effort exerted; dependence on help	Persistence at problem solving after incorrect attempts; dependence on encouragement messages	Low state of arousal combined with low effort and gaming	Increased problem solving time and increased error rate after some time in the tutoring session

Table 1. Independent behavioral variables and dependent variables used in emotion research

3. Responding to student emotion

Our goal is to structure a tutor's responses based on individual affective variables. We are investigating five independent affective variables, Table 1: *frustration* (feelings, thoughts, and behaviors associated with not achieving a particular goal), *motivation* (initiation, direction, intensity and persistence in an activity), *self-confidence* (belief in one's powers, abilities, or capacities), *boredom* (restlessness, or irritability that results from a lack of stimulation) and *fatigue* (mental weariness or decreased capacity to function normally because of excessive stimulation). We have two sets of dependent variables, those that track student engagement (or flow) and those that track negative emotions detrimental to learning (being stuck).

We have used a variety of heuristic policies to respond to student's emotions. For example, the heuristic policies listed in Table 2 have been applied when the specific emotional state is detected. We have also used machine learning optimization algorithms to search for policies for individual students in different affective and cognitive states, with the goal of achieving high learning and positive attitudes towards the subject, compared to pre-defined heuristic policies. We used a reinforcement learning agent to help discover optimal ways to react in specific affective states to maximize the long term goal of achieving high learning, by allowing posterior analysis of the policies that RL comes up with for students in different emotional states.

We use off-line unsupervised learning, meaning that we will learn optimal policies from student data from the previous year.

We are analyzing the dependency of specific behavioral variables, diagnosing emotions and synchronizing this information with the learning task to train classifier algorithms. We use a small subset of these variables to build a prototype model where we draw on the relations between emotional state engagement and actions. We identify and work with the most powerful predictors (e.g., motivation and frustration). In addition to the non-invasive techniques (wireless sensors and machine learning) and a survey, we may interrupt students in-between problems to ask about their feelings and attitudes for the data-collection phase. The steady state goal however is to measure emotions on line without any invasive techniques. We measured how feedback variables interact to promote learning in context (characteristics of the learner, aspects of the task). Instructional feedback will be varied according to type (explanation, hints, worked examples) and timing (immediately following an answer, after some elapsed time) (Shute, 2006).

Feedback for self-confidence. We measured the impact of feedback on student self-confidence. One version of our tutor selected new problems based on student proficiency prediction using ML and an HMM, and a second version provided friendly comments (graphs, tips, offering help) (Arroyo et al., 2004). Responses to questions such as "How will you do in mathematics next year?" show significant differences in the two intervention groups. Students in the ML version had higher confidence and motivation, higher self-concept and liked mathematics more. Students in the motivational version showed improvement in attitude and had better feelings towards the system ("The tutor is friendly/smart"). Both groups learned more and perceived that they learned more ("How much did you learn?").

Feedback for frustrated and bored students. We are using our hardware/software research platform to distinguish between bored and frustrated students. Stress sensors (mouse and chair) help identify frustration (arousal and hyperactive behavior) and boredom (containing gaming but with low arousal). In conjunction with activity behavior (hint asking, 'gaming the system' detection) pattern matching methods help infer these states in real time. For cases of frustration, we provide motivational and empathetic feedback to support students to understand failure and use it to move the student forward. For cases of boredom, we provide alternative activities (animation and exploratory modules) or more challenging projects.

*Feedback for self-concept***.** We are investigating student self-concept (assessment of current performance in a discipline), which is related to academic outcomes and motivation (Narciss, 2004). Students differ in their task specific self-concept and tend to explain their success or failure based on internal (their original talents) or external (originating in our environment) factors. Sadly, people with low self-concept attribute their failures to themselves and the reverse happens for people with high self-concept. We will use external responses ("That problem was really hard") when students of low self-concept fail, and use internal responses ("Congratulations, you did an amazing job with that!") when they succeed, hopefully reversing their negative beliefs.

