An affect-sensitive intelligent tutoring system with an animated pedagogical agent that adapts to student emotion like a human tutor

A thesis presented in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy
in Computer Science

at Massey University, Albany,
New Zealand.

Samuel Thomas Vaughan Alexander

2007

Abstract

One of the established strengths of human tutors is their ability to recognise and adapt to the emotions of students. This is a skill that has traditionally been lacking from Intelligent Tutoring Systems (ITSs); despite their ability to intelligently model and adapt to aspects of the student's cognitive state, ITSs are generally completely unable to detect or adapt to aspects of the student's affective state.

In response to this shortcoming, this thesis explores the pioneering development of an emotion-sensitive ITS. With the empathy of effective human tutors as our blueprint, we investigate how an artificial tutor should adapt to the affective state of students, and develop an original affective tutoring strategies method. As a validation of the feasibility of an emotion-sensitive tutoring system, we implement and test our method in a functional Affective Tutoring System (ATS) for counting and addition, Easy with Eve, featuring an empathetic animated pedagogical agent, Eve. Eve is able to detect student affect using an in-house real time facial expression analysis system.

To inform the system's adaptation to student affect, the novel method for student modelling and emotion-sensitive tutoring strategies has been developed using a fuzzy, case-based reasoning approach. This approach is used to mine data about human tutor adaptations to student affect that was generated by an observational study of human tutors that was carried out in a local primary school.

To test the impact of emotion detection and the presence of the animated agent, four different versions of the ATS were tested in local primary schools with a total of 59 participants. The findings from the study indicate that adding the detection of facial expressions to the student model did not improve student short-term performance, but there was mixed evidence that the presence of the animated agent Eve may cause students to perceive the system slightly more positively (a *persona effect*). This effect was marginally greater when the animated agent was enabled to detect and adapt to the affective state of students, which tentatively shows that emotion detection in an ATS may have a positive effect on student motivation.

Acknowledgements

Primarily, I am indebted to Abdolhossein Sarrafzadeh, my main supervisor, who has been an unceasing source of advice, feedback, guidance and kind support over the time that I have worked with him; without his assistance, preparing this thesis would have been many, many times more difficult. Secondly, I am also very grateful to Stephen Hill, my co-supervisor, who has also been a rich source of insight and good judgement, and who has provided a much valued psychological perspective on my research.

The development of Easy with Eve was greatly assisted by Chao Fan, who implemented the real-time facial expression recognition system that was used to detect student emotion. I am also thankful for the help that I received from Charlotte Bostin, who recorded the voice tracks for Eve and who also helped with the inter-rater reliability study in Chapter 3.

Thank you to Gail and Micah Williams, and Tabitha and Luke Beaumont, who helped with the pilot tests of Easy with Eve. Thank you also to the teachers, students and parents who assisted with the observational study of human tutors and the study of the effectiveness of Easy with Eve. I am especially grateful to Karen Spicer of Murrays Bay School and Sheila Brown of Torbay School, who helped with the organisation of the study of the effectiveness of Easy with Eve.

Thank you also to Antonija Mitrovic and Stellan Ohlsson, for their timely advice near the beginning of this research. I am also grateful for the financial assistance that I received from Massey University, the Institute of Information and Mathematical Sciences, the New Zealand Vice-Chancellors’ Committee and Education New Zealand.

On a personal note, the other PhD students in the Computer Science and Information Systems departments have helped to make my time in Albany bright and enjoyable, and I will miss our little chats. Finally, thank you to all of my family, and especially my parents, for their consistent support and encouragement that has helped to make this all possible.

List of publications

Refereed journal articles

- **Alexander, S. T. V.**, Sarrafzadeh, A., & Hill, S. (2007). Foundation of an affective tutoring system: Learning how human tutors adapt to student emotion. *International Journal of Intelligent Systems Technologies and Applications* (in press).

- Sarrafzadeh, A., **Alexander, S. T. V.**, Dadgostar, F., Fan, C., & Bigdeli, A. (2007). "How do you know that I don't understand?" A look at the future of intelligent tutoring systems. *International Journal of Computers in Human Behavior* (in press).

Refereed international conferences

- Sarrafzadeh, A., **Alexander, S. T. V.**, Dadgostar, F., & Fan, C. (2006). See me, teach me: Facial expression and gesture recognition for intelligent tutoring systems. *IEEE International Conference on Innovations in Information Technology*. Dubai, United Arab Emirates.

- **Alexander, S. T. V.**, Sarrafzadeh, A., & Hill, S. (2006). Easy with Eve: A functional affective tutoring system. *Workshop on Motivational and Affective Issues in Intelligent Tutoring Systems*, Intelligent Tutoring Systems 2006, Jhongli, Taiwan.

- **Alexander, S. T. V.** (2005). Modeling emotions from non-verbal behaviour in an affective tutoring system. *User Modeling*. Edinburgh, Scotland.

- **Alexander, S. T. V.**, Hill, S., & Sarrafzadeh, A. (2005). How do human tutors adapt to affective state? *Workshop on Adapting the Interaction Style to Affective Factors*, User Modeling 2005. Edinburgh, Scotland.

- Sarrafzadeh, A., Fan, C., Dadgostar, F., **Alexander, S. T. V.**, & Messom, C. (2004). Frown gives game away: Affect sensitive tutoring systems for elementary mathematics. *IEEE International Conference on Systems, Man and Cybernetics*. The Hague, The Netherlands.

- **Alexander, S. T. V.**, & Sarrafzadeh, A. (2004). Interfaces that adapt like humans. *Asia-Pacific Computer-Human Interaction (APCHI 2004)*. Rotorua, New Zealand.

- **Alexander, S. T. V.**, Sarrafzadeh, A., & Fan, C. (2003). Pay attention! The computer is watching: Affective tutoring systems. *E-Learn Conference*. Phoenix, Arizona.

Other publications

- **Alexander, S. T. V.** (2006). Implementing Easy with Eve. *Institute of Information and Mathematical Sciences Postgraduate Conference*. Massey University, Albany, New Zealand.

- **Alexander, S. T. V.** (2005). Lifelike tutors that adapt to facial expressions. *Institute of Information and Mathematical Sciences Postgraduate Conference*. Massey University, Albany, New Zealand.

- **Alexander, S. T. V.** (2004). Emulating human tutor empathy. *Institute of Information and Mathematical Sciences Postgraduate Conference*. Massey University, Albany, New Zealand.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

The strange thing about life is that though the nature of it must have been apparent to everyone for hundreds of years, no one has left any adequate account of it. The streets of London have their map; but our passions are uncharted.

Virginia Woolf, Jacob's Room

This thesis is about "charting the passions" in education.

In this introduction we firstly give a brief background to the history of emotions in Intelligent Tutoring Systems, and then we clarify what exactly is meant when we talk about the "affective state" of students. Then we list the goals and contributions of this research, and finish with a summary of the structure of this thesis.

1.1 Statement of the problem

Computers in education is an immense issue in today's schools, universities and businesses, with an annual global revenue now measured in tens of billions of dollars (eMarketer.com, 2007). This global increase in e-learning expenditure has also been reflected in the quickly mounting body of academic research in this area; recent years have witnessed dozens of international conferences and journals related to computers in education, and this trend seems likely only to gather yet more momentum. Therefore, with such a vast increase in the amount of e-learning research, the question is increasingly asked: can computers teach as well as people?

In fact, e-learning has been a significant issue dating right back to the late 1950s, when American psychologist B. F. Skinner pioneered “programmed instruction”, where students used “teaching machines” to do basic flashcard exercises (Hill, 1997). As technology improved, the teaching machines were superseded by computers, and so programmed instruction developed into Computer Assisted Instruction (CAI). But what both programmed instruction and CAI systems both lacked was a means of personalising their instruction (Urban-Lurian, 2003); this meant that all students were treated at all times in exactly the same way. The strengths, the weaknesses and the learning styles of individual students simply never came into consideration.

As a reaction against this weakness, the successor of CAI was the Intelligent Tutoring System (ITS), which came into existence during the early 1970s and is still in vogue today, albeit mostly in academia. ITSs are characterised by their “intelligent” ability to adapt their tutoring strategies to the knowledge and abilities of individual learners (Urban-Lurian, 2003). ITSs have certainly proved to be effective, and some studies have shown that in some domains ITSs even outperform traditional classroom instruction (e.g. Anderson, Corbett, Koedinger, & Pelletier, 1995). But even so, competent one-to-one human tutoring remains the most effective means of instruction – and were there enough one-to-one tutors in the world for everyone, and enough money for everyone to pay them, we can safely assume that both CAI and ITSs would have received significantly less attention than they have done.

This then raises the question: why is it that ITSs are not as effective as one-to-one human tutoring? Doubtless there are many answers to this question that we could investigate, but in this thesis we focus on just one: a severe lack of *empathy* in ITSs. Many researchers now feel strongly that traditional ITSs completely overlook one of the human tutor’s greatest allies, because they do not understand the student’s affective state (e.g. Alexander, Sarrafzadeh, & Hill, 2005; Picard, 1997; Kort, Reilly, & Picard, 2001). Such is the nature of human communication, that tutors unconsciously process a continuous stream of rich affective information that can guide their tutoring; for this reason, competent human tutors can adapt their tutoring according to both the cognitive *and* affective states of their students. Indeed, many researchers are now increasingly working towards augmenting existing ITSs with emotional awareness, sensitivity and expressivity, believing that this will significantly enhance their effectiveness (e.g. Burleson, 2006; Conati & Mclaren, 2004; D'Mello, Craig, Gholson, Franklin, Picard, & Graesser, 2005). However, although various studies highlight the role of affect in learning, a completely functional emotion-sensitive tutoring system has not yet been reported in the literature.

We began this section by posing the question whether computer tutors could be just as effective as human tutors. The answer to that question lies far beyond the scope of this thesis, but what we *do* explore is the pioneering development of an emotion-sensitive ITS – an Affective Tutoring System (ATS). With the empathy of effective human tutors as our blueprint, we investigate how an ATS should adapt to the affective state of students, and develop an affective tutoring strategies method. To gauge the feasibility of an emotion-sensitive tutoring system, we implement and test our method in a functional ATS for counting and addition, Easy with Eve, featuring an empathetic animated pedagogical agent, Eve.

However, before we focus on the particular goals and contributions of this research, it is important that we clarify exactly what we mean by the *affective state* of students.

1.2 A definition of affect

Despite the fact that people from all walks of life talk about emotions almost every day without misunderstanding, defining exactly how emotions function is one of psychology's notoriously challenging questions, that has led to many different theories (Sloman, 1991; Strongman, 2003). In fact, by the early 1980s there had been towards a hundred different definitions of emotions recorded (Kleinginna & Kleinginna, 1981), and doubtless there have been many more since then.

However, for the sake of this thesis we can broadly define affect using the explanation given by Parkinson (1995, p.19): "An emotion is a relatively short-term, evaluative state focused on a particular intentional object (a person, an event, or a state of affairs). Good examples are anger, fear, love, and hate." Other examples of short-term emotions that we might find in a learning context could be confusion, frustration, interest, boredom or happiness (Picard, 1997) – longer-term feelings such as moods are not considered in this thesis. As is perhaps already clear, we will use the terms *affect* and *emotion* interchangeably; the term *affective state* will be used to describe a person that is experiencing a particular emotion at a point in time.

The particular goals and contributions of this research will now be outlined in detail in the two following sections.

1.3 Research goals

As stated earlier in Section 1.1, the overall aim of this research is to explore the feasibility of an ATS that is capable of recognising both the cognitive and affective state of students; also studied is an approach to the development of emotion-sensitive tutoring strategies that makes use of this information. As well as recognising expressed emotions using an existing in-house image processing system, the tutoring system should also be able to show emotions through an animated pedagogical agent. The aim of the ATS is summarised in the model presented in Figure 1.1: the left hand side of the diagram at time t_i represents the cognitive and affective state of the student immediately following a student action, where the affective state is identified by detecting the student's facial expression. At time t_{i+1} the system's animated agent should then respond to the cognitive and affective state of the student by adapting both its tutoring and its own facial expression, with the intention of mapping the student's cognitive and affective states to a particular desired state.

In particular, this overall aim can be broken down into the following four main sub-goals:

1. To gather data about how human tutors adapt to the affective state of students in a one-on-one tutoring scenario. This involves an observational study where several human tutors were videoed as they interacted with primary school students. This goal is addressed in Chapter 3.

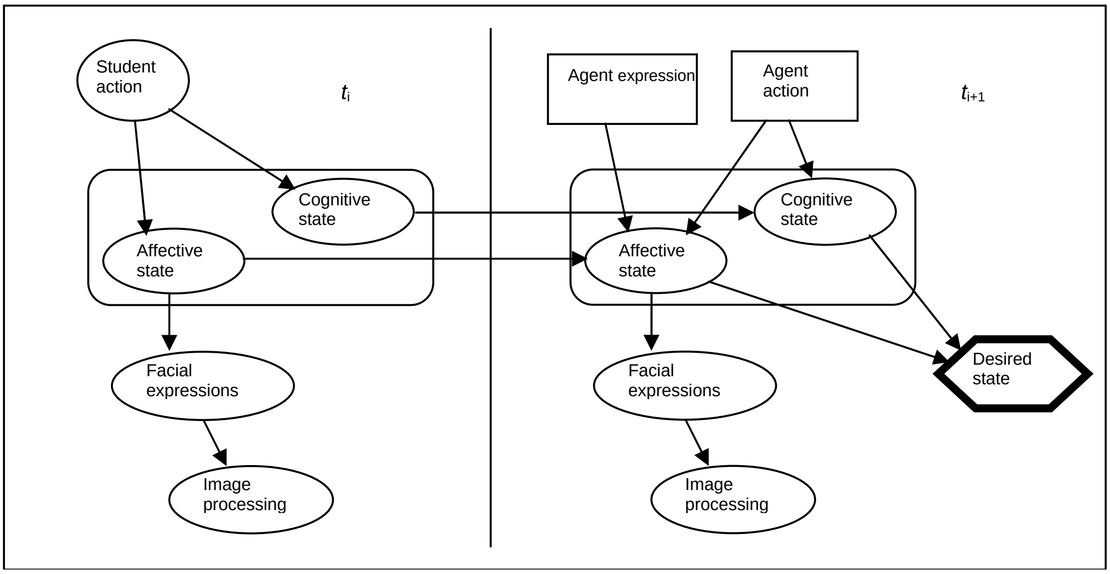


Figure 1.1. Model of the ATS, adapted from Conati (2002).

2. To develop a method for adapting the tutoring strategies of an ATS in accordance with both the affective and cognitive states of students. The affect-sensitive method for adapting to students is based on the data that was gathered from the observational study of human tutors in Chapter 3. This goal is addressed in Chapter 4.

3. To determine the feasibility of an emotion-sensitive tutoring system by applying the tutoring strategies approach from Chapter 4 in a functional ATS. The ATS is in the domain of primary school addition, and to feature an animated pedagogical agent. This goal is addressed in Chapter 5.

4. To test whether the ATS has a significant effect on the learning of primary school students. This is to gather data about the effectiveness of the emotion-sensitive tutoring strategies approach, and about the impact of agent-based vs. text-based feedback. This goal is addressed in Chapter 6 and discussed in Chapter 7.

1.4 Contributions of the thesis

Mapping to the four research sub-goals in the previous section, the following are the contributions to knowledge that this thesis makes:

1. An observational study of human tutors has been carried out that has gathered data about how human tutors adapt to students. This study considered both affective and non-affective student and tutor actions. This study and its findings are presented in Chapter 3.

2. A novel method for student modelling and emotion-sensitive tutoring in an ATS has been developed that uses a fuzzy, case-based reasoning approach. This approach has been applied in an emotionally intelligent tutoring strategies module that can recommend tutoring actions to an ATS based on the data from the observational study by comparing how human tutors acted in similar scenarios. The fuzzy, case-based approach and the implemented tutoring strategies module are presented in Chapter 4.

3. The feasibility of an emotion-sensitive tutoring system has been validated through the implementation of (to the knowledge of the author) the first ever functional ATS, Easy with Eve, that helps primary school students with counting and addition. The system features Eve, the first of a new type of emotion-sensitive animated pedagogical agent that can recognise, express and adapt to selected student emotions. The design and implementation of Easy with Eve is presented in Chapter 5.

4. The ATS was also the first to be tested in local schools; the aim of the testing was to examine the effectiveness of the tutoring strategies approach and the impact of agent-based vs. text-based feedback. The methodology and results of the study are presented in Chapter 6, and these results are discussed in Chapter 7.

1.5 Structure of the thesis

This thesis is structured as follows:

Chapter 2. Here we review the relevant literature to the implementation of an ATS. We begin by discussing the links between student learning and student affective state; this is followed by a review of the field of affective computing, and finally a review of work leading towards actual ATSs.

Chapter 3. To gather information about how an ATS should adapt to student emotion, Chapter 3 presents the methodology and results of an observational study of three human tutors who were videoed while interacting one-on-one with students at a local primary school.

Chapter 4. In this chapter we present the case-based reasoning tutoring strategies approach that was developed to mine the data from the observational study in Chapter 3 to generate emotionally-sensitive tutoring suggestions for the ATS.

Chapter 5. Here we present the implementation of the innovative, functional ATS, Easy with Eve.

Chapter 6. In this chapter we present the methodology and results of a study testing the effectiveness of the ATS, Easy with Eve, and in particular the effectiveness of the tutoring strategies approach that was discussed in Chapter 4. The study also tests the impact of agent-based vs. text-based feedback.

Chapter 7. In this chapter we discuss the results and lessons learned from Chapters 3, 4, 5 and 6.

Chapter 8. Finally, we bring our thesis to a close by drawing conclusions about the work presented in this thesis, and by suggesting opportunities for future research.

Chapter 2

Literature review

In this chapter we review the literature relevant to our overall research goal of an Affective Tutoring System (ATS). In particular, this thesis studies three areas: affect and learning, affective computing, and affective computing in the context of Intelligent Tutoring Systems (ITSs). Each of these areas will now be discussed in turn: in Section 2.1 we explore the link between student learning and emotions, that gives a strong rationale for the development of an affect-sensitive tutoring system; in Section 2.2 we look at the broader research area of affective computing; and finally in Section 2.3 we review the progress of other groups working towards the development of an ATS.

2.1 Affect and learning

In this section, we address the significance of affect to learning in a tutoring scenario. The importance of affect to tutoring can be broken down into two main issues: firstly that the affective state of the learner influences the cognitive activities required for learning, and secondly that affective feedback given by a learner is a critically important cue by which human tutors constantly regulate their teaching. We shall briefly consider each of these issues in turn.

2.1.1 How student emotions affect learning

The emotions of a student can be shown to have a significant influence on his/her ability to learn new information or to solve problems. Although researchers vary in exactly how they interpret particular student affective states, and although much of the research in this area is still reasonably tentative, we show in this section that a sound case can be made for a strong link between student emotion and student learning.

The link between affect and performance on cognitive tasks is becoming increasingly well documented in both the neuroscience and psychology literatures. Lisetti (1999) claims that emotions have a significant influence on a wide range of cognitive tasks, such as decision-making, planning, adapting to new environments and learning. Picard (1997) agrees, claiming that "emotions play an essential role in rational decision making, perception, learning, and a variety of other cognitive functions" (p. x).

Picard's conclusions are on the basis of work such as Damasio (Saver & Damasio, 1991), who cites the now famous case of "Elliot", who suffered damage to a portion of his frontal lobe (the ventromedial prefrontal cortex) as a result of a tumour. This damage had a profound effect on Elliot's personality – he had lost the ability to feel emotion. He was aware that he had *used* to feel emotion, but he was no longer able to feel it any more. This meant that even though he still scored above average on standard IQ and social intelligence tests, his decision making had become detached from any form of emotional input. This had tragic consequences in Elliot's case, as he lost all sense of "gut feeling"; he completely lacked any emotional aversion to disagreeable consequences. Without any "emotional wisdom" to draw upon from previous failures or successes, Elliot soon embarked on a programme of disastrous decision making that lost him his wife, his next wife, a series of jobs, and his entire life savings. Based on cases like Elliot, and accompanying neuropsychological research, Damasio and his colleagues were able to conclude that emotions are in fact an essential element of "rational" decision making. Indeed, when comparing Elliot to other patients with similar frontal lobe damage, Damasio consistently found a lack of emotion to be linked with reliably poor decision making: in Damasio's words, "the powers of reason and the experience of emotion decline together" (1994, p. 54). These findings led to Damasio's (1994) formulation of the "somatic marker hypothesis", which states that the process of decision-making is dependant on emotion; further evidence has since been amassed that also supports this hypothesis (Bechara, Damasio, & Damasio, 2003). Therefore, the cases of Elliot and the other similar patients clearly illustrate that performance on cognitive tasks – such as *learning* – cannot be considered distinct from affective state, as the latter unquestionably shades the former.