Frustrated student	**Low motivation**	**Low confidence**	**Bored student**	**Fatigued student**
• Empathetic response: "That was frustrating. Let's move to something easier" • Give students control: "Would you li to choose the next problem? What kind would you like?"	• Agent changes voice, motion and gestures; • Presents graph, hints, adventures	• Provides encouragement; • Indicate student performance level • Link performance to student effort • Attribute failure to external (hard problem) and success to internal reasons (you are doing great)	• Increase challenge level of activities • Empathy messages: "Maybe you agree with me that this is quite boring? Let's move to something more challenging"	• Empathy message: "I am pretty tired of this. Let's switch to something more fun" • Change in scenario, e.g., adventures, animation, game
Not-frustrated No response	**High motivation** No response; Encouragement, praise.	**High-confidence** No response	**Not-bored** No response	**Not-fatigued** No response

Table 2. Responses of the tutor to student affect

We are investigating the dependency of specific behavioral variables and attitudes/emotions, the latter diagnosed with items from validated instruments at the end and in between the tutoring session. We specify in detail which behavioral variables and which interactions between variables help predict attitudes. We induce both *static* (based on demographics and emotion instruments) and *dynamic* (that include real-time sensor data as well as inferred hidden variables) student models (McQuiggan & Lester, 2006). Students are tested on validated instruments to measure emotion, and will be outfitted with sensors that wirelessly and inexpensively measure physiological and movement data while they work with different intelligent tutors. After completing the tutor (a matter of several days) students are again presented with emotion instruments to measure long-term changes in their motivation, self-confidence and self-efficacy. Both static and dynamic models have been shown to independently predict students' real-time levels of motivation, self-confidence and self-efficacy at between 70 – 80% accuracy. This research enhances the power of these individual models and the addition of physiological and movement data should further enrich their predictive power.

4. Instructional feedback research

Instructional feedback can be a powerful motivator when delivered in response to goal-driven efforts (Shute, 2006). Feedback can be used to support learning, and we are examining different types (and possibly timing) of feedback in relation to effects on affective states and ultimately learning. Towards that end, we are exploring innovative, non-invasive ways to measure (dynamically & statically) affective states and to evaluate different kinds of feedback to get learners back on track, in the zone, flowing. The contention is that being in the zone will increase the probability that the student will actually learn.

It is not immediately clear how best to engage students. Methods include adding fantasy contexts and pedagogical agents (Lester et al., 1997; Johnson et al., 2000; McQuiggan & Lester, 2006) as well as optimizing learning efficiency by peeling off unnecessary elements in the system's interfaces, such as fantasy and context, to improve learning efficiency. Understanding this and similar contradictions are in part the goals of this research.

Appropriate feedback does improve learning in human-to-human instruction. Feedback can reduce uncertainty about how well (or poorly) the student is performing and motivate strategies aimed at reducing that uncertainty (Ashford, et al., 2003). Sweller et al. (1998) showed how the presentation of worked examples reduces the cognitive load for low-ability, novice or struggling students. Feedback also provides useful information for correcting inappropriate task strategies, procedural errors, or misconceptions. One general recommendation is that immediate feedback for students with low achievement levels in the context of either simple (lower-level) or complex (higher-level) tasks is superior to delayed feedback; while delayed feedback is suggested for students with high achievement levels, especially for complex tasks.

Few intelligent tutor projects mix task-based and affect-based support in their learning environments. We will research the role of each support type individually and together (with several forms of 'together'). With so many interesting issues to investigate we are likely to uncover more questions than answers. For example we found gender differences in middle school students for task vs. affect based responses (Burleson, 2006). Boys and girls have differing responses to the social support of the Companion. Girls were more frustrated if they did not receive both task and affect based support and when they did receive both their frustration was reduced. Studying the controlled mixing of affect and task support is difficult to do in human-human interactions and is highly leveraged by use of an intelligent tutor.

5. Feedback in computer tutors

A number of computer projects have tackled sensing and modeling emotion in learning and educational gaming environments (Kort et al., 2001; Kapoor and Picard, 2001). A dynamic decision network was used to measure a student's emotional state based on variables such as heart rate, skin conductance and eyebrow position (Zhou & Conati, 2003). A probabilistic model applied Decision Theory to choose the optimal tutor action to balance motivation and student learning. This model linked hidden nodes that indexed personality, goals, and emotional states to observable variables captured by sensors and heart monitors. Eye tracker data was used, which is almost always noisy as students gaze at non-relevant information. The structure and parameters of the model, in the form of prior and conditional probabilities, were set by hand and not estimated from data. Lack of engagement has been shown empirically to correlate with a decrease in learning (Baker, et al., 2004). In this study, automated diagnosis in a latent response model classified student actions as either gaming the system or not. However the tutor response elicited negative feelings from the students, in part because it blocked students who were presumed to be gaming,(Aleven et al., 2005).

Many of the systems above did not provide appropriate instructional feedback to move students to a state of increased learning. Some did not use fully adaptive learning environments and others were games or web pages of text-books. On the other hand, the research described here integrates emotion with intelligent tutors as part of classroom learning.

6. Intelligent tutors and affective agents

The hardware and software technologies described above are being implemented in several intelligent tutors. The potential of tutors that recognize affect has barely been tapped and the development of emotionally intelligent tutors is still in its infancy (Picard, 2000). The emphasis till now has been on improving reasoning about factual, procedural or cognitive knowledge, not about affect. Research involving affect has been primarily in laboratory studies involving single students and invasive sensors.

The technologies we are developing recognize human emotion, possess emotional intelligence (e.g. mirror emotion) and are domain-independent. Domain independence is demonstrated by evaluating the same intelligent affective agent across multiple domains and content areas. We examine agent performance with in two intelligent tutors to demonstrate how such affective agents will work. We are using two tutors, a closed world geometry tutor that expects explicit answers to geometry problems and an open-ended biology tutor that supports student exploration in human biology.

Geometry Tutor. Affective technologies have been integrated into Wayang Outpost, which prepares students for the mathematics section of the standardized state exams[2] (Arroyo et al, 2004). The student model represents geometry knowledge and recognizes which skills a student has learned. It also uses machine learning to model student affective characteristics, e.g., interest in a topic, amount of challenge in learning and whether the student wants to continue with tutor.

This tutor improved learning gains (an average 12% improvement from pretest to posttest) after only 2 class periods and students passed the state standard exam at a higher rate (92%) as compared with students not using the tutor (76%). The tutor currently uses information about each student's cognitive skills and dynamic multimedia techniques (sound and animation) to customize instruction and improve performance. It incorporates knowledge of student group characteristics (e.g., profile of cognitive skills, gender) to guide instruction. The tutor customizes the choice of hint type for individual students based on their cognitive profile, gender, spatial ability, and math fact retrieval speed.

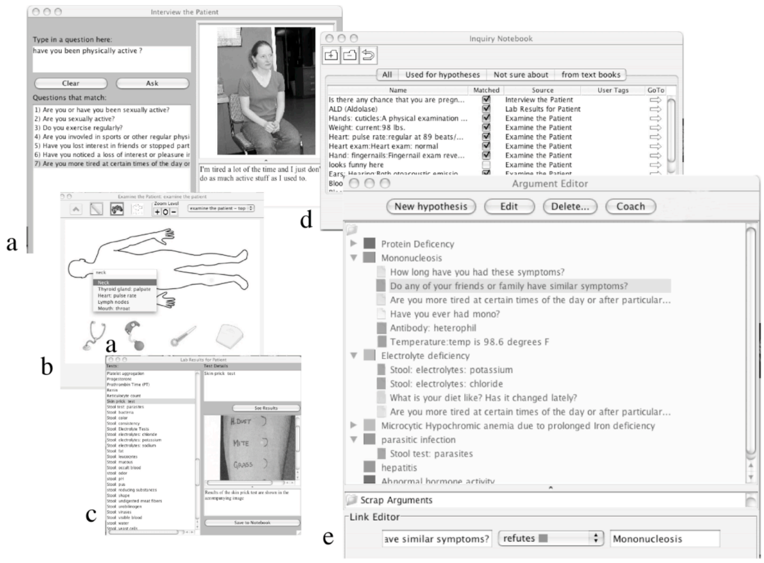

Figure 3. Human Biology Tutor.