Another example of the way in which affect can impact on cognitive tasks is the consequences of *positive affect*, which has been shown to facilitate tasks such as creative problem solving and cognitive organisation, and increases intrinsic motivation for performing tasks (Isen, 2000; Isen & Reeve, 2005). In one set of studies on positive affect, participants were required to solve the so-called candle problem (Duncker, 1945) in a test of creative problem solving. The aim of the problem was to attach a candle to the wall so that it would burn without dripping wax onto the table or ground, using nothing but a box of tacks and matches; the solution was to place the candle inside the box, which is tacked to the wall – this is creative because the box has to be used in a novel way. Across several studies, participants in whom positive affect had been induced significantly outperformed control participants (Greene & Noice, 1988; Isen, Daubman, & Nowicki, 1987), which shows that positive affect can increase performance on creative problem solving.

Another study that illustrates the impact of positive affect involved asking a set of participants to rate, from one to ten, the degree to which particular examples fit into a given category – for example, *shirt* is a much better match for the category *clothing* than *cane* (Isen et al., 1987). Participants in whom positive affect had been induced rated fringe examples such as *cane* higher than participants in whom positive affect had not been induced. This suggests that positive affect assists cognitive organisation, because even reasonably diverse concepts were seen to be related, such as *cane* and *clothing*. One possible theory for explaining this is that the neurotransmitter dopamine, which is associated with positive affect, is also associated with the activation of regions of the brain related to the ability to switch cognitive perspective (Ashby, Isen, & Turken, 1999); in any case, positive affect participants were able to cognitively organise the concepts in such a way as to maximise their similarities. Similarly, further studies have also hinted at the power of positive affect, finding that people in whom positive affect has been induced can also organise concepts to maximise their differences (Isen &

Daubman, 1986). All of these sets of studies illustrate the idea that affect can be a major factor in performance on cognitive tasks.

However, even if positive affect is the best state for problem solving and organisation, Kort, Riley, and Picard (2001) argue that both positive and negative affective states are important in the learning process. They present a four quadrant model of how students learn that features two axes, dividing the model into four quadrants. As shown in Figure 2.1, the vertical axis represents the learning that is taking place, and the horizontal axis represents the valence of a particular emotion. Regarding the emotional axis, Kort et al. have identified five emotion sets, anxiety-confidence, boredom-fascination, frustration-euphoria, dispirited-encouraged, and terror-enchantment, that range from a positive valence of 1 to a negative valence of -1. Thus, for any given moment you could plot five different points for a student by using each of the five different emotion sets as the horizontal axis.

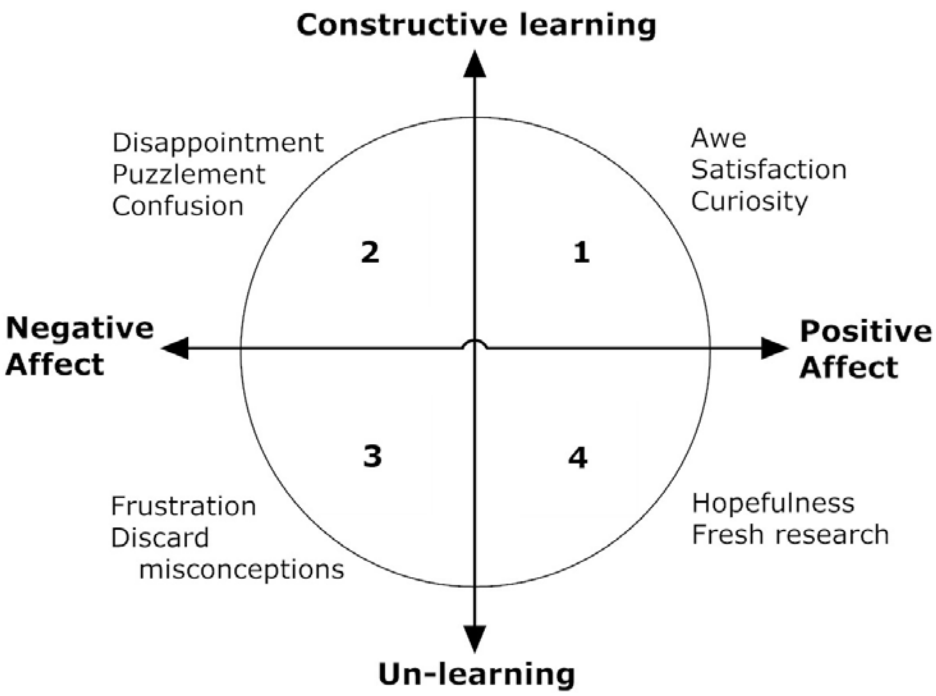


Figure 2.1. The four quadrant model of student affect and learning (Kort et al., 2001).

The main assertion behind the four quadrant model is that students typically orbit the quadrants in an anti-clockwise direction as they attempt to solve a problem. For example, a student might begin a new problem in Quadrant 1, in a state of curiosity and interest; the student will then move into Quadrant 2, in a state of confusion as to how to solve the problem. When an attempt to solve the problem fails, the student will then move into Quadrant 3, in a state of frustration as he discards his current approach; the student should then enter Quadrant 4, in a fresh attitude of hopefulness as he tries to come up with a new idea. When he does get a new idea, he will re-enter Quadrant 1, and the cycle will begin again. Of course, remembering that there are five different emotion axes at any given time, the student could be in several different quadrants at once with respect to particular emotional states. This model certainly appears to be supported by other studies of student learning in the ITS literature, which have found that moments of learning – that implies a positive affective state – tend to be preceded by moments of impasse – that implies a negative affective state (e.g. VanLehn, Siler, Murray, Yamauchi, & Baggett, 2003).

Therefore, the four quadrant model suggests that both positive and negative emotions are a natural part of a student's normal learning cycle. This further reinforces our point that emotion and learning are related, and also has an important implication for tutoring systems as the ability to identify a student's quadrant will help an intelligent tutor to give the most appropriate feedback to a student. If a student's learning process intrinsically involves a journey through distinct emotional states, then identifying these emotional states will help an intelligent tutor to speed this learning process on its way.

In conclusion, whatever we are to make of all of this, the most important point is that affective state *is* related to successfully tackling cognitive tasks. As shown by cases like Damasio's Elliot, emotional intelligence is critical to the ability to carry out cognitive tasks, and positive affect in particular has been shown to assist in problem solving and cognitive organisation. However, Kort et al. argue that negative affect is also a part of the learning process, and from this we conclude that tutors can help the student's learning process through an awareness of the student's affective state. Therefore a student's affective state *is* important to tutoring, and an affect-sensitive tutoring system is a valid endeavour.

2.1.2 How student emotions affect human tutoring

In the previous section we asserted that an awareness of student emotion would assist an intelligent tutor to help students through the learning process; this leads directly to the second issue that we will consider, that affective feedback is a major cue that *human* tutors use to constantly adapt their teaching strategies (Lepper, Aspinwall, Mumme, & Chabay, 1990; Lepper & Chabay, 1988; Fox, 1991; Merrill, Reiser, Ranney, & Trafton, 1992). Human tutors use the affective information they receive from learners to individualise the tutoring session, so that the learning is as effective and efficient as possible for each individual student (Kort et al., 2001). For example, a human tutor who can see that a learner is visibly confused will naturally try to simplify the material being presented, or perhaps revisit prerequisite material; this immediately meets the learner's need, and leads to greatly increased learning.

The affective feedback from students that human tutors adapt to can take many forms, including facial expressions, gestures and vocal inflections (Picard, 1998). Noteworthy is the fact that almost all affective communication is either non-verbal or paralinguistic; people seldom directly say "I am happy" or "I am disgusted", but usually this information is conveyed nonetheless through any or all of media described above. In fact, one famous study has estimated that up to 93% of communication takes place either non-verbally or paralinguistically (Mehrabian, 1971). The nature of these media is such that learners constantly transmit affective information during a tutoring session, whether consciously or not, because every facial expression, every gesture and every vocal inflection each carries affective data. In fact, due to non-verbal communication, it would be extremely difficult for a student *not* to communicate. Therefore human tutors constantly receive a significant amount of affective feedback during a tutoring session, which tells the tutor whether the learner is happy, confused, frustrated, surprised, bored or content with the material that the tutor is presenting. Human tutors use this information to adapt their teaching to match the affective state of the learner (Sarrafzadeh, Fan, Dadgostar, Alexander, & Messom, 2004).

The implication of emotions in human tutoring for artificial tutors is hard to overstate: if human tutors constantly adapt to the affective state of students, and if human tutors are generally accepted to be the most effective tutors that there are (as we saw back in Section 1.1), then it follows that enabling intelligent tutors to adapt to emotions too could potentially significantly enhance student learning. This provides a further basis for the current research.

Unfortunately though, extremely little is known about exactly *how* human tutors adapt to emotion. It is commonly accepted that human tutors cannot help but adapt to emotion, and it is also commonly accepted that good human tutors can do this very well, but very little research has actually addressed the details of *how* human tutors adapt to particular affective signals in particular teaching scenarios. This fact was significant when we came to ponder exactly how an ATS should respond to student emotion, and is discussed in more detail in the following chapter.

2.1.3 Summary

We conclude from the previous two sections that student affect is important to a tutoring scenario for two main reasons: firstly because emotions influence performance on cognitive tasks such as learning, and secondly because human tutors use affective feedback as a basis for the adaptation of their teaching strategies. It is clear that good human tutors appreciate the pedagogical implications of affect, yet existing ITSs adapt their teaching strategies based only on a model of the knowledge state of the learner. Therefore with human tutoring as our guide, we assert that the selection of teaching strategies by a tutoring system would be significantly enhanced if its model of the learner was extended to incorporate the affective state as well as the knowledge state of the learner. However, as we saw in the previous paragraph, emotions in human tutoring is a rough guide at best, and much more work needs to be done in this area; this issue is addressed in the following chapter.

2.2 Affective computing

Before we come to review ITSs that adapt to student emotion, we now consider the more general field of *affective computing*. Affective computing is generally traced back to Rosalind Picard's (1997) book of the same name, where she defines affective computing to be "computing that relates to, arises from, or deliberately influences emotions" (p. 3). Picard divides affective computing into four main areas, but for our purposes we can group the last two together:

1. Computers that recognise emotions;

2. Computers that express emotions; and

3. Computers that have emotions, and computers that have emotional intelligence.

We will now discuss previous work on each of these issues in turn, particularly in the context of an affect-sensitive ITS.

2.2.1 Recognising emotion

The first step towards making an affect-sensitive ITS is facilitating adding emotion to the system's student model; in other words, there is no way to adapt to the affective state of the student without first knowing what it is. Therefore recognising user emotion is absolutely foundational to any attempted implementation of an ATS.

Recognising user emotion relates to picking up affective signals, whether the user sends them consciously or not. When a user does experience a particular emotion, their body potentially transmits this information in many different ways, some of which are more obvious to humans than others. For instance, the most obvious media for transmitting affective state include facial expressions (which are the focus of this research), vocal inflections, touch, language, gestures and eye gaze. However, there are also media for transmitting emotions that mostly go unnoticed by humans, such as heart rate and skin conductance, which are also valid measures of affective state (although these are better as measures of general arousal than indicators of the kind of emotion being experienced).

Fortunately, a significant and ever increasing body of work has investigated these areas of emotion detection, and we briefly summarise some of this work now.

<u>Vocal inflections.</u> The prosodic qualities of speech, or vocal inflections, carry significant emotional content quite apart from the meanings of the words themselves. However, one study has shown that even humans find recognising emotions from neutral speech to be challenging, with a success rate of only about 60% (Scherer, 1981). Nonetheless, particular acoustic-prosodic features of speech such as pitch, energy, duration, tempo and pausing can all be used to gather affective state information

(Litman & Forbes, 2006), and encouraging progress is being made in these areas (e.g. Pantic & Rothkrantz, 2003; Shafran, Riley, & Mohri, 2003). Litman and Forbes (2006) present a system that is capable of distinguishing between positive, neutral and negative affect in student speech with up to 66% accuracy.

Motor and physiological cues. Some interesting work has been done to use motor cues to detect user interest and affect. Mota and Picard (2003) have developed a special chair that detects interest based on a user's posture with an accuracy of 87.6% when tested on new subjects. The affective computing lab at the Massachusetts Institute of Technology (MIT) has also developed a pressure sensitive mouse that can be used as a measure of frustration (Picard, 1997).

Physiological cues. Wearable sensors can be used to detect physiological cues such as skin conductance, heart rate and temperature (Picard, 1997; Prendinger & Ishizuka, 2007). Burleson (2006) used a wrist-band sensor to detect a user's skin conductance, which carries information about the level of the user's arousal. As they point out, a drawback of these technologies is that they require the user to attach the sensors to their body, which could potentially be considered intrusive (although perhaps a case could be made that *all* of these methods are intrusive, to a greater or lesser degree).

Probabilistic models of emotion. Another approach to including affect in the student model is to predict it, rather than detect it, on the basis of the student's interaction with the system. Conati (2002) presents one such model based on a Dynamic Decision Network: based on an individual student's personality traits and goals and their interaction with an ITS, it is possible to assess the most likely student emotional state, and the effect that tutor actions may have on this emotional state. Evaluations of probabilistic models have given promising results, although it is ultimately intended that these models also be integrated with information from sensors that detect affect as well (Conati & Maclaren, 2004; Conati & Maclaren, 2005).

Facial expression analysis. Finally, the other major method used to detect affective state is facial expression analysis, which is an area where significant progress has been made in the last decade (see Fasel & Luettin, 2003; Pantic & Rothkrantz, 2003 for surveys). Most approaches to facial expression analysis involve detecting minute facial movements referred to as *action units*, based on the *facial action coding scheme* (FACS) that was developed by Paul Ekman (Ekman & Friesen, 1978). The success of facial expression analysis systems can be tested using standardised databases of facial expression images such as the Cohn-Kanade or Ekman-Hagar datasets (Kanade, Cohn, & Tian, 2000; Donato, Bartlett, Hager, Ekman, & Sejnowski, 1999), and increasingly impressive results are being reported in the image processing literature (Fan, Sarrafzadeh, Dadgostar, & Gholamhosseini, 2006; Bartlett, Littlewort, Frank, Lainscesk, Fasel, & Movellan, 2006; Lucey, Matthews, Hu, Ambadar, De la Torre Frade, & Cohn, 2006).

Facial expression analysis was the approach that was chosen for the current research, as our research group at Massey University had already developed an in-house real time facial expression analysis system (Sarrafzadeh et al., 2004). The emotion classification is achieved by using support vector machines, and is able to detect the 6 basic facial expressions that are defined by Ekman (1992). The module uses a facial feature extraction algorithm that is not only able to extract all important facial features, but is also fast enough to work in real time. Unlike other facial expression analysis systems, facial information is automatically detected without the need to manually mark facial features. Initial tests on artificial data sets have yielded a successful recognition rate for the current version of the system of approximately 90% (Fan, Sarrafzadeh, Dadgostar, & Gholamhosseini, 2006).

2.2.2 Showing emotion

The second major area in affective computing is *showing* emotion; in other words, as well as recognising emotion, it is desirable for a computer system to be able to show emotion too. In fact, this second area of affective computing is much further developed than the first; although there are as yet relatively few systems that can recognise emotions, many have been developed with the ability to display simulated emotions. In the context of ITSs, emotions are generally shown through lifelike tutoring characters, also known as animated pedagogical agents (Prendinger & Ishizuka, 2004).

Animated pedagogical agents are “lifelike autonomous characters that cohabit learning environments with students to create rich, face-to-face learning interactions” (Johnson, Rickel, & Lester, 2000, p. 47). Gulz and Haake (2006) divide animated agents into three main groups:

- agents that take the role of teacher, e.g. *AutoTutor* (McCauley, Gholson, Hu, Graesser, & The Tutoring Research Group, 1998), *Whizlow* (Lester, Zettlemoyer, Gregoire, & Bares, 1999), *Adele* (Shaw, Johnson, & Ganeshan, 1999), *Cosmo* (Lester, Voerman, Towns, & Callaway, 1999), *Herman the Bug* (Lester, Stone, & Stellin, 1999), *Laura* (Bickmore, 2003) and *PPP Persona* (André, Rist, & Müller, 1998);

- agents that take the role of learning companion, e.g. *Trouble Maker* (Aïmeur, Dufort, Leibu, & Frasson, 1997) and *Steve* (Rickel & Johnstone, 2000); and

- agents that simply act as a guide, e.g. *Olga* (Beskow & McGlashan, 1997), *Jack* (Noma & Badler, 1997) and *Will* (Churchill, Cook, Hodgson, Prevost, & Sullivan, 2000).

Other examples of animated agents include *Baldi,* (Massaro, 2003), *GRETA* (Berry, Butler, & de Rosis, 2005) and *HCA* (Corradini, Mehta, Bernsen, & Charfuelan, 2005). All of these and the above agents are able to display emotions through channels such as facial expressions, gestures, language, posture and prosody.

The primary benefit of animated characters is that they carry a “persona effect”; the presence of a lifelike character can strongly influence students to perceive their learning experiences positively, over and above the positive effects that feedback from a text-based agent might give (Lester, Converse, Kahler, Barlow, Stone, & Bhogal, 1997; Prendinger, Mayer, Mori, & Ishizuka, 2003; Clark & Choi, 2005). The social qualities that an animated agent possess are believed to raise its believability and trustworthiness, which is important for users to take them seriously; the social nature of an animated agent also helps to engage the student, which means that the student will pay closer attention to the tutoring system (van Mulken, André, & Muller, 1998).

But although the persona effect has been shown to increase learner motivation, especially in technical domains, its overall benefits remain unclear. A study by van

Mulken, André, and Muller (1998) using *PPP Persona* found that the presence of an animated agent did not lead to a significant increase in student performance. They suggest that motivation and distraction may have acted as antagonising factors – the increase in motivation caused by the presence of the agent was offset because the agent actually distracted students from the task. If this is the case, there are two main ways that an agent could maximise the persona effect:

- by being as believable, trustworthy and engaging as possible,

- whilst at the same time being as unobtrusive as possible, and focussing the student on the task at hand.

2.2.3 Having emotion, and being emotionally intelligent

The other issues in affective computing relate to a computer's ability to have emotion, and its ability to act in an emotionally intelligent manner – that is, to be "skilled at understanding and expressing its own emotions, recognising emotions in others, regulating affect, and using moods and emotions to motivate adaptive behaviours" (Picard, 1997, p. 76). The first of these issues is a fascinating philosophical quandary – for instance, will users ever take affect-sensitive animated characters seriously if they know that the characters do not *really* feel the emotions that they simulate? However, fascinating though this question is, it is beyond the scope of the current research, and will have to be left for future work – except perhaps very briefly to comment that the famous *Media Equation* of Reeves and Nass (1996) may suggest that humans tend to interact with computers as if they *are* human, when they know that they are not.

On the other hand, the second issue, of computers acting with emotional intelligence, does have a significant bearing on our research into affect-sensitive tutors, for all the reasons outlined in Section 2.3.1 and especially Section 2.3.2. There is no need to repeat here what has been written earlier in those sections, but suffice to say that acting with emotional intelligence is absolutely essential for an affective tutor to be believable and positively engaging to students. In other words, even with the ability to accurately recognise and convincingly show emotions, an affect-sensitive tutoring system must know how to use those abilities in harmony with the user for either of them to be of any use. However, as we noted previously, exactly what emotional intelligence should mean to an affective tutor remains a very largely unanswered question; the next chapter of this thesis will deal with this subject in more detail.

2.2.4 Summary

In this section we have reviewed the field of affective computing, which we have divided into three main areas: recognising emotion, showing emotion, and having emotion/emotional intelligence. Of these three areas, research on showing emotion in ITSs is well established through the implementation of animated pedagogical agents, and research on recognising emotions is offering promising results through a number of different approaches, but little work has been done so far to describe emotional intelligence in the context of an artificial tutor.