Students interview the patient (a); perform an examination (b); or order lab tests and imaging studies (c). Data collected and the site of the observations from the *Interview Room* (a) or *Examination Laboratory* (b), are recorded automatically in the *Inquiry Notebook* (d). Students posit several diagnoses in the *Argument Editor* (e) using a drag and drop feature to move data from the *Inquiry Notebook* (d) to the *Argument Editor* (e) to link evidence that supports or refutes each diagnosis.

Human biology inquiry tutor. Affective technologies will be integrated into the Rashi Inquiry Tutors, which invite students to posit theories and recognize when data does or does not support their hypotheses, Figure 3. The environment tracks student investigations (e.g., questions, hypotheses, data collection and inferences) and helps students articulate how evidence and theories are related. These tutors have been used with hundreds of students in biology, geology and ecology (Woolf et al., 2003; 2002; Murray et al., 2004; Bruno, et al., 2000; Bruno & Jarvis, 2001). The tutor provides an expansive set of tools to help students access and organize information, collect data, organize evidence and construct arguments, within a central data repository.

For example, inquiry cases in human biology support students to gain a solid understanding of analytical approaches to solving medical problems and of how diseases are transmitted, their physiological effect, and the immune response to disease-causing microorganisms.[3] Students search the Web to learn more

[2] http://wayang.cs.umass.edu/Wayang/flash/

[3] Human Biology Inquiry Tutor: http://ccbit.cs.umass.edu/Rashihome/projects/biology/index.html

about the patient's symptoms and about medical tests. Patients' complaints form an initial set of data from which students begin the diagnostic process, by "interviewing" the patient about symptoms, examining her or ordering laboratory tests, Figure 3 b, c. Some data is made visible by student action, e.g., asking for chest x-rays, asking the patient a question, or using a blood-pressure tool. Some data is interpreted for the student (e.g., "x-ray normal"); the student interprets other data. Though similar to BioWorld (Lajoie, 1998) in that students diagnosis a disease and utilize a wide variety of information about the patient, Rashi is distinct in that it is a domain-independent (implemented in four domains) and provides cognitive tools to support students to generate hypotheses and formulate arguments.

7. Discussion

Both the tutors above will be enhanced with tools that sense emotion and respond to affect with the goals of supporting student reflection and increasing teaching efficiency. Earlier research integrated ML with student models (Conati, 2002; Murray & VanLehn, 2000; Baffes & Mooney, 1996; Mayo & Mitrovic, 2001). However, most focused on short-term decisions. A central focus of this work is to support long-term pedagogical decisions and to view a series of student actions, not simply to provide single-shot responses. Emotion recognition can significantly improve our long-term planning if teaching is sometimes directed toward eliciting long-term experiences, which might entail sacrifices in immediate performance. For example, maintaining a level of student engagement in the tutoring process should be a priority. Enhancing students' attention and willingness to continue may imply sacrificing students' learning at times. If the long-term goal is to have students learn and 'stay' with the tutor, it may be important to sacrifice immediate learning by interleaving multimedia 'adventures,' for example, when observing signs of boredom or confusion to recover students' engagement with the system.

In the long term, we will evaluate several predictions, specifically that affective companions perceived as 'caring' will: help students persevere longer through frustrating learning episodes, out motivate agents perceived as neutral, and contagiously excite learners with passion for a topic, leading to greater effort to master the topic (Picard, 2006).

8. Acknowledgement

This research was funded in part by awards from NSF: 'Learning to Teach: The Next Generation of Intelligent Tutor Systems' Woolf, B. P., and I. Arroyo with A. Barto, S. Mahadevan, and D. Fisher, Research on Learning and Education and Engaging Undergraduates in On-Line Inquiry Learning: A Case-Based CyberLibrary in Human Biology, Merle Bruno PI, Beverly Woolf Co-PI. .Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

9. References

Aleven, Vincent; Roll, Ido; McLaren, Bruce; Ryu, Eun Jeong; Koedinger, Ken. (2005). An architecture to combine meta-cognitive and cognitive tutoring: Pilot testing the Help Tutor. *Proceedings of the 12th International Conference on Artificial Intelligence in Education*. 2005

Arroyo, I., and Woolf, B. (2005). Inferring Learning and Attitudes from a Bayesian Network of Log File Data. *In Proceedings of the Twelfth International Conference on Artificial Intelligence in Education*.