The challenge for an ATS will be to successfully combine all three of these areas of affective computing: to create a system that can recognise emotion, that can show emotion, and that can respond to the student in an emotionally intelligent way. Much work remains to be done, but the state of the art in ATSs will now be reviewed in the following section.

2.3 Affective Tutoring Systems (ATSs)

The idea that ITSs could be enhanced by adapting to student emotion has spawned the developing field of Affective Tutoring Systems (ATSs): ATSs are ITSs that are able to adapt to the affective state of students in the same ways that effective human tutors do (Alexander, Sarrafzadeh, & Hill, 2006; de Vicente, 2003). The history of ATSs is remarkably short; it seems that the term *Affective Tutoring System* was first used only several years ago (Alexander, Sarrafzadeh, & Fan, 2003; de Vicente, 2003), although the popular concept of an ITS adapting to perceived emotion can be traced back at least as far as Rosalind Picard's book *Affective Computing* (1997). In this section we review the research in this area.

2.3.1 Related work

MOODS. de Vicente (2003) has developed a method of diagnosing the level of a student's motivation that incorporates both self-report and motivation diagnosis rules. The motivation diagnosis rules were generated by having participants study a recorded interaction between students and an ITS; the participants were instructed to make inferences about the student's motivational state, and to give reasons for their inference. These reasons were then molded into the set of motivation diagnosis rules, which were further refined by subsequent validation testing with expert teachers. de Vicente's motivation diagnosis methods were applied in *MOODS*, a simulated ITS environment, with encouraging results that showed a reasonable level of accuracy in determining the motivational state of students. Exactly how accurate these methods are compared to the emotion recognition methods that we discussed in Section 2.2.1 would certainly be a very interesting study.

Building on previous work by del Soldato (1994), MOODS includes a set of motivational planning rules and a set of affective dialogue rules that form the core of the system's adaptations to students. However, the focus of de Vicente's work was diagnosing motivation, not facilitating it, so the effectiveness of these rules were not tested. Also, the rules were particularly concerned with motivation – which is only a subset of affect – so they do not explicitly consider affective states such as confusion or frustration etc.

AutoTutor. The Tutoring Research Group (TRG) at the University of Memphis is adding an emotional component to their ITS *AutoTutor*. AutoTutor is a natural language-based tutor that has been successfully tested on about 1000 computer literacy and physics students, with significant learning gains in deeper level explanations as well as surface knowledge (Graesser, Lu, Jackson, Mitchell, Ventura, Olney, & Louwerse, 2004). AutoTutor has also performed well in a bystander Turing test, where participants were unable to distinguish between AutoTutor's responses and the responses of a real human tutor (D'Mello, Craig, Gholson, Franklin, Picard, and Graesser, 2005). Through an "emote-aloud" study of students using the system, it was found that the most significant affective states for AutoTutor students were frustration, confusion and boredom, and further investigation determined that expert judges were much more accurate than untrained peers in their analysis of learner emotions. They are currently working towards an "emotion classifier" that is able to reliably detect a student's affective state (D'Mello, Picard, & Graesser, 2007); they plan to gather this information from the student using real-time facial expression and posture analysis, as well as conversation cues (as AutoTutor uses natural language). Their method will incorporate both standard and biologically motivated classifiers to optimize the accuracy of their results.

Prime Climb. Conati (2002) has developed a probabilistic model of determining student affect that has been applied in *Prime Climb*, an educational game designed for 11 year old maths students that was developed by the Electronic Games for Education in Math and Science (EGEMS) group at the University of British Columbia. The model relies on a Dynamic Decision Network (DDN) to identify student affect, which features two types of assessment: diagnostic assessment and predictive assessment. The diagnostic assessment focuses on the effects of student emotion, which are the outward signs of feelings that were discussed in Section 2.2.1; when fully complete, the system will use tools such as facial expression analysis and biometric sensors to feed real-time information into the DDN. However, Conati argues that a diagnostic assessment alone will not always be enough, so the second half of the DDN is based on a predictive assessment of student emotion. The predictive assessment is based on the Ortony, Clore and Collins (OCC) model (Ortony, Clore, & Collins, 1988), where emotions are considered to be the result of an appraisal of how a current situation fits one's goals and preferences. Therefore the DDN assesses the affective state of the student based on how the student is likely to feel given his/her particular goals and preferences in a current situation within Prime Climb. The appraisal model of how particular goals, preferences and situations interrelate is designed based on relevant work in psychology, several Wizard of Oz studies, and simple common sense; an individual student's preferences and goals can be assessed in real time as the student interacts with the system.

A study evaluating the accuracy of the predictive assessment and the real time goal assessment produced encouraging, if not perfect, results (Conati & Maclaren, 2004), and it was found that the most significant weakness of the predictive assessment would be elegantly complemented by the diagnostic assessment. The predictive assessment has since been further refined (Conati & Maclaren, 2005), and future work will be to integrate the DDN together with accurate diagnostic sensors (such as biometrics), and then to integrate the whole of the probabilistic model with the student learning model that is being developed in parallel (Conati & Maclaren, 2004). Integration between the probabilistic model and eye tracking (as a measure of user attention patterns) is also being explored (Conati, Merten, Amershi, & Muldner, 2007; Conati & Merten, to appear).

<u>Affective Learning Companion.</u> Burleson (2006) at MIT has developed an *Affective Learning Companion* that is able to mirror several student non-verbal behaviours believed to influence persuasion and liking. The agent is attached to a game based on the Towers of Hanoi problem, and mirrors students in real time based upon input from several nonverbal communication sensors: a pressure mouse, a wireless skin conductance wrist band, a posture analysis seat and a camera that detects upper facial features (but not lower facial features that convey expressions such as smiling and tension). To test the effects of the mirroring, the real-time mirroring agent was compared to an agent that used prerecorded behaviours instead of real-time adaptation – the prerecorded behaviours were chosen by identifying the most common inputs from the nonverbal sensors in a previous pilot test and using those as the basis for an "average" response. However, testing with 11-13 year old students found no significant difference between the strength of the social bond created by the mirroring and prerecorded agents; this may have been because the prerecorded version was, after all, based on commonly occurring student affective states in the previous pilot study. The agent was also capable of two types of feedback: an affective response and a task-based response – it was found that the girls in the study were significantly more likely to favour the affective response than the boys.

The Affective Learning Companion is also able to detect frustration/help seeking behaviour from students using input from the nonverbal sensors with an accuracy of 79%, but that result is based on offline computation. Future work will be needed to refine the system to accurately sense these frustration/help seeking behaviours in real time, so that the agent can be continuously aware of the student's affective state as he/she completes the Towers of Hanoi exercise. Exactly how and when the agent will choose to intervene (or not) will be based upon the real time information from the sensors as well as the personality type of the student.

ITSPOKE. Litman and Riley (2006) continue their work with *ITSPOKE*, an ITS that helps students with physics problems. They plan to make the system able to adapt to the affective student by analyzing acoustic-prosodic features of student speech, in conjunction with natural language processing. However, they are still at the stage of improving the speech analyser, and research into how ITSPOKE will adapt to student emotion remains a task for future work.

Other. Researchers at Essex University and Shanghai Jiao Tong University plan to test a special emotion-aware tutoring system on students in China, with a view towards augmenting distance learning technology (Simonite, 2007). Students will wear a Bluetooth-enabled sensor ring on their finger, which will transmit heart rate, blood pressure and skin conductance data to the tutoring system; this data will be analysed to assess the user’s boredom and confusion levels.

2.3.2 Summary

Research towards a fully functional ATS is still in its infancy, across all of the work described above, although exciting work is being done in this area. Several groups have had impressive results in developing ITSs that are capable of recognising emotion, though even the most impressive of these results leave significant room for improvement.

However, of particular note is that the question of *how* to adapt to student emotion once it has been recognised has received little attention. Of the work reviewed above, the Affective Learning Companion at MIT is the only tutor to have designated affective feedback; however, this feedback has not been tested, as their real time emotion detection does not yet run in real time. Similarly, although MOODS does include rules for generating affective dialogue, these have not been tested either, and in any case they focus on motivation rather than affect as a whole. No system has yet been developed that responds to the affective state of students based on solidly grounded tutoring strategies; neither has an ITS yet been developed that fully integrates both the detection and the display of emotions through an affect-sensitive animated agent. Perhaps this is because how to adapt to student emotion has generally been seen as an issue to be tackled after problems of detecting student emotion have been resolved; in any case, investigating a method for how to adapt to the affective state of students forms a major goal of this thesis.

Chapter 3

Video study of human tutors

As we saw in Chapter 1, the overall aim of this thesis is to develop an Affective Tutoring System (ATS) that is able to detect student emotion, to display emotion of its own through an animated pedagogical agent, and to *adapt* to the student emotion that it is able to detect. This chapter describes an observational study of human tutors that was the first step in addressing this issue of enabling the ATS to adapt to student emotion, which was the first sub-goal of this thesis as outlined in Section 1.2. We will later show in Chapter 4 how the results of the observational study presented in this chapter were then used as the basis for a case-based reasoning tutoring strategies module that could recommend real time tutoring adaptations based on student affect. The case-based reasoning tutoring strategies module was in turn the core of the implementation of the ATS, Easy with Eve, that we will present in Chapter 5.

We begin this chapter by reiterating the aims of the observational study and by giving a brief background to previous studies of human tutors; then we give the methodology and results of the observational study. This is followed by the results of an inter-rater reliability study that validated the data from the observational study, before we conclude with a brief summary of the implications of the observational study to the overall development of Easy with Eve.

3.1 Aims of the study

As we saw in Chapter 2, it has now been possible for some years for Intelligent Tutoring Systems (ITSs) to express emotions through an animated agent, or even through text (e.g. Johnson, Rickel, & Lester, 2000; Prendinger & Ishizuka, 2004), and increasingly ITSs are also able to detect student emotions with reliable accuracy (e.g. Burleson, 2006; D’Mello, Picard, & Graesser, 2007). But even if an ATS could detect student emotion with 100% accuracy, it would still need to know what to do with this information before it could actually adapt its tutoring usefully. Similarly, the most expressive and believable animated agent in the world might be of little use if it did not know how to use its facial expressions to match the current tutoring scenario. However, as we also discussed in Chapter 2, very little, if any research has focussed directly on this issue of *how* artificial tutors should adapt to particular student affective states. In fact, even the ways in which human tutors adapt to the affective state of students have yet to be adequately explained.

Therefore, the aims of this observational study were twofold, with one minor and one major aim respectively:

- firstly, to gain some general insights into the ways in which student affect and human tutoring adaptations interact; and

- secondly, and most importantly, to gather a body of data that could be used as the basis for the tutoring strategies module of an ATS.

However, both of these aims assume the validity of using human tutors as the basis for an affect-sensitive tutoring strategies module, which is an issue that deserves elaboration. Therefore, before we present the methodology and results of the observational study itself, in the following section we give a rationale for the study of human tutors and a brief overview of previous work in this area.

3.2 Human tutors

Given that there is little research that has focussed on *how* an artificial tutor should adapt to student affect (as we saw in Section 2.3), the rationale for carrying out an observational study of how human tutors adapt to student affect had two main components: firstly, we know that human tutors are very effective at adapting to students, including adapting to student affect; and secondly there was no established body of research that addresses *how* exactly human tutors adapt to student affect on the level of individual facial expressions or other outward displays of emotion. Also, a third issue was that there was an indirect benefit of basing the tutoring strategies module of the animated agent on the tutoring actions of human tutors in that it would also help to maximise any persona effect – as discussed in Section 2.2.2 – that the animated agent might cause. We now discuss each of these three issues in turn.

The first reason for studying human tutors was because it is well established that human tutors are a very effective means of instruction, causing an improvement of two standard deviations in student performance over traditional classroom instruction (Bloom, 1984; Cohen, Kulik, & Kulik, 1982). In fact, the concept of the effectiveness of human tutors is extremely widely accepted in the ITS community, and as such has been the inspiration for a large body of ITS research (e.g. Merrill, Reister, Ranney, & Trafton, 1992; de Vicente, 2003; Kort & Riley, 2002; Person & Graesser, 2003; Heffernan, 2001; VanLehn, Siler, Murray, Yamauchi, & Baggett, 2003; Derry & Potts, 1998). To explain the effectiveness of human tutors, a number of studies have focussed on particular aspects of human tutoring, such as scaffolding, eliciting student self-explanation, helping students to correct their own misconceptions, and timely intervention (Chi, Siler, Jeong, Yamauchi, & Hausmann, 2001; VanLehn, Siler, Murray, Yamauchi, & Baggett, 2003; Fox, 1993).

However, human tutors have also been shown to demonstrate particular abilities at adapting to the affective and motivational states of the student. Kort, Riley and Picard (2001) assert that human tutors use their assessment of the affective state of students to help them progress through the emotional stages of learning in their four quadrant model of student learning, as we saw in Section 2.1.1. Similarly, Lepper, Aspinwall, Mumme, and Chabay (1990) argue that the highly interactive nature of human tutoring is intrinsically motivating, and that tutors are able to pitch material in such a way as to make it intriguing and challenging, whilst at the same time minimising negative emotions associated with failure; comparable results were found by Fox (1991) and Chi et al. (2001). Therefore, as human tutors are effective at adapting to the emotional state of students, they provided a logical model on which to base the tutoring strategies of an ATS; this was the first reason why an observational study of human tutors was carried out in the current research.

The second reason for studying human tutor adaptations to student displays of affective state was that previous work in this area does not explain how tutors adapt to students on the low-level of individual tutor and student expressions. This information was especially critical to the development of the ATS, as the animated agent needed to appear to adapt in real time to low-level displays of student emotion, but this had not, to the author's knowledge, ever been studied in detail before. For instance, the previous literature on motivation and affect in learning does not discuss appropriate tutor responses to individual student facial expressions that convey emotions such as happiness or frustration. The previous research most closely related to the study of low-level tutor adaptations to student affect is work towards the ITS *AutoTutor* (Person & Graesser, 2003) and the interaction coding of Chi et al. (2001), but while these studies do examine human tutor and student actions on a low-level, turn-by-turn basis, they do not explicitly consider the turn-by-turn affective state of either students or tutors. Therefore the current observational study of human tutors was necessary because there was no previous work that had directly addressed the question of how human tutors adapt to student displays of affective state on a low-level, turn-by-turn basis.

Finally, there was also an indirect benefit of using human tutors as a model for the tutoring strategies of the ATS in that it may have helped to increase the persona effect of the animated agent. Since by definition an animated tutoring character pretends to act like a human tutor, the more human-like it appears to act (through appropriate emotional awareness and expressions), the more believable the agent will be. As we have seen, this is important because the believability of animated agents is the foundation of the persona effect (van Mulken, André, & Muller, 1998). Therefore given a student's cognitive and affective state, the ATS would not only present the most appropriate tutoring material for these student states, but it would do so via an especially believable animated agent that aims to maximise the persona effect. When the ATS was tested, the participants were given a questionnaire to gauge the impact of the persona effect – this questionnaire and its results are presented and discussed in Chapters 6 and 7 respectively.

In summary then, the main aim of this study was to gather data about how human tutors adapt to student emotions, as well as student actions; this data was to be used to help develop the tutoring strategies module for the ATS, which will be discussed in Chapter 4. The main reasons behind this approach were twofold: firstly, that human tutors are very effective at adapting to students; and secondly that the observational study was necessary to fill a hole in the existing literature. Also, as an indirect benefit of this approach, the animated agent of the ATS would have a better chance of benefiting from the persona effect. We now present the methodology of the observational study.

3.3 Methodology

3.3.1 Participants

The observational study of human tutors involved videoing several tutors as they tutored students individually. There were three tutors altogether – two of the tutors were teachers at the school where the research was undertaken, and the third was a professional tutor. There were nine student participants, all of whom were 8 or 9 year old students at a school in Auckland, New Zealand. Five participants were tutored by the professional tutor; the other two tutors tutored two students each. All of the three tutors were female; the nine students comprised four males and five females.

Ethical approval was sought and obtained from the Massey University Human Ethics Committee before the participants were contacted and the study carried out. A copy of the ethics application and the accompanying consent forms and information sheets are given in Appendix G.

3.3.2 Procedure

The day before the study began, a familiarisation session was held between the professional tutor and the nine students to ensure that all the students would be familiar with the exercise that would be used for the tutoring session. The familiarisation session also gave the students a chance to meet the professional tutor so that the performance of the five students that the professional tutor tutored would not be significantly affected by shyness at working with a stranger.

The study was carried out over two days at the local school that the nine students attended. Students were called into a meeting room one at a time; the other people present were two researchers and one of the three tutors. Each participant was tutored for about 20 minutes as they worked on a mathematics exercise, and both the student and the tutor were videoed for the duration of the tutoring session. There was a total of approximately 3 hours of video footage.

3.2.3 Tutoring exercise

The domain that was chosen for the observational study was the concept of part-whole addition. The study used an existing exercise developed by the New Zealand Numeracy Project (2003) that encourages students to add numbers by transforming the initial equation to make the first addend up to the next 10. For example, 17 + 6 would become 17 + 3 (to make 20) + 3 = 23. Students learn this reasoning by manipulating tens frames and counters, as shown in Figure 3.1: in this example the student should move three counters from the tens frame on the right across to make the middle tens frame up to ten. As students improved at using the counters to add the two addends, the exercise became gradually more difficult, with the counters being either covered by a sheet of card – to force the student to imagine manipulating the counters – or taken away altogether. These different levels of difficulty in the exercise are discussed in further detail in the context of the development of Easy with Eve in Section 5.2.1.

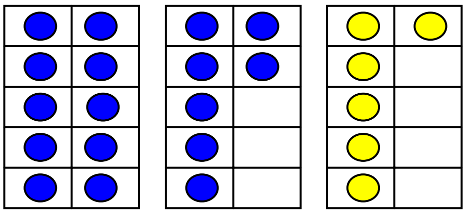


Figure 3.1. Tens frames and counters in the mathematics exercise.

3.2.4 Coding the videos

To analyse the videos, a coding scheme was developed that expands on previous work by Person and Graesser (2003). This scheme was used to extract data from each tutoring video to describe the behaviours, facial expressions and expression intensities of students and tutors.

The coding scheme was used to divide each tutoring video into several hundred clips, with each clip indicating a single student or tutor “turn” (behaviour) in the tutoring dialogue; this approach was validated by the positive results of the inter-rater reliability study that is discussed in Section 3.4. The student and tutor turns that were used for the coding scheme are shown in Tables 3.1 and 3.2 respectively. The facial expressions that were used for the coding scheme are shown in Table 3.3. The intensity of each expression was initially rated on a scale from 1 to 5 (from low to high); afterwards these were reclassified as *low* for values 1 to 3 and *high* for values 4 to 5. Neutral expressions were assigned a low intensity by default. The coding scheme was applied to each of the clips in the videos, thus generating the raw data of the study; the reliability of this data was established by an inter-rater reliability study that is described below in Section 3.4.

Table 3.1. Frequency of student turns used in the coding scheme.

CODE	STUDENT TURN	FREQ	%AGE
1	Correct answer / action	820	74
2	Partial answer	23	2
3	Error-ridden answer	106	10
4	No answer	70	6
5	Related question	14	1
6	Unrelated question	0	-
7	Valid statement	20	2
8	Invalid statement	5	-
9	Reminding example	0	-
10	Meta-comment	7	1
11	Acknowledgement	20	2
12	Complaint	0	-
13	Thinking aloud	5	-
14	Thinking silently	0	-
15	Other	16	1
	TOTAL	**1106**	**100**

Table 3.2. Frequency of tutor turns used in the coding scheme.

CODE	TUTOR TURN	FREQ	%AGE
1	Pose initial problem	8	-
2	Pose harder problem	63	3
3	Pose easier problem	6	-
4	Pose similar problem	71	4
5	Ask new question	336	18
6	Ask about error	5	-
7	Request clarification	55	3
8	Pump for additional information	454	24
9	Assess knowledge of topic	0	-
10	Global assessment	0	-
11	Positive immediate feedback	276	14
12	Positive delayed feedback	1	-
13	Neutral immediate feedback	285	15
14	Neutral delayed feedback	2	-
15	Negative immediate feedback	3	-
16	Negative delayed feedback	1	-
17	Reminding example	7	-
18	Hint	64	3
19	Answer own question	14	1
20	Answer student question	10	1
21	Rearticulate / discuss problem	38	2
22	Rearticulate / discuss question	44	2
23	Rearticulate / discuss solution	123	6
24	Comment about tutor ability	1	-
25	Comment about student ability	15	1
26	Comment about problem	6	-
27	Complaint	0	-
28	Other	22	1
	TOTAL	**1910**	**100**

Table 3.3. Frequencies of student and tutor expressions.

EXPRESSION	FREQUENCY (%)	
	STUDENT	TUTOR
Neutral (low)	62	86
Neutral (high)	-	-
Happy (low)	23	9
Happy (high)	6	3
Confused (low)	3	-
Confused (high)	-	-
Frustrated (low)	-	-
Frustrated (high)	-	-
Disappointed (low)	-	-
Disappointed (high)	-	-
Bored (low)	-	-
Bored (high)	-	-
Surprised (low)	-	-
Surprised (high)	-	-
Apprehensive (low)	4	1
Apprehensive (high)	-	-
Disgusted (low)	-	-
Disgusted (high)	-	-

3.2.5 Formatting the data

The nine tutoring videos were divided into over 3000 sequential clips of student and tutor turns. Each of these clips was coded to describe the actor's behaviour, facial expression, and intensity of facial expression.

Whenever sequences of tutor turns or student turns appeared in the data, they were compressed into a single turn so that there was a 1:1 ratio between student and tutor turns. This led to the creation of 6 new kinds of student turn and 112 new kinds of tutor turn, such as "tutor gives positive immediate feedback and pumps for additional information" – a composite of the existing tutor turns 11 and 8 (see Table 3.2). This made the data much easier to work with. The complete lists of composite tutor and student turns, and their frequencies in the data, are given in Appendix A.

3.3 Results

As the data from the video study are mined using case-based reasoning in the tutoring strategies module (Chapter 4), it was not strictly necessary to manually carry out a detailed analysis of the data. Nonetheless, to give a flavour of the interactions between the tutors and students in the study, in this section we discuss:

1. The overall frequency of student and tutor turns;

2. The frequency of tutor turns following correct/incorrect student answers;

3. The overall frequency of student and tutor expressions;

4. The frequency of tutor expressions following particular student expressions.

3.3.1 Overall frequencies of student and tutor turns

Tables 3.1 and 3.2 show the respective frequencies of student and tutor turns in the coding scheme. Perhaps unsurprisingly, almost all student turns were related to answering questions, with these turns occurring a total of 92% of the time. The occurrences of tutor turns were more widely spread across the coding scheme, *although ask new question*, *pump for additional information*, *positive immediate feedback* and *neutral immediate feedback* between them totaled 71% of all tutor turns. Negative feedback was almost never used by tutors.

3.3.2 Frequencies of tutor turns following correct/incorrect student answers

Figures 3.2 and 3.3 below show the frequencies of the most common tutor turns following correct and incorrect student answers respectively (as there were well over a hundred tutor turns it is not convenient to show the frequencies of them all). Unsurprisingly, the most common tutor responses to a correct student answer involve combinations of neutral or positive feedback and asking the next question. In particular, *pump for additional information* was the most frequently occurring response, at 18%. On the other hand, the most common response to an incorrect student answer was *hint*, at 23%, comfortably ahead of *rearticulate/discuss question* at 13%.

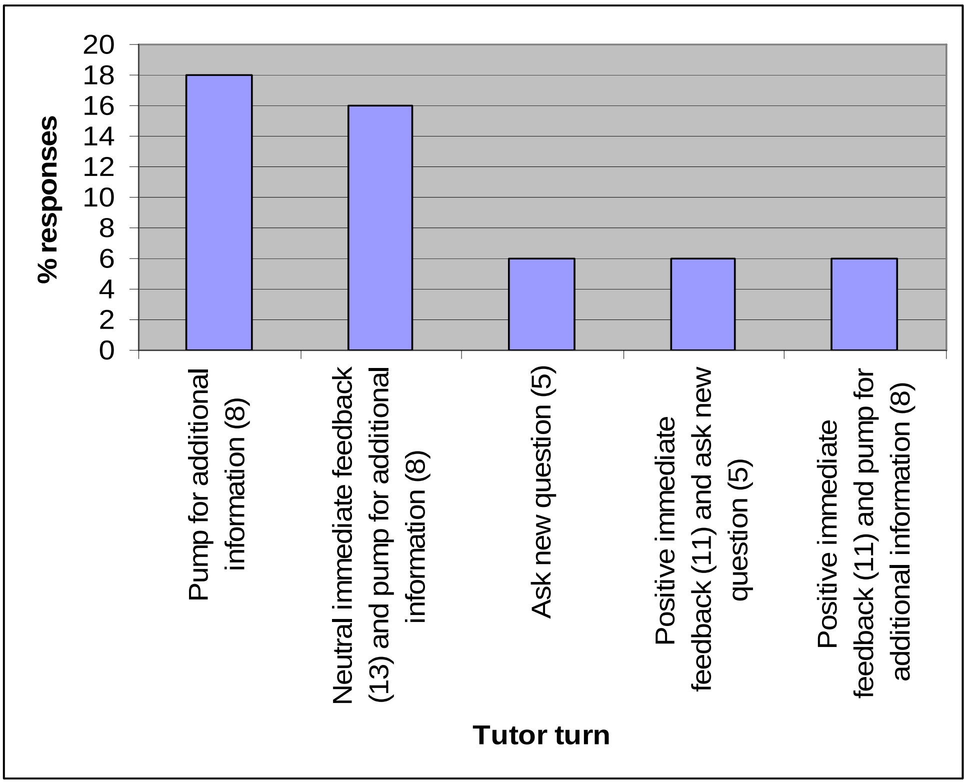


Figure 3.2. Frequencies of the most common tutor responses following correct student answers. The codes for the tutor turns are given in brackets.

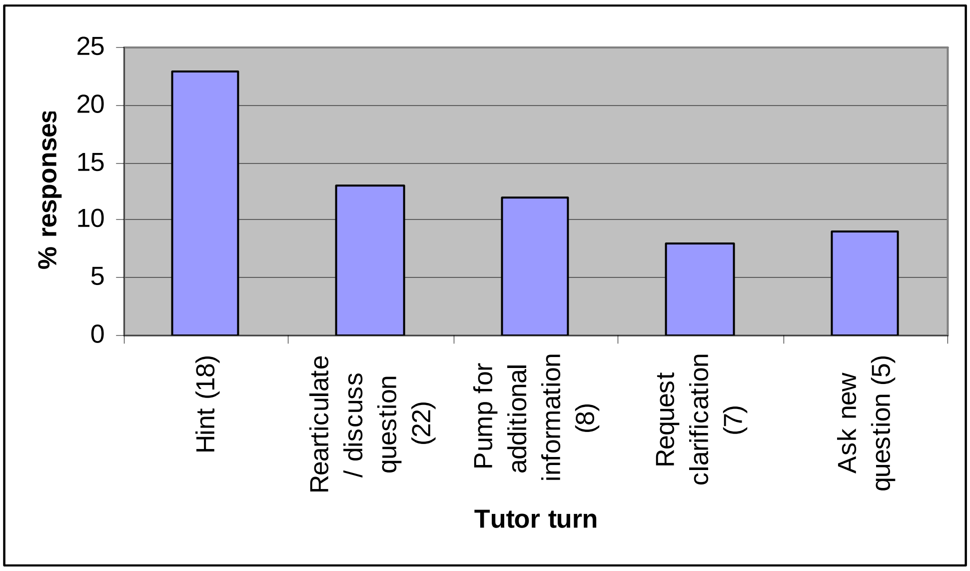


Figure 3.3. Frequencies of the most common tutor responses following incorrect student answers. The codes for the tutor turns are given in brackets.

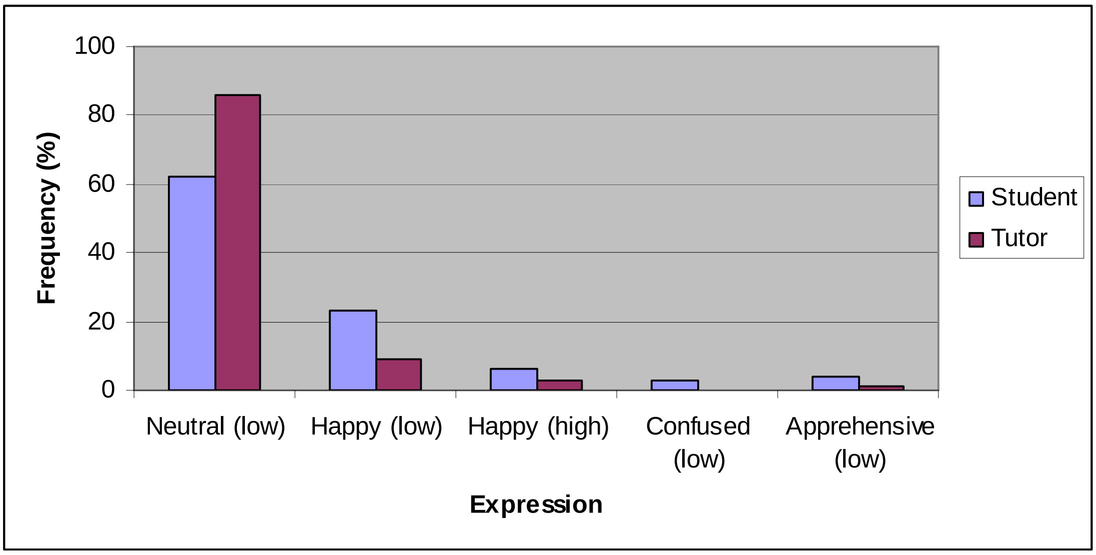


Figure 3.4. Frequencies of facial expressions for tutors and students.

3.3.3 Overall frequencies of student and tutor facial expressions

The frequencies of student and tutor facial expressions are shown in above Table 3.3, and graphically in Figure 3.4. For both students and tutors, neutral expressions were by far the most common: this was especially the case for tutors, for whom 86% of all expressions were neutral. The second and third most commonly appearing expressions were also the same for both students and tutors, with smiling (low) the second and smiling (high) the third most common expressions. However, smiles were much more common for students than tutors: students smiled for a total of 29% of student turns, whereas tutors smiled for only 12% of tutor turns. Students also appeared apprehensive (low) for 4% of turns and confused (low) for 3% of turns, but the combined occurrences of all other student expressions totaled only 2% of student turns. Similarly for tutors, the combined occurrences of all expressions other than neutral and smiling (including confusion and apprehension) totaled only 2% of tutors' turns. Expressions of frustration, disappointment, boredom, surprise and disgust were almost entirely absent for both students and tutors.

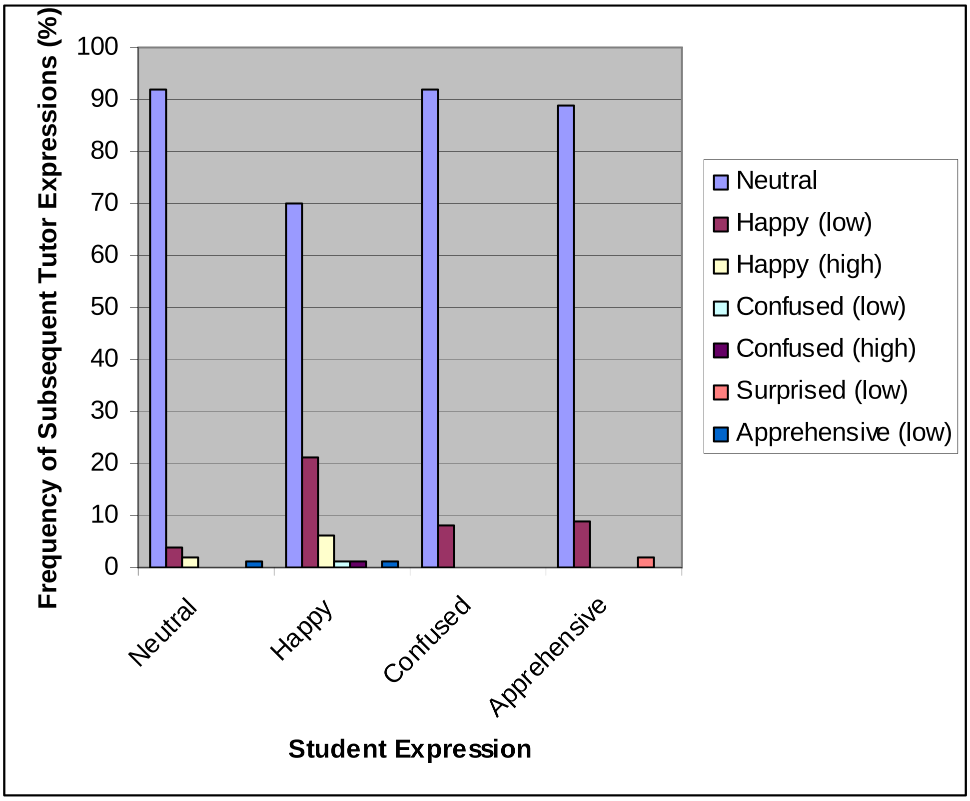


Figure 3.5. Frequencies of tutor expressions following student expressions.

3.3.4 Frequencies of tutor expressions following particular student expressions

Tutor expressions were clearly influenced by the student expressions that immediately preceded them. For instance, as shown above in Figure 3.5, confused and apprehensive student expressions almost always resulted in neutral tutor expressions, as did neutral student expressions: all three expressions were followed by a neutral tutor expression over 90% of the time. Figure 3.6 shows how the likelihood of a tutor smiling was significantly affected by whether or not the student was smiling, and the intensity of student smiles. Combining low and high intensity tutor smiles, tutors smiled in only 6% of turns following neutral student expressions – exactly half the average of 12% across all student expressions – but this frequency increased to as much as 45% when students smiled with high intensity. Therefore the probability of tutor smiles significantly rose as students smiled with increasing intensity.

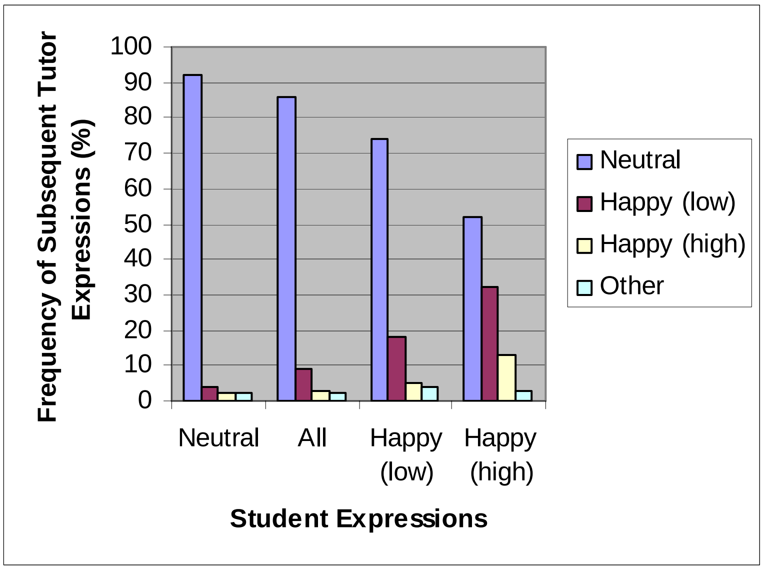


Figure 3.6. Frequency of particular tutor expressions following neutral, all, smile (low) or smile (high) student expressions.

3.4 Inter-rater reliability study

To validate the reliability of the data gathered by the video study of human tutors, a small inter-rater reliability study was carried out. The result of the study was that the data was found to be reliable for use in the ATS.

3.4.1 Methodology

One undergraduate student was first trained in the coding scheme and then tested on an initial trial set of 25 turns. As the initial trial was successful, the coder proceeded onto the actual coding.

Approximately 20% of the entire set of video clips was then re-coded following the description in Section 3.2.4, with 75 clips re-coded from 8 of the 9 student videos from the observational study. The sets of clips that were recoded were selected to make the sample as representative of the whole as possible, and included a range of expressions and turns from both students and tutors.

3.4.2 Results

Unfortunately it was not possible to use kappa values to measure the inter-rater reliability in this study, as there were differences in the numbers of categories that were used (it is not possible to use a kappa value when one rater uses a particular category but another coder never uses it.) However, we were able to measure the percentage similarity between the two sets of data, and examine any anomalies in the results on a one-by-one basis.

The results for the three areas of the coding scheme will now be discussed individually.

<u>Turns.</u> The overall rate of identical classifications for tutor and student turns was 325/600, or 54.2%. This can be broken up into tutor and student turns: tutor turns were classified the same at a rate of 146/373, or 39.1%; student turns were classified the same at a rate of 179/227, or 78.9%.

On first the glance, the percentage of identical tutor turn ratings seems to be extremely low, which in turn makes the overall rate of identical classifications seem low. However, the distinctions between several of the turns in the coding scheme were not clear, and were accordingly consistently classified differently. The two examples of this were between turns 5, 6, 7 and 8, to do with asking another question (see Table 3.1 above), and turns 13 and 23. In the case of turns 5, 6, 7 and 8, it was clear that they are all to do with asking another question, and are all “close enough”; in the case of turns 13 (*neutral immediate feedback*) and 23 (*rearticulate / discuss solution*), it became evident that the second rater was not given clear enough instructions about what constituted negative feedback, and that there was no real discrepancy between what both coders had understood to have taken place. Also, the second coder was poorly placed to distinguish between turns 1, 2, 3 and 4, which all related to the difficulty of a new problem, as she was reasonably unfamiliar with the mathematics exercise in the videos. Finally, turns classified as 28 (*other*) were notoriously difficult to classify, and should be disregarded as they related to conversation unrelated to the mathematics exercise; again, confusion here was caused by a lack of clarity in the coding scheme. When all of these factors were taken into consideration, the rate of matching tutor turn classifications rose to 287/366, or 78.42%, which is acceptable.

Similarly, there was understandable confusion between student turns 1 (*correct answer*) and 2 (*partial answer*), which both related to correct answers to questions. In hindsight, the coding scheme here needed to be more clearly defined, as each *partial answer* was also by definition a *correct answer* to some extent. On the other hand, it was very clear when the student answered incorrectly (turn 3), or not at all (turn 4). When this distinction was allowed for, the rate of matching student turn classifications rose to 202/223, or 90.58%, which is good.

<u>Expressions.</u> The overall rate of identical expression classifications was 455/600, or 75.83%, which is acceptable. However, of the 145 non-matches, a surprisingly large 80 non-matches occurred when the first coder saw a neutral expression but the second coder saw a smiling expression at the lowest possible intensity ('1' on a scale from 1 – 5), which is a slender difference. If those 80 non-matches are disregarded, the rate of identical classifications becomes 455/520, or 87.50%, which is good.

Possibly there was a slight gender bias here between the two coders, as the first coder was male and the second coder was female; there is much relevant literature that suggests that females are better at detecting facial expressions than males (e.g. Brody, 1999; Rotter & Rotter, 1988; Briton & Hall, 1995; Thayer & Johnsen, 2000), which may explain why the second (female) coder often saw slight smiles when the first (male) coder did not.

<u>Intensity.</u> The rate of identical expression intensity classifications was 581/600, or 96.83%, which is very high.

3.4.3 Statement of reliability

In summary, when slight anomalies in the coding scheme were allowed for, the inter-rater reliability for classifying both tutor and student turns was found to be acceptable. The inter-rater reliability for classifying facial expressions was acceptable, and when allowing for a potential gender bias it was good. The inter-rater reliability for classifying the intensity of expressions was good. Therefore it was considered acceptable to incorporate the data in the tutoring strategies module of the ATS, whilst bearing in mind that the coding scheme could be improved. These improvements will be considered as part of Chapter 7.

3.5 Summary

This chapter has presented an observational study of human tutors that was designed to gather data about tutor adaptations to students that could be used in the development of a tutoring strategies module for an ATS. The data were validated by an inter-rater reliability study, although this study has helped to identify improvements that could be made to the coding scheme. Several interesting findings from the study were as follows:

- Students, and especially tutors are predominantly neutral in their facial expressions;

- Apart from smiles, non-neutral expressions were rare for both students and tutors. Smiles were much more common for students than tutors;

- Tutors were more likely to smile following a student smile; they were very unlikely to smile following a student neutral expression.

However, the main outcome of the study was simply the collection of the data itself; this data is mined by the case-based tutoring strategies module of the ATS, which is discussed in the next chapter.