Arroyo, I.; Beal, C.; Murray, T.; Walles, R.; and Woolf, B. (2004). Web-based Intelligent Multimedia Tutoring for High Stakes Achievement Tests. *In Proceedings of the Seventh International Conference on Intelligent Tutoring Systems* 468–477.

Ashford, S. J., Blatt, R., & VandeWalle, D. (2003). Reflections on the looking glass: A review of research on feedback-seeking behavior in organizations. *Journal of Management, 29*, 773-799.

Baffes, P., & Mooney, R. (1996). Refinement-based student modeling and automated bug library construction. *Journal of Artificial Intelligence in Education, 7*(1), 75-116.

Baker, R., Corbett, A.; and Koedinger, K. 2004. Detecting Student Misuse of Intelligent Tutoring Systems. *In Pro- ceedings of the Seventh International Conference on Intelligent Tutoring Systems* 531–540.

Block, J. (1995). On the relation between IQ, impulsivity and delinquency. *Journal of American Psychology*, 104, 395-398.

Bruno, M. (2000). Student-active learning in a large classroom, *Project Kaleidoscope 2000 Summer Institute*. Keystone, Colorado.

Bruno, M., & Jarvis, C. (2001). It's fun, but is it science? Goals and strategies in a problem-based learning course. *The Journal of Mathematics and Science: Collaborative Explorations, 4*(1), 9-24.

Burleson, W., (2006), Affective Learning Companions: strategies for empathetic agents with real-time multimodal affective sensing to foster meta-cognitive and meta-affective approaches to learning, motivation, and perseverance, MIT PhD Thesis, available at http://affect.media.mit.edu/pdfs/06.burleson-phd.pdf

Conati, C., Gertner, A. and VanLehn, K. (2002). Using bayesian networks to manage uncertainty in student modeling. *User Modeling and User-Adapted Interaction, 12*(4).

Cytowic, R.E. *Synesthesia: A Union of the Senses*. Springer-Verlag, New York, 1989.

Dennerlein, J., T. Becker, et al. (2003). Frustrating Computers Users Increases Exposure to Physical Factors. Proceedings of the International Ergonomics Association.

Goleman, D. (1996). Emotional Intelligence: why it can matter more than IQ. London : Bloomsbury.

Haro, A., I. Essa, et al. (2000). Detecting and Tracking Eyes by Using their Physiological Properties. *Computer Vision and Pattern Recognition*.

Johns, J. and Woolf, B. (2006). A Dynamic Mixture Model to Detect Student Motivation and Proficiency, In *Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI-06)* 2-8. Menlo Park, Calif.: AAAI Press.

Johns, J., Mahadevan, S. and Woolf, B. (2006). Estimating Student Proficiency using an Item Response Theory Model, in M. Ikeda. K Ashley and T.-W. Cahn (Eds.): *ITS 2006*, Lecture Notes in Computer Science, 4053, pp 453-462, Springer-Verlag Berlin Heidelberg.

Johnson, L., J. Rickel, et al. (2000). Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments. *International Journal of Artificial Intelligence in Education* 11: 47-78.

Kapoor, A. and Picard, R. W. (2001). A real-time head nod and shake detector, Workshop on Perceptive User Interfaces, Orlando, FL.

Kapoor, A. and Picard, R. W. (2002). Real-Time, Fully Automatic Upper Facial Feature Tracking. *Proceedings of the 5th International Conference on Automatic Face and Gesture Recognition*, Washington D.C., May 20-21, 2002.

Kapoor, A., Burleson, Winslow and W Picard, Rosalind (to appear) Automatic Prediction of Frustration," International Journal of Human-Computer Studies, to appear. IJHCS Online by subscription

Kort B, Reilly R and Picard R W. (2001) An affective model of interplay between emotions and learning: reengineering educational pedagogy-building a learning companion, in *Proceedings of International Conference on Advanced Learning Technologies* (ICALT 2001), Madison Wisconsin.

Lajoie, S. P. (1998). Promoting argumentation in face to face and in distributed computer-based learning situations: Constructing knowledge in the context of BioWorld . In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the 20th Annual Conference of the Cognitive Science Society* (p. 5), Mahwah, NJ: Erlbaum.