Chapter 4

Case-based reasoning tutoring strategies module

This chapter describes a novel, fuzzy case-based reasoning approach for developing the tutoring strategies module of an Affective Tutoring System (ATS). The aim of the tutoring strategies module is to consider both the student's affective state and answers to questions, and thus to suggest appropriate tutoring actions for the ATS to carry out. As a basis for ensuring that the suggested tutoring actions were indeed appropriate, the tutoring strategies module was based on the actions of real human tutors; the module uses the data gathered from the study of human tutors that was discussed in Chapter 3.

We begin by giving a background to existing approaches to tutoring modules; then we discuss why case-based reasoning was chosen as the approach for this research, and how this approach was novel. Then we talk about the first version of the case-based tutoring strategies module, which searched the data for exact tutoring scenario matches, and the second, fuzzy version of the tutoring strategies module which searched for similar matches as well as exact matches. We conclude with a summary of the chapter.

4.1 Background

Tutoring strategies have been an essential component of Intelligent Tutoring Systems (ITSs) since they were first conceived several decades ago (Self, 1999). This is because as well as maintaining information about the state of the student, it is also vital that an ITS should know *what to do* with this information before it can adapt intelligently. In other words, tutoring strategies are the part of an ITS responsible for guiding the interaction between the tutor and the student; this underscores the critical importance of the soundness of tutoring strategies in an ITS.

In the past a variety of methods have been used to represent tutoring strategies, most notably including methods such as procedures, plans, constraints, model tracing, and rules (Murray, 1999; du Boulay & Luckin, 2001; Merrill, Reiser, Ranney, & Trafton, 1992). However, all of these methods share a common theme in that they presuppose a correct course of action given a certain situation: for example, if ever the student is in state X, then event Y is a correct course of action for the ITS to carry out. An illustration of this can be seen in the tutoring strategies for the ITS *AutoTutor* (Person, Graesser, Kreuz, Pomeroy, & The Tutoring Research Group, 2001), which are made up of 15 production rules in the format:

IF [condition based on current student state/action]
THEN [tutor action]

However, so far as this thesis (about adapting to affect) is concerned, a fundamental weakness of these common methods of representing tutoring strategies is that they assume knowledge on the part of the developer about the best ways to adapt to a student in a given scenario (Elorriaga & Fernández-Castro, 2000). What if this knowledge about the best way to adapt to a student is shaky at best, or at worst, does not exist at all? As we discussed in Section 2.3.2, this is very much the case when we come to consider the best tutor adaptations to student emotions. For instance, it is currently not at all clear what the best tutor response is to a student who is smiling, or appears to be confused; this means that it would be impossible to deal with this situation in an "if-then" approach, as nobody is quite sure what the "then" would be. Of course, this is not to say that rules for adapting to the affective state of students could *not* be derived from an in depth analysis of human tutors interacting with students, but such a study has yet to be carried out, and would certainly be an extremely significant undertaking if it was. For instance, the 15 production rules in AutoTutor were a decade in the making, involving hundreds of hours of videotaped tutoring sessions and thousands of pages of transcripts

(Graesser, Person, Harter, & The Tutoring Research Group, 2001); similarly, it is likely that classifying helpful tutoring adaptations to student emotion into some kind of rule-based method would be a difficult and lengthy undertaking.

Therefore, it was important that a different method for determining tutoring strategies for adapting to student emotion was used in the current research, that would sidestep the lack of existing knowledge in the ITS, psychology and education literatures in this area. The method that was chosen to meet this need was case-based reasoning; this approach is now discussed in the following section.

4.2 A case-based approach

In this section we discuss the case-based reasoning approach that was used to implement the tutoring strategies module based on the observational study of human tutors in Chapter 3. We begin with an overview of case-based reasoning; then we discuss in further detail exactly why case-based reasoning was chosen for our approach, and how our particular application of case-based reasoning was novel.

4.2.1 An overview of case-based reasoning

Case-based reasoning is defined as a strategy that “uses an explicit database of problem solutions to address new problem solving situations” (Luger, 2002, p. 275). Many cognitive psychologists believe case-based reasoning to be widely used in everyday life (e.g. Griggs & Cox, 1982); for instance, if someone’s current problem was to cook a batch of blueberry muffins, they might base their solution on their previous experience (cases) of cooking muffins. Lawyers use case-based reasoning when they study the history of cases that are similar to their current case, and programmers use case-based reasoning when they reuse old code to solve a new problem. Another common metaphor for case-based reasoning is the way that doctors review past cases when making a new diagnosis: *CASEY* (Koton, 1988) and *PROTOS* (Bareiss, Porter, & Weir, 1988) are two classic examples of case-based reasoning systems that have been applied to medicine. In fact, human tutors themselves use case-based reasoning all the time when they decide how to teach a particular lesson based on their case-history of teaching similar lessons or students.

There are four commonly agreed-upon stages when using case-based reasoning to solve a problem (Luger, 2002):

1. Retrieve: search in memory for cases that are relevant to the current problem.

2. Reuse: modify the solution from a previous case so that it can be used to solve the current problem.

3. Revise: try the solution out, and see what happens. If the outcome is not desirable, a revision to the solution may be suggested.

4. Retain: store the current problem, solution and outcome in memory for future reference.

The main benefit of case-based reasoning is that it allows problems to be solved without needing to reference a set of rules that describe a domain; creating sets of rules can be notoriously time-consuming for non-expert developers in complicated domains (such as tutor adaptations to student emotion). This means that the system does not need to understand *why* a solution does or does not work, so long as it can tell from its database of past cases what *will* or will not work. Of course, on the other hand, this strength of case-based reasoning systems is also a weakness, as they are unable to provide deep explanations for why a solution is valid (Luger, 2002). However, if these deep explanations were easy to come by using some other method, which they often are not, then there would be much less need for case-based reasoning in the first place.

Case-based reasoning has been applied to ITSs in the past for several different purposes, such as modelling students by evaluating their current problem solving ability based on their success with previous, similar problems (González, Burguillo, & Llamas, 2006; Han, Lee, & Jo, 2005; Shiri, Aimeur, & Frasson, 1998); case-based reasoning has also been used as a teaching tool in its own right by reminding the student of previous cases related to a problem (Khan & Yip, 1995; Schank & Edelson, 1990). Notably, it has also been applied as a tool for guiding instructional planning (Soh, 2006; Du & McCalla, 1991; Riesbeck & Schank, 1991; Khan & Yip, 1995), which can be considered a broad form of tutoring strategy, with which this chapter is concerned.

However, a broad form of tutoring strategy such as instructional planning would not help much with the problem of continuously adapting to a student's affective state in real time. What was needed was a much finer application of case-based reasoning that could inform the system about the minute tutoring actions and displays of empathy that would be appropriate based on a case-history built from minute descriptions of tutor-student interactions. A summary of exactly why case-based reasoning was chosen as our approach, and how it was novel, is given in the following section.

4.2.2 A novel application of case-based reasoning

Given the foregoing overview of case-based reasoning, the rationale for adopting case-based reasoning as our approach for developing the tutoring strategies module of the ATS can be summarised by three main reasons:

1. It allowed the suggestion of tutoring adaptations based on what real human tutors had done in similar scenarios in the observational video study discussed in Chapter 3;

2. It would not require a deep understanding of how human tutors adapt to student emotions (which is helpful because as we mentioned in Section 2.3.2, nobody *does* fully understand this); and

3. Eventually it would allow the system to learn over time as it added new cases to its memory (although adding this functionality to the tutoring strategies module remains future work).

The reason that our application of case-based reasoning to the guidance of tutoring strategies was novel was because of the level of detail with which the actions of the tutoring system are prescribed. Previous case-based instructional planners have tended to work at a high level, relating to relatively broad issues such as which problem to present next to a student; in contrast, our application of case-based reasoning focuses at a very low level, guiding every distinct tutoring action of the tutoring system. Similarly, whereas cases in previous case-based instructional planners have tended to be broad in scope, relating to overall problems, cases in our tutoring strategies module are narrow in scope, relating to low-level interactions between the tutor and the student.

Exactly how case-based reasoning was applied in the implementation of the tutoring strategies module is presented in the following two sections, beginning with the earliest version of the module.

4.3 The basic case-based module

The basic module of the tutoring strategies module searched for exact matches in the case-base to the current scenario (the final version of the module incorporated fuzzy rules to also search for similar matches, which is described in Section 4.4). In this section we first describe the way that the current scenario was represented as a sequence of tutor-student interactions, and then we present the algorithm for recommending a tutoring action based on searching for exact matches to the current scenario in the case-base.

4.3.1 Scenarios as sequences of interactions

As an introduction to the design of the case-based tutoring strategies module, it is important to first describe the flexibility with which cases are represented in the module: as this section shall explain, the data that is stored in the case-base is the sequence of tutor-student interactions from the observational study in Chapter 3, and any sequence of interactions of any length in this data can be considered as a distinct case.

We saw in Chapter 3 that the observational study of human tutors yielded some 3000 interactions between the tutors and students, and it would be tempting in a case-based reasoning system to consider these interactions in isolation from all but the one directly preceding interaction. Indeed, as we saw in Section 3.3, even at this simple level the data from the observational study of human tutors contain a wealth of information about the interaction between tutors and students during the tutoring process. For instance, for any given combination of student turn, facial expression and intensity of facial expression, the following information is readily available:

- the frequencies in the data of all the tutor turns that immediately followed this combination of student states, and

- the frequencies in the data of all the tutor facial expressions (and intensities) that immediately followed this combination of student states.

However, a human tutor's response to a tutoring scenario is influenced by the *history* of his/her interactions with the student throughout the tutoring session – not just by what the student has only just done. Implicit in the data is the way that a human tutor's adaptations vary according to the history of his/her interactions with a student during a tutoring session; it was necessary that this information was mined for the tutoring strategies module to be realistic. Therefore, as well as considering the current cognitive and affective states of the student, the tutoring strategies module of an ATS also needed to consider the *sequence* of interactions leading up to any given tutoring scenario.

This means that *cases* in the tutoring strategies module can relate to any possible sequence of interactions in the data, regardless of their length. In other words, a particular interaction between a tutor and a student could be searched for as a case in its own right, or it could be searched for as part of a case that is a sequence of interactions that is 2 or 5 or 10 interactions long.

Thus the "current scenario" in the context of a tutoring situation refers to a sequence of interactions between the tutor and student: this is the sequence that is searched for in the case-base in the tutoring strategies module. In the basic version of the tutoring strategies module, only exact matches with the entire sequence of interactions were searched for; this algorithm is presented in the following section.

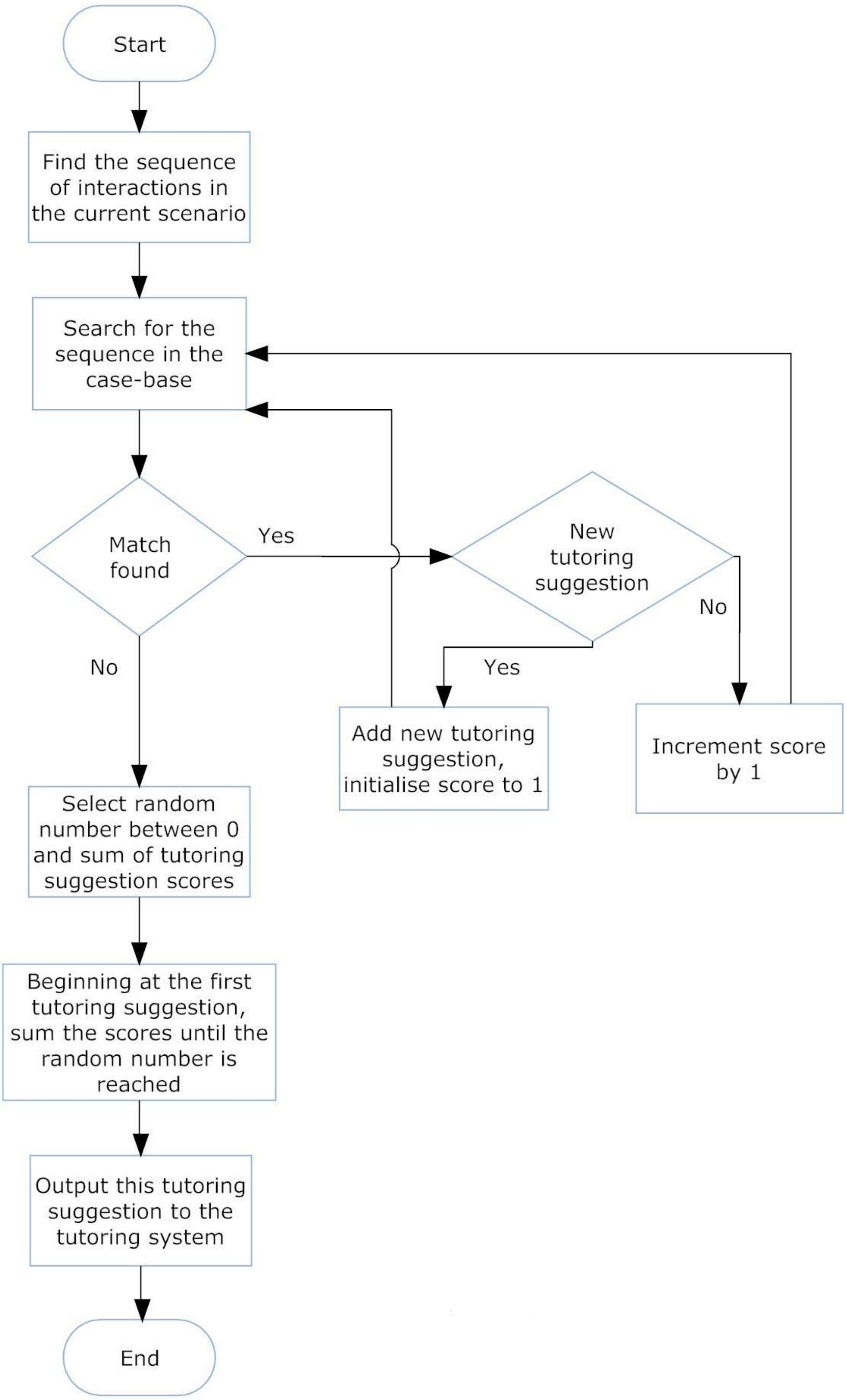


Figure 4.1. Flow chart of the algorithm for the basic version of the tutoring strategies module.

4.3.2 Using case-based reasoning to adapt to tutoring scenarios

The basic version of the case-based tutoring strategies module searched the data from the video study of human tutors for exact case matches with the current scenario, and then output a recommended tutoring action based on what the human tutors did next in the same scenario. As illustrated in Figure 4.1, the particular steps in the algorithm for the basic version of the tutoring strategies module were as follows:

1. Get the sequence of interactions in the current scenario. This will be a sequence of student and tutor turns, expressions and intensities in the same coding scheme format as the data from the study of human tutors.

2. Search for this sequence in all of the data from the human tutor study (which forms the case-base for the module).

3. Whenever a match is found, add what the human tutor did next (the tutor turn, expression and intensity) to a list of potential suggested tutoring actions, and assign that particular tutoring action a score of 1. If the tutoring action is already in the list of potential suggested tutoring actions, simply add 1 more to that tutoring action's score.

4. If no matches at all are found in the data, then shorten the length of the sequence by one entry by removing the oldest interaction from the sequence. Then repeat Steps 2 to 4 as required. If one or more matches *has* been found, then once all the data has been searched a list of weighted potential suggested tutoring actions will have been generated.

5. Take the sum of all the scores of the potential suggested tutoring actions, and then choose a random number between 0 and the sum.

6. Then beginning at the first member in the list of potential suggested tutoring actions, keep adding the scores in the list to a cumulative total of all the scores in the list. Stop when the random number that was generated in Step 5 has been reached, and make a note of the current member in the list.

7. Output the current member at the end of Step 6 as the suggested tutoring action for a tutoring system to implement.

Thus the data from the observational study of human tutors are searched for matches with a current tutoring scenario. This search generates a weighted list of potential suggested tutoring actions, and one of these actions is then chosen randomly – yet also according to the weights of each suggested tutoring action. The higher the score of an individual tutoring action, the more likely it is to be selected as the chosen suggested tutoring action. A sample interaction with the program is shown in Figure 4.2: the first column of the input represents whether the actor in the turn was a tutor or a student, the second column represents the student or tutor turn (see Tables 3.1 and 3.2), and the third column represents the facial expression of the actor (see Table 3.3) and whether the intensity of the expression was low or high.

```
Please enter a hypothetical history of interactions:

Student    Complaint                             Happy (low)
Tutor      Pump for additional information       Neutral
Student    Error-ridden answer                   Happy (low)

No matches for this history. New history is:

Tutor      Pump for additional information       Neutral
Student    Error-ridden answer                   Happy (low)

Tutoring suggestions for this history:

Tutor      Ask new question                      Neutral
Tutor      Hint                                  Neutral
Tutor      Pump for additional information       Neutral
Tutor      Rearticulate / discuss question       Neutral
Tutor      Request clarification                 Confused (high)
```

Figure 4.2. Interaction with the basic tutoring strategies module.

For example, without going into the detail of decoding the interactions, the scenario in Figure 4.2 shows an initial situation that is a sequence of 3 interactions long. When this interaction was searched for in the data from the observational study, no matches were found, so the sequence was shortened by removing the oldest interaction. Then when this shortened interaction was searched for there were 6 matches in the data, spread across 5 distinct tutoring actions. A random number would be chosen between 0 and 6, which would be used to select which tutoring suggestion would be output to be carried out by the tutoring system.

However, it was noted that often this basic version of the tutoring strategies module failed to find many matches for interaction sequences that were any more than several sequences long. Therefore it was decided that this algorithm could be significantly improved by adding a fuzzy element to the process that considers turns and expressions that are *similar* to, if not exactly the same as, the precise turns and expressions that are input to be searched for. This approach is discussed in the next section.

4.4 The fuzzy case-based tutoring strategies module

Due to the lack of matches found by the basic version of the tutoring strategies module, the aim of the fuzzy case-based tutoring strategies module was to also search the data for similar matches to the current scenario as well as exact matches. This section discusses the fuzzy version of the tutoring strategies module, which was the version later used in the ATS, Easy with Eve.

4.4.1 Why incorporate a fuzzy approach?

Fuzzy logic was pioneered by Lofti Zadeh (1973) at the University of California, Berkeley in the 1970s. Briefly defined, it is a method for representing “in a straightforward way ‘common sense’ knowledge and skills, or knowledge that is subjective, ambiguous, vague or contradictory” (Kasabov, 1998, p. 15); it is an approach that has been used in the past for case-based reasoning systems where the boundaries between categories are somewhat blurred (Pal & Shiu, 2004). Fuzzy logic has also been widely applied in ITSs, especially to areas of student modelling where student states do not always fit neatly into precise boxes (Hwang, 2003; Xu, Wang, & Su, 2002); for example, the production rules in AutoTutor discussed above are also based on a fuzzy assessment of student state (Person, Graesser, Kreuz, Pomeroy, & The

Tutoring Research Group, 2001). Finally, fuzzy logic has also been previously used in tandem with case-based reasoning as an approach to student modelling (Tsaganou, Grigoriadou, & Cavoura, 2002).

A fuzzy approach was appropriate to the current research as many of the distinctions between the tutor turns and expressions in the coding scheme can indeed be slightly blurred, as we saw in Section 3.4.2. For instance, the tutor turn *ask new question* with a low-intensity smile is almost the same as *ask new question* with a high-intensity smile. Similarly, *give neutral feedback and ask new question* is almost the same as *ask new question*. By including similar sequences in the search, it was possible for the program to make a more balanced estimate of likely tutor responses. A fuzzy approach would also make the data go further, as much more of the data is relevant to any given search than would otherwise be the case.

Therefore, the key difference between the basic tutoring strategies module and the fuzzy tutoring strategies module was that the fuzzy module searches for similar sequences to the current scenario as well as exact matches; the mechanisms by which these similar sequences are generated and weighted constitute the fuzzy rules of the fuzzy case-based tutoring strategies module. The methods used to generate and weight these similar sequences are now discussed in turn.

4.4.2 Generating similar sequences

The first aspect of the fuzzy approach is that a set of new sequences is generated that are all similar to the actual (real life) sequence of interactions between the tutor and the student. These similar sequences are generated in three different ways:

1. By varying the turns in the interactions;

2. By varying the expressions in the interactions; and

3. By varying the lengths of the interactions.

Regarding the first two of these ways of generating similar sequences, each student and tutor turn is linked to a set of other similar turns, and each combination of expression and intensity is also linked to a set of other similar combinations of expressions and intensities. So the first step is to generate new sequences by varying each of turns and expressions in the initial sequence according to the links between particular tutor and student turns and expressions. For example, given that the student turn *partial answer* (coded as 2 in the coding scheme – refer to Table 3.2) is linked to the student turns *error-ridden answer* (coded as 3) and *no answer* (coded as 4), and that the combination of expression *happy* and intensity *high* is linked to the combination of expression *happy* and intensity *low*, Figure 4.3 shows all the similar interactions that are generated for the following interaction: *partial answer, happy, high*.

	Actor	Turn	Expression	Intensity
Initial interaction:	S	2	2	2
Generated interactions:	S	3	2	2
	S	3	2	1
	S	4	2	2
	S	4	2	1
	S	2	2	1

Figure 4.3. Generated interactions that are similar to the initial interaction.

These new sets of interactions are calculated for every interaction in the initial sequence; the final set of generated sequences is the complete set of all the permutations that can be found as the individual interactions that make up the initial sequence are varied.

Regarding the third way to generate similar sequences, a new group of further similar sequences is then generated by varying the lengths of all the sequences that were generated by varying the turns and expressions. For example, if a sequence was made up of 5 different interactions, four new sequences would be generated by gradually shortening the length of the sequence by removing the oldest interaction. For example, a sequence 5 interactions long would first be shortened to 4 interactions long to create a new sequence; then it would be shortened to 3 interactions long to create a new sequence; then it would be shortened to 2 interactions long to create a new sequence; and then it would be shortened to 1 interaction long to create a new sequence. However, these new sequences are only created if they are unique, as otherwise this would often lead to the creation of many duplicate new sequences.

The final result from these three methods of generating new sequences is a (usually large) set of sequences that are all similar to some degree to the initial sequence that was input to the tutoring module. All of these new sequences can then be searched for in the data from the observational study, as well as the initial sequence. However, the less similar a sequence is to the initial sequence, the lower its weight: the method used to weight the generated sequences is now given in the following section.

4.4.3 Weighting similar sequences

In the previous section, an outline was given of how a set of sequences is generated that are similar to an initial sequence; in this section we explain the weighting system that was used to describe the closeness of the similarity between a generated sequence and the initial sequence. As we shall explain, the weights of generated sequences depend on two factors: the extent to which the interactions in the generated sequence vary from the initial sequence; and the extent to which the length of the generated sequence is shorter than the initial sequence.

Firstly, the weights of generated sequences were affected by the extent to which their individual interactions were different from those in the initial sequence. This was calculated by first determining a weighting of the similarity between each corresponding interaction in the generated sequence and the initial sequence: this in turn was determined by a weighting factor associated with each of the links between similar turns or combinations of expressions and intensities that were referred to in the previous section. Thus the total weight of a generated sequence was found by calculating the weights of all the interactions (between 0 and 1) that made up the sequence, and by multiplying those all together to arrive at a final score – which would be somewhere in between 0 and 1.

The specific weights that link similar tutor turns and similar student turns were chosen based on familiarity with the data; they were found by experimenting until the best solution was found. A better solution would have been to survey teaching faculty on what they considered to be the closeness of links between particular tutor and student and turns, but this was hampered by the fact that, including composite turns, there were a total of 140 unique tutor turns. To survey faculty on the weight of links between 140 turns with 139 other turns would have resulted in a matrix with almost 20,000 questions to fill out, which was clearly not ideal. However, devising an objective measure of determining the weights of links between similar tutor and student turns would be a good topic for future work. As it is, the weights of the links between tutor turns are all either 0.6 or 0; the weights of the links between student turns are all either 0.8 or 0.

The specific links between similar expressions and intensities all have a weight of 0.7 or 0. The criteria for determining the weights was as follows:

- If the expression was not neutral and the intensity was low, then there were two similar expressions: the same expression with a high intensity expression; and a neutral expression with a low intensity expression.

- If the expression was not neutral and the intensity was high, then there was one similar expression: the same expression with a low intensity.

- All other links between combinations of expressions and intensities were assigned a weight of 0.

Secondly, once the generated sequences were weighted based on the similarity of their interactions to the interactions in the initial sequence, the weights were also affected by the length of the generated sequences in comparison to the initial sequence. For each interaction that a generated sequence was shorter than the initial sequence, the weight of the generated sequence was reduced exponentially. All of the weighted generated sequences were then searched for in the data, along with the initial sequence; this is discussed in the following section.

4.4.4 Searching for the generated sequences

As we have seen from the previous two sections, the final result of generating the new sequences was a set of sequences of interactions that are all similar, to varying degrees, to the real current tutoring scenario; each sequence has a weighting between 0 (low) and 1 (high; however, only the real scenario has a weighting of 1).

Then each of these sequences is searched for in the data using the algorithm in Section 4.3 – except that in the fuzzy module the scores for potential suggested tutoring actions are not incremented by 1, but by the weighting of the sequence that provided the match for the potential suggested tutoring action. This way the final recommendations of the tutoring strategies module are affected both by the weights of the sequences that result in the matches as well as the frequency of those matches for each generated sequence.

```
Please enter the next interaction:

Student   Error-ridden answer               Happy (low)

Length is: 10, number of generated sequences is: 8751

There were 195 matches in the data to generated sequences


Tutoring suggestions and scores:

Tutor     Hint                              Neutral       64.02
Tutor     Rearticulate / discuss question   Neutral       40.24
Tutor     Ask question                      Neutral       31.86
Tutor     Pump for additional information   Neutral       30.36
Tutor     Pump for additional information   Happy (low)   15.06

(etc...)
```

Figure 4.4. Interaction with the fuzzy tutoring strategies module.

For example, a sample of the output of the system is given in Figure 4.4: given a sequence of 10 interactions, ending in this case with a student answering a question incorrectly whilst smiling with low intensity (coded as “S 3 2 1”), the program generated 8751 similar sequences of interactions. When these sequences were searched for in the data, 195 matches were found; these matches were used to generate the list of suggestions at the bottom of Figure 4.4.

4.4.5 Efficiency

Initially a major obstacle to the success of the case-based method was the efficiency of generating (and thus having to search the data for) up to hundreds of thousands of similar sequences for any given set of tutoring interactions. The method needed to run in real time, but could sometimes take almost a minute to generate the set of tutoring suggestions.

Therefore, to keep a lid on the number of sequences that are generated, the maximum length of a sequence was restricted to the last 10 interactions between the tutor and the student (the minimum sequence length is 1). Also, generated sequences with a weight so low as to render them insignificant were discarded. With these two safeguards, the method now runs in real time; a description of how the tutoring strategies module runs in conjunction with the rest of tutoring system Easy with Eve is given in the following chapter.

4.5 Summary

This chapter has presented a case-based reasoning tutoring strategies approach that uses the data from the study of human tutors to recommend tutoring actions based on the sequences of interactions between tutors and students, taking turns, expressions and intensities into account. Two versions for the tutoring strategies module have been presented: the basic version, that only searched the data for exact matches with the current tutoring scenario; and the fuzzy version, that also searches the data for a set of similar sequences that are generated based on weighted links between turns, expressions and intensities. The fuzzy method runs in real time.

The fuzzy tutoring strategies module that this chapter has presented is used in the ATS for maths, Easy with Eve, that is presented in the following chapter. As an important part of the implementation of Easy with Eve, the ways in which the recommendations of the tutoring strategies module relate to the actual content that is delivered by the ATS will be discussed in the following chapter.

The effectiveness of the tutoring strategies module was assessed at the same time as the ATS, Easy with Eve, was assessed; the evaluation study of Easy with Eve is presented in Chapter 6.

Chapter 5

Design and implementation of Easy with Eve

This chapter presents the design and implementation of an affect sensitive Intelligent Tutoring System (ITS), Easy with Eve.

To briefly recap, we saw back in Chapter 1 that the overall aim of this research is to develop an affect-sensitive ITS that features an emotionally expressive animated pedagogical agent. It was seen that this had never been done before, and so Easy with Eve is the first tutoring system of its type. However, one fact that we noted in Chapter 2 was that little research had been done to address *how* tutors should adapt to student affective states, and this fact led to the observational study of human tutors that was discussed in Chapter 3. Then Chapter 4 presented a fuzzy case-based tutoring strategies approach that addresses the issue of ITS adaptation to student emotion by mining the data that was collected from the observational study of human tutors.

Now in this chapter we present the design and implementation of Easy with Eve, an Affective Tutoring System (ATS) that was developed using Borland C++ Builder as a test-bed for the fuzzy case-based tutoring strategies module. The system adapts to the student via an emotionally expressive animated pedagogical agent, named Eve. We begin this chapter by briefly presenting the overall architecture of the system, and then we discuss each of the components of the system in turn.

5.1 Architecture

As Figure 5.1 below shows, the overall architecture of Easy with Eve is made up of four main components: a facial expression analysis module; a student model, which is made up of both a set of proficiency scores for several mathematics concepts and a coded representation of the interaction history; a tutoring strategies module; and a set of tutoring actions in the form of agent animations and tens frame and counter animations. As Figure 5.1 shows, both the actions and the emotions of the student are taken into account by the tutoring system. The architecture of Easy with Eve has been designed so that it is extensible; this is discussed in Appendix E.

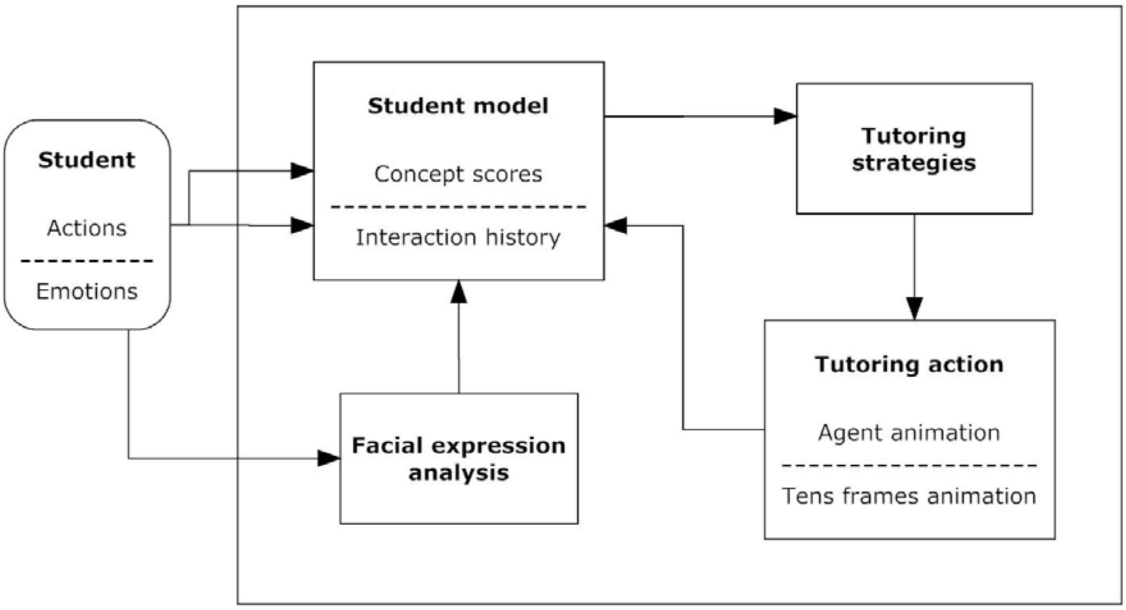


Figure 5.1. Architecture of Easy with Eve.

Next we explain the domain of the ITS and the different levels in the tutoring exercise, and then we discuss the facial expression recognition system that is used by the ITS for emotion detection. We then explain in turn how the Eve and the tens frames were animated, the way that the student model works, and how the student model interfaces with the tutoring strategies module. We finish the chapter by giving an example of an order of events when Easy with Eve runs, and by reporting the results of initial pilot tests.

5.2 Domain

Easy with Eve is designed to help primary school students about the age of 8 with exactly the same New Zealand Numeracy Project (2003) exercise that was used in the observational study of human tutors that was presented in Chapter 3. As described in Section 3.2.3, the Numeracy Project exercise involves learning to add two numbers together by subtracting from the second addend to make the first addend up to the next ten. Two examples of the tens frames, counters and screen that are used in Easy with Eve are shown below in Figures 5.2 and 5.3.

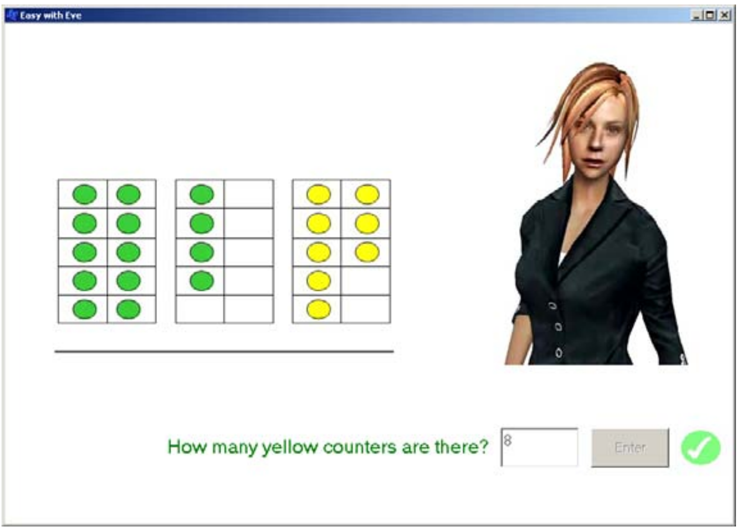


Figure 5.2. Screenshot of the interface of Easy with Eve at Level 1.

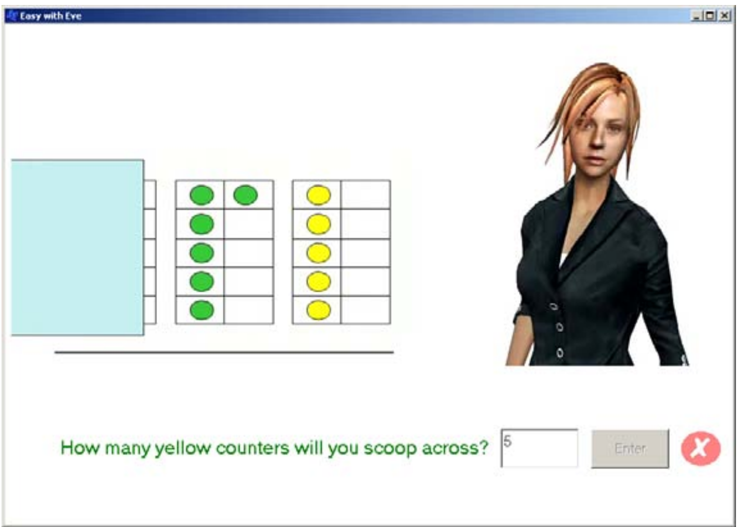


Figure 5.3. Screenshot of the interface of Easy with Eve at Level 2, showing the screen as it covers the counters.

5.2.1 Levels in Easy with Eve

The addition exercise in Easy with Eve is divided into 4 levels of difficulty; as students succeed, they progress through the levels, but if they struggle with a level over several problems they go back to an easier level. The four levels are now described in turn.

1. In a problem at Level 1, the student can see the counters and the tens frames; the first addend is between 10 and 20 and the second is between 1 and 9. The student is asked a series of questions that leads them to rearrange the counters so that they can easily identify the sum. A screenshot of the interface at Level 1 is shown above in Figure 5.2.

2. In a problem at Level 2, the tens frames and counters are now covered by a piece of card during the questions. The card is briefly taken away each time the student answers a question to give them a chance to re-visualise the counters. A screenshot of the interface at Level 2 is shown above in Figure 5.3.

3. In a problem at Level 3, the tens frames and counters are removed altogether; now the first addend is between 20 and 89, but the second addend is still between 1 and 9. Similarly to previous levels, the student is still asked a series of questions that leads them to rearrange the addends so that the sum is made obvious. For example, the problem 33 + 8 would become 33 + 7 (=40) + 1, which would become 40 + 1, which is obviously 41.

4. In a problem at Level 4 the tens frames and counters are still removed, and now both addends are double-digits. This time there are no leading questions asked, and the student is required to apply the part-whole addition principle by him/herself to add the two numbers together.

5.2.2 Moving between levels

Students move between levels according to a simple scoring system. As students begin each level they are assigned a score of 0 points, and each time that they complete a problem perfectly (remembering that in Levels 1 to 3 each problem is made up of a series of questions), their score increases by 1 point. Whenever a student's score reaches 1, the student will progress to the next level – unless the student is already at Level 4, in which case the student will remain on Level 4. However, whenever a student fails to complete a problem perfectly, the score will decrease by 1 point. Similarly, whenever a student's score reaches -2, the student will go back to the previous level – unless the student is already at Level 1, in which case the score remains at -2.

5.3 Facial expression analysis

As a student interacts with Easy with Eve, Eve is able to detect several student emotions. This is achieved by using a real time facial expression analysis system that has been developed in-house at Massey University (Sarrafzadeh, Fan, Dadgostar, Alexander, & Messom, 2004; Fan, Sarrafzadeh, Dadgostar, & Gholamhosseini, 2005), as was discussed in Section 2.2.1. The emotion classification is achieved by using support vector machines, and is able to detect 8 different facial expressions: neutral, smiling, laughing, surprised, angry, fearful, sad and disgusted. A screenshot of the facial expression recognition system is shown in Figure 5.4.

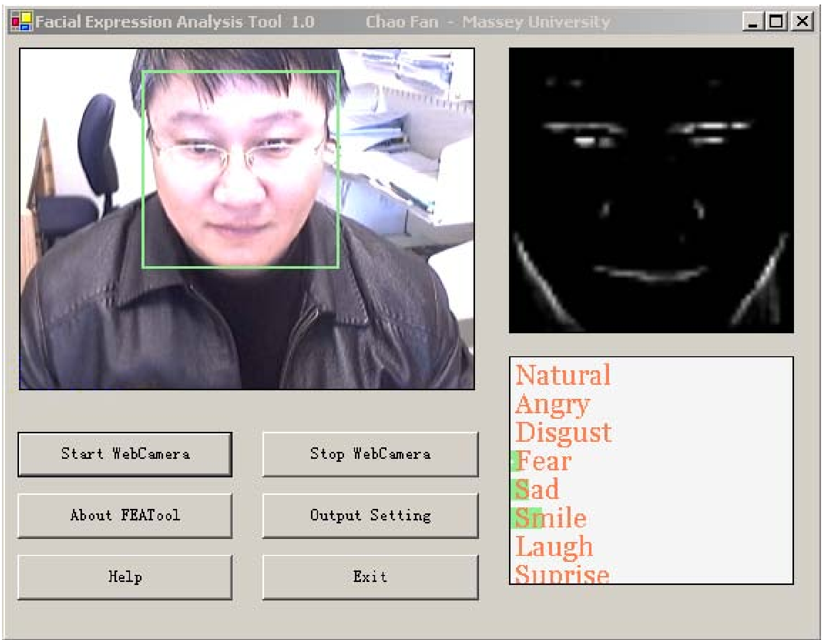


Figure 5.4. Screenshot from the facial expression recognition system.

The module uses a facial feature extraction algorithm that is not only able to extract all important facial features, but is also fast enough to work in real time. Unlike other facial expression analysis systems, facial information is automatically detected without the need to manually mark facial features. Initial tests on artificial data sets (using computer generated faces) have yielded a successful recognition rate of approximately 90% (Fan et. al, 2005).

The facial expression analysis system runs in the background while the student interacts with Easy with Eve. A file is updated every time that the system detects a facial expression, and this file is checked by the tutoring system for changes.

5.4 Animation

Easy with Eve features two different types of animation: firstly, Eve is an animated pedagogical agent; and secondly, the tens frames and counters used in the problems are animated to help the student understand the part-whole addition concept (this was explained in Section 3.2.3). We will now briefly discuss these two different types of animations.

5.4.1 Agent animation

The affect-sensitive pedagogical agent Eve has been animated so that she is able to display a range of emotions through facial expressions, as well as deliver spoken teaching content. Screenshots of Eve are given above in Figures 5.2 and 5.3.

Approach. To do this, all possible tutoring actions of Eve were first animated, and saved as videos that could be imported into the ATS – each of these videos had identical frames at the beginning and end so that they could be played seamlessly back to back.

Selecting which tutoring actions to animate. Ideally, it would have been good to separately animate every possible combination of tutor turns and expressions (according to coding scheme in Section 3.2.4) for several different problems – and in fact it would have been necessary to animate several different videos for each one of those combinations to allow for slight variations in the dialogue (to reduce the potential for the appearance of repetition). However, that would have led to animating well over 5000 separate videos one by one, which would have taken an unfeasibly long time.