Lepper, M. R., & Hodell, M. (1989). Intrinsic motivation in the classroom. In C. Ames & R. E. Ames (Eds.), Research on motivation in education (Vol. 3, pp. 73-105). New York: Academic Press.

Lester, J. C., S. A. Converse, et al. (1997). The persona effect: Affective impact of animated pedagogical agents. *CHI'97*, New York, NY, ACM Press.

Mayo, M., and Mitrovice, A. (2001). Optimising ITS Behaviour with Bayesian Networks and Decision Theory. *International Journal of Artificial Intelligence in Education* 12:124–153. van der Linden, W., and Hambleton, R., eds. 1997. *Hand- book of Modern Item Response Theory*. New York, NY: Springer-Verlag.

McQuiggan, S. and Lester, J., (2006) Diagnosing Self-Efficacy in Intelligent Tutoring Systems: An Empirical Study, In *Proceedings of the 8th International Conference on Intelligent Tutoring Systems* (ITS 2006), Jhongli, Taiwan, June 26-30, 2006.

Murray, R. C., & VanLehn, K. (2000). *DT tutor: A decision-theoretic, dynamic approach for optimal selection of tutorial actions.* Paper presented at the Intelligent Tutoring Systems: 5th International Conference, Montreal, Canada.

Murray, T., Woolf, B., & Marshall, D. (2004). Lessons learned from authoring for inquiry learning: A tale of three authoring tools, *7th International Conference on Intelligent Tutoring Systems*. Maceio, Brazil: Springer Lecture Notes in Computer Science 3220.

Narciss, S. (2004). Task specific self-concept, learner control and informative tutoring feedback - How do they affect motivation and achievement in concept learning? In H.W. Marsh, J. Baumert, G. E. Richards, & U. Trautwein (Eds.), *Proceedings of the 3rd International Biennial SELF Research Conference: Self-concept, Motivation and Identity: Where to from Here?*.

Norman D A. (1981) Twelve issues for cognitive science, in Norman D A (Ed): *Perspectives on cognitive science,* pp 265—295, Erlbaum, Hillsdale, NJ

Ortony, A., Clores, G.L, and Collins, A. (1988). *The Cognitive Structure of Emotions*. Cambridge University

Press, Cambridge, MA.

Picard, R., (2006) Building an affective learning companion, Invited keynote talk, International Conference of Intelligent tutoring Systems, Taiwan,http://www.its2006.org/ITS_keynote/ITS2006_01.pdf - search=%22Building%20an%20affective%20learning%20companion%20ppt%22

Picard, R.W., S. Papert, W. Bender, B. Blumberg, C. Breazeal, D. Cavallo, T. Machover, M. Resnick, D. Roy and C. Strohecker (2004), Affective Learning--A Manifesto, *BT Technical Journal*, Vol. 22, No. 4, October 2004.

Reynolds, C., 2001. The sensing and measurement of frustration with computers. Master's thesis, MIT.

Shute, V. J. (2006). *Focus on formative feedback.* ETS Research Report, Princeton, NJ.

Strauss, M., Reynolds, C., Hughes, S., Park, K., McDarby, G., Picard, R. W.,2005. The handwave bluetooth skin conductance sensor. In: *ACII*. pp. 699–706.

Sweller, J., Van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10,* 251-296.

Woolf, B., Marshall, D., Mattingly, M., Lewis, J., Wright, S., Jellison, M., et al. (2003). Tracking student propositions in an inquiry system, *Artificial Intelligence in Education 2003*. Sydney, Australia.

Woolf, B., Reid, J., Stillings, N., Bruno, M., Murray, D., Rees, P., et al. (2002). A general platform for inquiry learning. In S. A. Cerri, G. Gouarderes & F. Paraguaca (Eds.), *Intelligent tutoring systems, 6th international conference, its 2002* (Vol. 2363, pp. 681-697): Springer Lecture Notes in Computer Science.

Zhou X and Conati C. (2003). 'Inferring user goals from personality and behavior in a causal model of user affect', *Intelligent User Interfaces 2003*, pp 211—218.