As a solution to this problem, choosing which tutoring actions that Eve would be able to perform was achieved by consulting the data about human tutoring actions from the observational study of human tutors that was presented in Chapter 3. The steps that were taken to choose Eve's possible tutoring actions are now given.

- Firstly, any tutor turn that occurred with a frequency of 0 in the data was immediately discounted (refer to Table 3.1), as clearly it was impossible for the case-based program from Chapter 4 to recommend a tutor turn that never happened as a suggested tutoring action.

- Then the tutor turns were grouped into categories based on similarities; for example, tutor turns 5, 6, 7 and 8 (refer to Table 3.1) were all related to asking a question, so they were all grouped together as a category.

- As a general rule, every one of these tutor turn categories that the human tutors performed with an overall frequency of 1% or greater was chosen to be animated. The only exception to this was the tutor turn *other*, in a category by itself, which only occurred in the observational study when the human tutor talked to the student about a topic unrelated to the tutoring exercise (such as the weather, or what they were going to do in the school holidays). This category was not chosen to be animated.

- Then each of these categories of tutor turns was animated for different facial expressions and expression intensities, again according to the data from the observational study. Each category of tutor turns that occurred with a frequency of 2 occurrences or greater for a particular expression and intensity was animated separately for that expression and intensity. However, if the sum of the low intensity and high intensity occurrences was less than 5 for a particular expression, then only one set of low intensity videos was animated for that expression, which in that case would also double as high intensity videos.

- Then each of this new set of videos for turn categories, expressions and intensities was animated differently for each different problem. There were four different problems at each level, although the same problems were used for Levels 1 and 2 (the difference being that a screen covered the counters in Level 2). However, non-problem-specific turns such as *positive feedback* were not required to be re-animated for each different problem.

- Finally, each of the combinations of tutor turn categories, expressions, intensities and problems were animated several times over, each with a slightly differently scripted dialogue. This was to allow for variation so that Eve did not appear repetitive.

- Several other videos were also animated, such as an explanation of each new level; in these cases only one version of the video was animated. In addition, several *idle* videos were also animated, which were to be played whenever the student was supposed to be answering a question. These would give the impression that Eve was patiently waiting, which would in turn increase her believability.

In total, almost 1000 different tutoring videos of Eve were animated; in other words, Eve had almost 1000 different possible tutoring actions. These tutoring actions included: giving positive or neutral feedback, asking questions, discussing problems or solutions, giving hints, or answering her own questions if required.

Animating the videos. The videos for the tutoring actions were animated using a software package called iClone, which is an animation tool developed by Reallusion (http://www.reallusion.com/iClone/, 2007). The software comes with several built in characters, as well as a range of built in animations and facial expressions that were also useful in animating the videos. The software also includes a motion editor that allows animations to be tailored, which was used to animate many of the videos. Each of the videos was given an identical beginning and end frame; this was to ensure that all of the videos could be played seamlessly back to back.

Eve’s dialogue in the videos was input as .wav files that had been individually pre-recorded for each combination of tutor turn and problem. To record the .wav files, a script was written with the dialogue for all of the tutor actions that was passed on to an assistant, who recorded herself reading from the script using standard Windows tools. This approach was chosen above using an artificial text-to-speech voice because a real voice sounded much more realistic and appealing; this was important since one of the key aims of the agent was to be as believable as possible to maximize the persona effect (van Mulken, André, & Muller, 1998; refer to Section 2.2.2).

When these sound clips were imported into iClone, the software automatically matched Eve’s lip movements to the sound clips to give the appearance that she was speaking. The initial videos that were output were in .avi format, but after some experimentation, it was decided to convert these into Flash movies. This was to avoid any issues relating to codecs when running the tutoring system on different computers.

5.4.2 Tens frame and counters animation

Animating the tens frames and counters was very straightforward compared to animating Eve. All that was required was to move a number of counters from the right-hand side tens frames across to the middle tens frame to make the middle tens frame up to the next ten, and back again, all depending on the particular stage of the problem that the student was at. Also, in Level 2 a screen was animated moving in from the left-hand side so that all of the counters were covered; then when the student answered a question the screen was taken away again. Videos for four different problems were animated – not enough for a complete version of the tutoring system, but quite enough for use as a test bed for the effectiveness of the agent and the tutoring strategies module.

Similarly to the agent videos, all of the counter videos were animated during the implementation phase as stand alone videos that were later imported into Easy with Eve. Again, this was chosen over real time animation of the counters because it seemed an easier, though no less effective option.

The counters were animated using Microsoft PowerPoint, which were then converted into .avi files using a piece of software called PresenterSoft PowerVideoMaker. Next, as with the agent videos, the counter videos were converted into Flash movies to avoid the potential for codec issues on different computers.

5.5 Student model

As shown previously in Figure 5.1, the student model in Easy with Eve is divided into two sections: the first is a set of proficiency scores for several mathematics concepts; the second is a representation of the history of interactions between the tutoring system and the student, to be used in conjunction with the case-based tutoring strategies module. We now discuss these two sections in turn.

5.5.1 Concept scores

The first section of the student model is a set of seven proficiency scores for concepts related to part-whole addition that were identified through personal communication with one of the researchers for the New Zealand Numeracy Project. The seven concepts are: counting the counters on the tens frame; making the middle tens frame up to the next ten using counters from the right tens frame; imagining the number of counters when they are covered by the card in Level 2; knowing basic facts; understanding what the correct addends are; and being able to carry one. When a student struggles with a concept in a problem, their proficiency score is adjusted accordingly, but this can be redeemed if they demonstrate the appropriate competence in subsequent problems. At the end of every session, the data in the student model are summarised in a report that could be used by teachers or parents to track the student’s progress.

The concept scores section of the student model does not play a role in determining the type of the agent’s tutoring actions, as these are generated using the case-based tutoring strategies module, as will be discussed in the following section. However, the concept scores can affect the content of Eve’s tutoring actions as the problems become more difficult and the scope for different errors increases. In other words, the concept scores do not determine the category of tutor turn that Eve will carry out, but they do affect *how* she goes about it once the student reaches Level 4. This will be touched upon in

Section 5.6.3, which addresses how Eve knows what to do next for any given tutoring scenario.

5.5.2 Maintaining a history of interactions

The second part of the student model is a representation of the history of interactions between the tutoring system and the student, to be used in conjunction with the case-based tutoring strategies module. As we saw in Chapter 4, the input required for the case-based reasoning tutoring strategies module is a sequence of interactions in the same format as the data from the observational study in Chapter 3; therefore the tutoring system must maintain a history of the interactions between Eve and the student that is coded according to the coding scheme presented in Section 3.2.4. In this section we discuss how student and tutor behaviour respectively can be maintained in the interaction history.

The student actions that need to be encoded include their turns, expressions and intensities. Firstly, student turns are coded according to their answers to questions, which are either no response, right, or wrong. Secondly, student expressions and intensities can be found by referencing the data file generated by the facial expression recognition system, which is frequently polled to check for changes. If a student changes his/her expression over the course of answering a question, then the most prevalent expression is selected to be added to the interaction history.

Tutor actions, on the other hand, already come packaged in the correct format for addition to the interaction history. This is because almost all tutor behaviour comes as a result of output from the case-based tutoring strategies module, which is inherently in the same format as the data from the observational study. How this tutor behaviour comes to be generated in the first place is now addressed in the following section.

5.6 Tutoring strategies module

So far in this chapter we have discussed the domain of the system, how it detects emotions through facial expression analysis, how the system was animated, and how the student model works. But in this section we discuss the engine-room of Easy with Eve that links all of these other parts together: the interface between the components of the tutoring system and the case-based tutoring strategies module from Chapter 4. This section addresses *how* Eve knows what to do next for any given scenario: first we describe how Eve generates tutoring actions from the tutoring strategies module; and then we describe how she puts these chosen tutoring actions into practice.

5.6.1 Choosing a tutoring action using the tutoring strategies module

The interaction history that was discussed in the previous section is used as an input to the case-based tutoring strategies module that was presented in Chapter 4. So, as soon as the student answers a question the student turn, expression and intensity is added to the interaction history, then the updated interaction history is passed to the tutoring strategies module.

Ignoring here its internal workings (that were described in detail in Chapter 4), the output of the tutoring strategies module is a set of suggested tutoring actions, each with weighted scores; Eve randomly selects one of the recommendations to follow according to the weights of the recommendations. Thus the recommendations with the higher scores are more likely to be selected, but it is impossible to know for sure which tutoring action will be selected to be performed. The reason for this approach is to introduce an element of slight unpredictability – and hence believability – to Eve, whilst still carrying out the most common tutoring responses from the observational study of human tutors most of the time. Generally, this one selected tutoring action will be what Eve carries out in response to the student's current behaviour.

However, sometimes the output of the tutoring strategies module needs to be regulated in order to fit within the limitations of a tutoring system, as opposed to the relative freedom intrinsic to the human-to-human interactions in the observational study. Following are several examples of the ways that the output needs to be regulated.

- The tutoring action must finish with the tutor asking the student the next question, or else it will be unclear for the student what he/she is meant to do next.

- If a student has completed a problem, and satisfied the conditions for a new level, then a description of the new level must be included in the tutoring action.

- If a student answers incorrectly three times in a row, then Eve must supply the answer herself to avoid a frustrating cycle of incorrect answers for student.

In these cases, the output from the tutoring strategies module is adjusted accordingly.

5.6.2 Carrying out the tutoring action

Once a tutoring action has been settled upon from the tutoring strategies module, it is put into action by mapping the tutor turn, expression and intensity to the appropriate agent videos, and then playing those videos. This mapping is done by using the same categories that were used to determine which agent videos to animate in Section 5.4.1; the expression and intensity of the tutor action are other factors in choosing which video to play. Also, where there is more than one version of the script for a video (to decrease the chance of repetition, as discussed in Section 5.4.1), one of the available videos is randomly selected. Once played, this video will become unavailable until the all variations in the script for this particular scenario have been exhausted. If the chosen tutoring action is a composite turn – that is, it is made up of a series of several tutor turns in a row – then several agent videos will be played back to back.

Also, as mentioned in Section 5.5.1, the concept scores can have an impact on the content of some of the tutoring actions in Level 4, and thus on which particular agent videos are played. There are several concepts in the student model that could be relevant to any given problem at Level 4; if the tutor turn *hint* is selected, then the content of the hint will be based upon which one of three concepts that the student has just misunderstood.

Finally, depending on the level of the problem and which tutor turn was selected, one of the counter videos may be played concurrently to tie in with the agent video. The counter videos and the agent videos are synchronised so that Eve can refer to the animation of the tens frames and counters as part of her dialogue.

5.7 An order of events

With the previous sections as a survey of the implementation of Easy with Eve, it may now be useful as a summary to review the order of events that occur within the system as it interacts with the student. As shown in Figure 5.5, whenever the student responds to a question, the following is the order of events:

- the concept scores are updated according to the student's response;
- the system determines the most prevalent facial expression and intensity of the student during the period that they took to answer the question;
- the interaction history is updated by classifying the student's response to the question and by using the information about the student's expression and intensity;
- the interaction history is input to the tutoring strategies module, which generates a set of weighted recommendations for Eve's next action;
- based on their weights, a tutoring action is semi-randomly chosen from the set of recommendations, and then regulated as required;
- the tutoring action is mapped to one or more agent videos, and possibly a counter video;
- these videos are played, by which Eve carries out the tutoring action;
- the history is updated with what Eve has just done;
- Eve appears to wait patiently for the next student response.

A sample interaction between a student and the tutoring system is given in Appendix F.

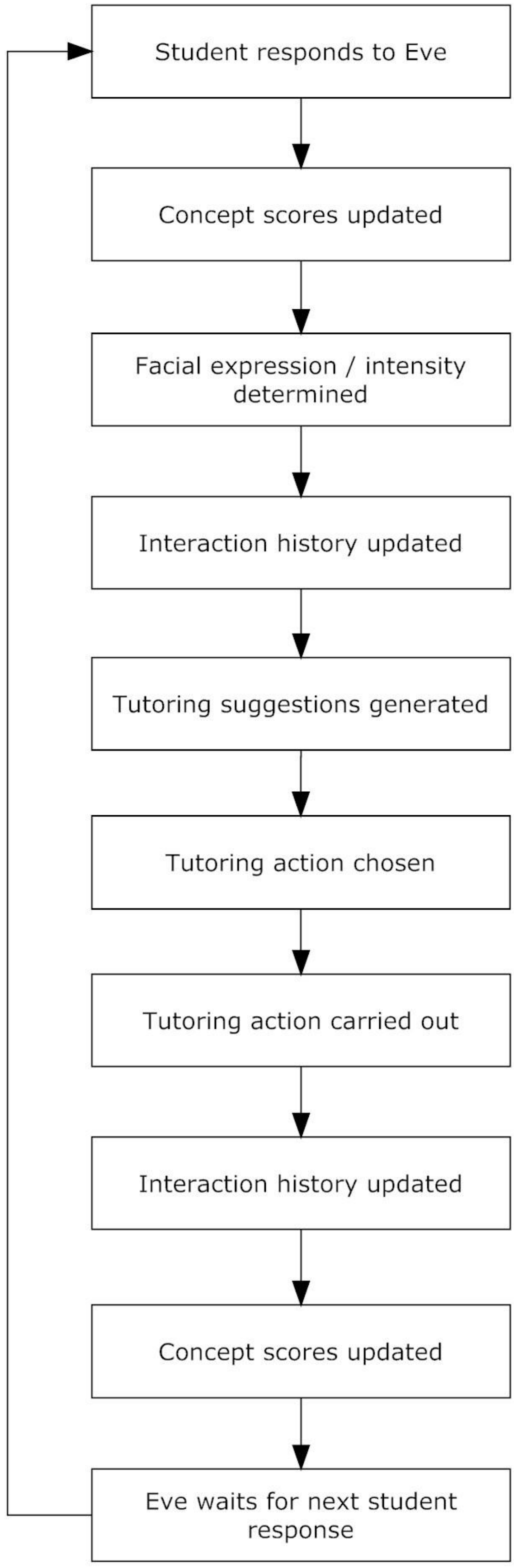


Figure 5.5. Order of events in Easy with Eve.

5.8 Pilot tests

Before going ahead with formal testing, Easy with Eve was first pilot tested on two different eight year old students over two separate tutoring sessions (one student per session). Each student was first greeted by Eve in an introductory video, who gave an brief overview of what the tutoring session would be like. Eve then explained a short 5 to 7 minute pre-test that the students were first required to take – the pre-test was made up of 15 questions that were taken from all of the four different levels in Easy with Eve. Eve disappeared as the pre-test began, and once all the questions were answered she reappeared to explain that the main tutoring session was about to begin. Students proceeded to use the tutoring system for a period of about 20 minutes, after which Eve explained that time was nearly up, and that the students would be finished after they completed a post-test; the post-test was exactly the same as the pre-test, except with different addends.

Some very useful feedback was gathered from the two tutoring sessions. Although both students were positive in their comments, and although both students improved their test scores from pre- to post-test, several limitations were noted. Most notably, students struggled with the language in the pre- and post-tests, as Eve was not present to read the questions like she was during the actual tutoring session. Also, it was seen that a slight change to how one of the questions is asked could significantly help the student understand the part-whole addition concept. Both of these issues were subsequently addressed: firstly by simplifying the language in the pre- and post-tests, and secondly by rephrasing the questions in the main system with the part-whole addition concept in mind.

5.9 Summary

This chapter has presented the implementation of Easy with Eve, including the domain and levels within the system, the way that emotions are detected by facial expression recognition, the animation of the agent and counter videos, and the design of the student model. Also, this chapter has described how the tutoring strategies module interfaces between all of these components, thus continually guiding Eve's next move. The method and results of several pilot tests have also been briefly described, which proved to be useful in preparation for the formal testing of Easy with Eve in schools. The methodology and results of this formal testing are given in the following chapter.

Chapter 6

Study of the effectiveness of Easy with Eve

This chapter describes the study that was carried out to test the effectiveness of the Affective Tutoring System (ATS), Easy with Eve. We begin by presenting the methodology of the study, then we give the results, and conclude with an overall summary. The results of the study are discussed in Chapter 7.

6.1 Methodology

Easy with Eve was user-tested at two local primary schools in Auckland to investigate its impact on student performance, motivation and perceptions of the tutoring experience. In this section we discuss the goals of the study, the participants in the study, the measures that were used, and finally a break down of the procedure that was followed to carry out the study.

6.1.1 Goals

The study carried out to test Easy with Eve was designed with several key goals in mind:

- To find out whether or not an ITS that adapts to student emotions as well as student answers to questions would significantly increase student learning and improve student performance.

- To find out whether or not an ITS that adapts to student emotions as well as student answers to questions would affect student perceptions of the tutor and the tutoring experience.

- To find out whether or not the presence of an animated tutoring character would increase student learning and improve student performance.

- To find out whether or not the presence of an animated tutoring character would affect student perceptions of the tutor and the tutoring experience; as discussed in Section 2.4.2, this phenomena is known as the *persona effect* (van Mulken, André, & Muller, 1998).

- To find out whether or not the ability to adapt to student emotion makes any difference to the persona effect of an animated agent.

As can be seen from this list of goals, the two independent variables that this study aimed to investigate were firstly whether or not the tutoring system would adapt to student emotion (i.e. student facial expressions) as well as student answers, and secondly whether or not the animated tutoring character (Eve) was present. The first independent variable, whether or not the system detected student emotion, was of particular interest as existing ITSs adapt only to student answers.

Therefore, a 2x2 between-subjects experimental design was chosen for this study, with four different experimental groups, as shown in Table 6.1. There were four different versions of the tutoring system, to match the four different experimental groups:

– in Group 1, the animated agent was present and facial expressions were detected (as a measure of student emotion to be used by the tutoring strategies module);

- in Group 2, the animated agent was not present (text-based feedback was used) but facial expressions were detected;
- in Group 3, the animated agent was present but facial expressions were not detected; and
- in Group 4, neither the animated agent was present or facial expressions detected.

Table 6.1. The four different groups in the 2x2 experimental design.

	With the animated agent, Eve	Without the animated agent; text-based feedback
Facial expressions detected	Group 1	Group 2
Facial expressions not detected	Group 3	Group 4

6.1.2 Participants

In total, there were 62 participants in the study, 30 males and 32 females. The participants were all 8-9 year old students at one of two local schools in Auckland; the students were evenly and randomly divided between the four experimental groups shown in Table 6.1. However, data from 3 of the students was unusable, leaving a total of 59 participants; the reasons for this will be discussed in Section 6.2.

Ethical approval was sought and obtained from the Massey University Human Ethics Committee before the participants were contacted and the study carried out. A copy of the ethics application and the accompanying consent forms and information sheets are given in Appendix G.

The participants were not told which experimental group they were going to be in. At the beginning of the testing each participant was told that the web-cam on the monitor (used by the facial expression recognition system) only worked half of the time – but the participants were not told whether their facial expressions were being detected or not. This was so that later on the participants could be asked as a part of a questionnaire to guess whether the camera was working or not, to see how much of a difference the facial expression detection made to student perceptions about the tutoring system. This questionnaire is discussed as one of the measures in the following section.

6.1.3 Measures

There were two measures chosen to investigate the effects of the two variables (whether or not the system adapted to student emotion and whether or not Eve was present): the first was a measure of student performance using a pre-test and a post-test, and the second was a measure of student opinions using a questionnaire. These will now be discussed in turn.

Difference between pre- and post-tests. The aim of the pre- and post-tests was to determine whether or not students improved their counting and addition through using Easy with Eve, and to generate a measure of exactly how much students improved (or regressed). Also, by comparing the improvements between the different experimental groups it would be possible to see if some groups improved significantly more or less than others.

Both the pre- and post-tests were given to the student on the computer with which they used Easy with Eve. In fact, both the tests were integrated into the tutoring system itself so that the entire pre-test, tutoring session, and post-test sequence was introduced and explained by the tutoring system; the system also explained the transitions between the pre- and post-tests and the tutoring system proper. In experimental groups where Eve was present, this was all explained to the student by Eve; the text-feedback groups used text.

The content of the tests was based on the same problems that the students completed when using Easy with Eve. In particular, the tests were made up of one problem from Level 1 (refer to Section 5.1.1 for a description of levels in Easy with Eve), one problem from Level 2, one problem from Level 3 and two problems from Level 4; there were 15 questions in total (the problems from Levels 1 to 3 were made up of more than one question), so they were scored out of a maximum of 15. The questions in the pre- and post-tests were more or less identical in style and difficulty, the only difference being changing the addends so that the answers were not exactly the same for both tests. The tests were designed to take about 5 to 7 minutes each.

Questionnaire. After completing the post-test, the students were given a brief questionnaire to fill out to evaluate their experience with Easy with Eve; the participants were required to tick one box per question on a five-point Likert scale, as shown in Figure 6.1 (from 5 at the left hand side to 1 at the right hand side). To make this as clear as possible for the students, a series of happy and sad faces were used; a similar rationale for the use of happy and sad faces in questionnaires designed for children was found in Burkitt and Barnett (2006). As listed below, the questionnaire had 5 questions:

Q1 Did you enjoy using Easy with Eve today?

Q2 Would you like to use Easy with Eve again?

Q3 How much do you think you learned today?

Q4 Remember how the camera on the screen only works half of the time – do you think it was working today?

Q5 Did Eve look real?

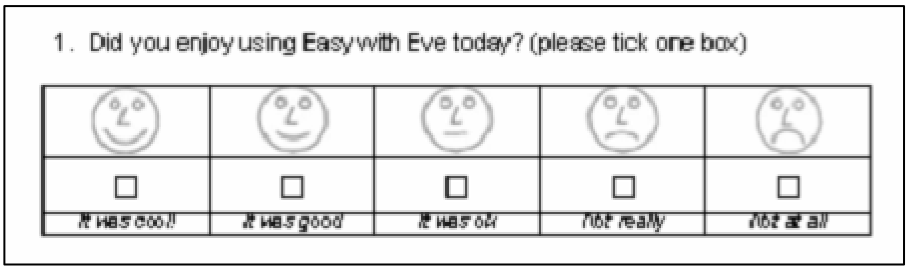


Figure 6.1. Example of the Likert scale used in the questionnaire.

Students in the two text-based experimental groups were given a questionnaire that did not include Question 5, as Eve was not present in their version of the tutoring system. Copies of both versions of the questionnaire are given in Appendix B.

The aim of the questions was to find out how students perceived their interaction with Easy with Eve: how much students enjoyed using the system and were motivated to use it again; how much students felt that they learned by using the system; whether or not students could tell whether or not their facial expressions were being detected; and whether Eve appeared real to those in the animated agent groups (Groups 1 and 3). In particular, the responses to the questionnaire were designed to give an insight into two distinct issues that are now discussed: whether or not Eve created a persona effect; and whether or not detecting facial expressions made any difference to how students perceived their interaction with the system and Eve.

- Firstly, if the persona effect were to hold for Easy with Eve then students should view the animated agent versions of the system as more enjoyable (Q1), more desirable to use again (Q2) and more helpful for learning (Q3) than the text-only versions, as well as conveying a believable persona to students (Q5).

- Secondly, for the questionnaire to show that facial expression detection enhances student perceptions of the tutoring system, students should rate the facial expressions detected versions of the system to be more enjoyable (Q1), more desirable to use again (Q2) and more helpful for learning (Q3) than the facial expressions not detected versions, as well as enhancing Eve's believability (Q5). It would also be expected that the appropriate responses of the tutoring system meant that students could tell that their facial expressions were being detected (Q4). Also, if the responses from the animated agent and facial expressions detected group (Group 3) were higher than the responses from the animated agent but facial expressions not detected group (Group 4) across the questions, then this would supply evidence that detecting facial expressions can strengthen the potential for a persona effect.

6.1.4 Procedure

This section details the set up and procedure that was followed to carry out the testing with the participants.

The testing was carried out on site at two local primary schools in Auckland, in a meeting room and a remedial teaching room respectively. As Figures 6.2 and 6.3 show, there were four computers set up to be used in the testing, with each computer running a different version of Easy with Eve for the four experimental groups depicted in Table 6.1. The computers were arranged so that there would be a minimum of distraction from the other participants.

Figure 6.2. Experimental set up used in the study at the first school.

Figure 6.3. Experimental set up used in the study at the second school.

Each of the four computers was equipped with a web cam and a set of headphones, regardless of whether or not Eve was present or facial expressions were being detected on that version of the tutoring system. The aim of this was to minimise the differences between the experimental groups, and also to make it possible to test whether or not students could tell whether or not their facial expressions were being detected based solely on the tutoring system's responses.

The procedure that was followed with the participants is now listed.

1. Up to four participants at a time were welcomed into the meeting/remedial teaching room, and given a brief overview of the pre-test, tutoring system and post-test.

2. Participants began the pre-test. Detailed instructions for this were given by the tutoring system itself; the test was designed to take approximately 5 to 7 minutes.

3. As participants completed the pre-test, the tutoring system introduced the next phase of the testing, which was the session with the tutoring system proper. This was designed to take approximately 18 to 20 minutes; participants began their last problem immediately after the first time that they completed a problem after 17 minutes had elapsed.

4. As participants interacted with the system during the tutoring session (i.e. after they completed the pre-test), the participants were also videoed from front-on, and a log of the interactions was kept by the tutoring system. The reason for this was to make it easier to review the tutoring sessions later on.

5. When participants had completed the tutoring session, they began the post-test. Detailed instructions for this were given by the tutoring system itself (refer to Appendix C for a transcript of these instructions); the test was designed to take approximately 5 to 7 minutes. The results of the pre- and post tests were written to a data file.

6. After the post-test was completed, the system exited; the participants were then given the questionnaire to fill out. This generally took no more than one minute.

7. After the questionnaire was completed, the participants were let back to class.

6.2 Results from the study

The results from the study fall into three categories: the overall improvement in test scores from pre- to post-tests across all groups; the effects of the presence of Eve and facial expression analysis on the improvement in test scores; and the data from the questionnaires that the participants completed. This section will present each of these three groups of results in turn; the raw data of these results is presented in Appendix D.

Firstly though, it should be noted that valid data was gathered from 59 participants out of the possible 62. The data that was invalid for the other 3 students was not useable for the following reasons.

- For one participant the tutoring system did not function appropriately.

- A second participant exited the tutoring system half way through his tutoring session – there was not enough time to begin the tutoring session over.

- The other participant struggled with the tutoring exercise to the extent that he would have needed an extra 20 minutes to complete the pre- and post-tests, which was not possible due to a tight testing schedule.

6.2.1 Overall effectiveness of Easy with Eve

The data showed a substantial overall increase in test scores between pre-tests and post-tests. As shown in Figure 6.4, a 2 (presence of agent) x2 (facial expression input) x2 (pre-/post-test) mixed design ANOVA revealed a main effect of test. Post-test scores (*M*= 9.34, *SD*= 2.23) were significantly greater than pre-test scores (*M*= 8.42, *SD*=

2.28), *F* (1, 55) = 7.48, *p* = 0.01. No other main effects or interactions were found for this analysis.

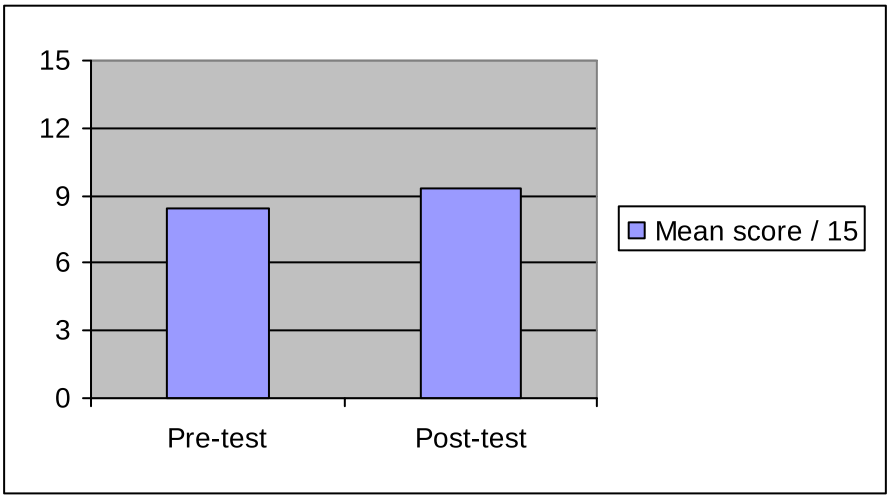


Figure 6.4. Graph showing the increase from pre-test mean to post-test mean.

6.2.2 Performance differences across experimental groups

The data were also analysed to determine if there were significant differences in participant performance, as measured by the mean increases between pre-test and post-test, caused by the two variables in the experimental groups: whether or not facial expressions were detected, and whether or not the animated agent was present.

Presence of animated agent. As shown in Figure 6.5, a 2 (presence of agent) x2 (facial expression input) ANOVA found no main effect of presence of agent, (Present *M*= 0.75, *SD*= 2.74 versus Absent *M*= 1.06, *SD*= 2.25), *F* (1, 55) < 1, n.s.

Facial expressions detected. Also, as shown in Figure 6.6, no main effect of facial expression input was detected, (Detected *M*= 1.10, *SD*= 2.76 versus Undetected *M*= 0.71, *SD*= 2.16), *F* (1, 55) < 1, n.s.

No interaction between the two variables. Finally, there was no evidence of an interaction between the presence of the animated agent and the detection of facial expressions, *F* (1, 55) < 1, n.s.

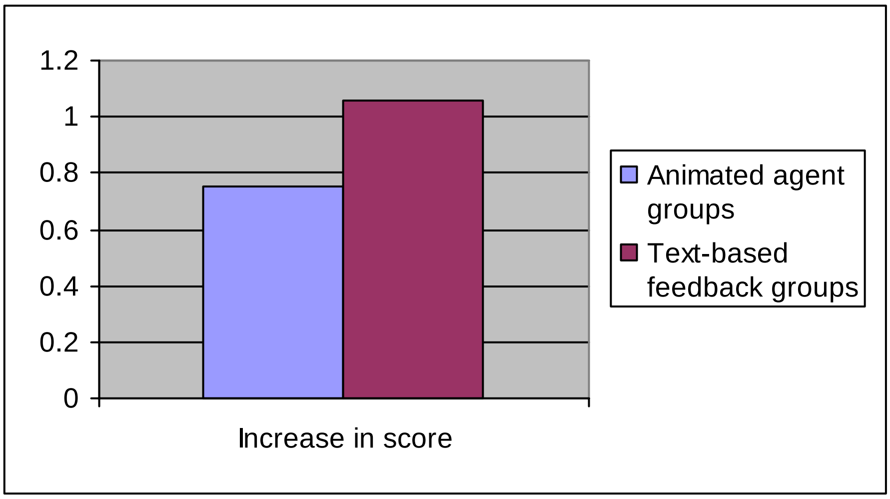


Figure 6.5. Graph showing the mean increase in score (post-test score minus pre-test score) for the agent present groups vs. the mean increase for the text-based feedback groups.

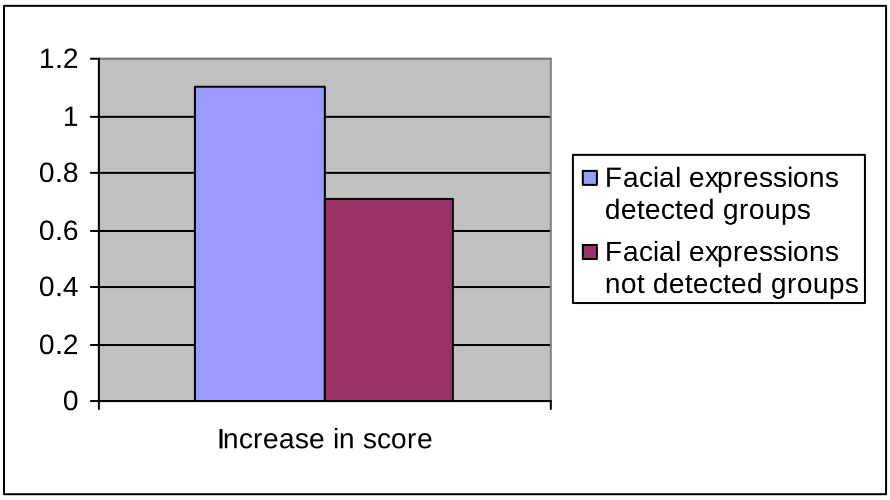


Figure 6.6. Graph showing the mean increase in score for the facial expressions detected groups vs. the mean increase for the facial expressions not detected groups.

6.2.3 Questionnaire responses

As was explained in Section 6.1.3, the questionnaire that was given to the participants to fill out had either four or five questions, depending on whether or not the participants were in one of the experimental groups where Eve was present. As with the pre- and post-test scores, the questionnaire data were analysed to determine if there were significant differences caused by the two variables in the experimental groups: whether or not the animated agent was present, and whether or not facial expressions were detected. The results from each of these questions will be presented in turn: a series of a 2 (presence of agent) x2 (facial expression input) ANOVAs were carried out to analyse data for the first four questions, and a *t*-test was carried out to analyse data for the fifth question. The data from the first four questions are shown in Table 6.2; significant and marginally significant *p* values are in bold.

Table 6.2. Main effect and interaction results from the first four questions showing means, *F* scores and *p* values for the three possible effects: agent present, facial expressions detected, and agent present and facial expressions detected. Significant and marginally significant *p* values are in bold.

Question	**Effects**	**Mean (SD) for present, absent conditions**	***F* score**	***p* value**
1	Agent present	4.36 (0.83), 4.74 (0.45)	5.84	**0.02**
	Expressions detected	4.68 (0.60), 4.43, (0.74)	2.62	0.11
	Agent present x Expressions detected		2.30	0.14
2	Agent present	4.29 (0.85), 4.35 (0.84)	0.18	0.68
	Expressions detected	4.39 (0.80), 4.25 (0.89)	0.53	0.47
	Agent present x Expressions detected		2.95	**0.09**
3	Agent present	4.11 (1.03), 4.39, (0.72)	2.11	0.15
	Expressions detected	4.42 (0.72), 4.07 (1.02)	3.10	**0.08**
	Agent present x Expressions detected		6.10	**0.02**
4	Agent present	4.07 (0.90), 4.00 (1.10)	0.05	0.83
	Expressions detected	4.13 (0.99), 3.93 (1.02)	0.63	0.43
	Agent present x Expressions detected		0.63	0.43

<u>Question one.</u> The first question asked the students whether or not they enjoyed using Easy with Eve. As shown in Table 6.2, students rated the text-based feedback versions

of Easy with Eve to be significantly more enjoyable than the animated agent versions (*M*= 4.74, *SD*= 0.45 versus *M*= 4.36, *SD*= 0.83), *F* (1,55) = 5.84, *p* = 0.02. However, there was no main effect of facial expression input, and neither was there an interaction between the two variables.

Question two. The second question asked the students whether they would like to use Easy with Eve again. As shown in Table 6.2, students did not rate the text-based feedback versions of Easy with Eve to be more or less desirable to use again than the animated agent versions. Similarly, there was no main effect of facial expression input, although there was a marginally significant interaction between the two variables, with facial expression detection increasing student ratings for the animated agent version more than for the text-only version, *F* (1, 55) = 2.95, *p* = 0.09.

Question three. The third question asked the participants how much they considered themselves to have learnt by using Easy with Eve. As shown in Table 6.2, students did not rate their learning using the text-based feedback versions of Easy with Eve to be significantly more or less than their learning with the animated agent versions. However, the facial expressions detected group did report marginally significantly higher learning than the facial expressions not detected group (*M*= 4.42, *SD*= 0.72 versus *M*= 4.07, *SD*= 1.02), *F* (1, 55) = 3.10, *p* = 0.08. There was also a significant interaction between the two variables, with facial expression detection increasing student ratings for the animated agent version more than for the text-only version, *F* (1,55) =6.10, *p* = 0.02.

Question four. The fourth question asked the students whether or not they believed that the web-cam was working (i.e. whether or not their facial expressions were being analysed). As can be seen from Table 6.2, the mean responses to this question were that the camera was probably (but not definitely) working; the overall mean across all groups was *M*= 4.03 (*SD*= 1.00), where a rating of 5 indicated certainty that the camera was working, and a rating of 1 indicated certainty that the camera was not working. As shown in Table 6.2, participants that used the text-based feedback versions of Easy with Eve did not respond significantly differently to this question compared to participants who used the animated agent versions. Similarly, there was no significant difference between the facial expressions detected groups and the facial expressions not detected groups and neither was there an interaction between the two variables.

Question 5. Question five asked participants in the two animated agent groups how realistic they deemed Eve to look. A *t*-test found no significant difference between the responses of the facial expressions detected group and the responses of the facial expressions not detected group (Detected *M*= 3.40, *SD*= 1.35 versus Undetected *M*= 3.38, *SD*= 1.26), F (1, 55) < 1, n.s.

6.3 Summary

This chapter has presented the methodology and results of a study of the effectiveness of Easy with Eve. Two key independent variables were explored in this study: whether or not the animated agent Eve was present, and whether or not the participant's facial expressions were detected. These variables were explored across two different measures: the increase in test scores between participants' pre-tests and post-tests, and a questionnaire that the participants completed after their tutoring session was complete. These results will be discussed in the latter half of the following chapter.

Chapter 7

Discussion

In this chapter we discuss all that has been presented so far in this thesis. The discussion falls into four sections, matching Chapters 3 to 6 respectively: the video study of human tutors from Chapter 3; the case-based tutoring strategies module from Chapter 4; the implementation of Easy with Eve from Chapter 5; and the study of the effectiveness of Easy with Eve from Chapter 6. These chapters will now be discussed in turn.

7.1 The video study of human tutors

In this section we discuss five issues arising from the observational study of human tutors that was presented in Chapter 3: whether or not it is possible to apply the results of the study across different domains; shortcomings in the amount of data that were generated; assumptions about the expertise of the human tutors; whether or not it is valid to group together data generated from different tutors given their potentially different tutoring styles; and improvements that could be made to the coding scheme that was used to describe the interactions.

7.1.1 Can the results of the study be applied to different domains?

As was shown in Chapter 3, the observational study of human tutors only directly involved the domain of part-whole addition. However, it may be the case that tutors interact differently with students in different domains – and if the domain had been introductory physics, or spelling, or art history, the data that were generated may very well have been different. Similarly, the age of the participants was between 8 and 9; the interactions between tutors and students may have differed if the students had been at secondary or tertiary level.

Therefore, it would be rash to too readily apply the results of the observational study to different domains. However, this posed no problem for the validity of Easy with Eve, as the same domain was used in Easy with Eve that was used in the observational study.

To find out whether or not tutors *do* interact differently with students depending on the domain or the age of the students, a set of similar studies would need to be carried out that sampled various domains and student ages. It would be important to keep the coding scheme consistent across all of the studies to make sure that any comparisons between the results of the studies were valid.

7.1.2 What can we assume about the expertise of the tutors?

The entire observational study was conducted on the assumption that the three human tutors that were videoed were sufficiently competent to be worth copying in a tutoring system. This was felt to be a safe assumption, as both the professional tutor and the two teachers from the school had significant experience with the New Zealand Numeracy Project, from which the tutoring exercise used in the study was taken. Measuring the expertise of the tutors was felt to be beyond the scope of this thesis; nonetheless, this should be noted as underlying assumption behind the study.

7.1.3 Are there enough data?

The data generated by the video study were made up of slightly over 3000 coded turns between tutors and students; on first glance, this seems like it should have been sufficient to start drawing conclusions about tutor-student interactions. However, as is apparent from the tables in Section 3.2.4, some of the turns in the coding schemes occurred only very rarely. For example, there are a lot of data about how tutors act after a student answers a question correctly, because that happened on over 800 separate occurrences – but there are very little data about how tutors act after a student gave a partial answer, as that occurred only 23 times. Similarly, there are plenty of data about how effective the tutors were when they pumped for additional information, which happened over 450 times, but very little about the effectiveness of reminding examples, which happened only 7 times.

Furthermore, the problem that some categories in the coding scheme are underrepresented becomes exponentially more concerning when we also come to consider facial expressions and intensities as well as tutor/student turns. For example, there are plenty of data about how tutors act following a student answering a question correctly with a neutral expression, because that happened very often – but there is no data at all about correct student answers with a high intensity confused expression, as that never happened at all. In other words, the chances that a category in the coding scheme is underrepresented becomes much greater when the turn categories are each split up further into the 9 expression categories, and then each of these new categories is then split up again into low and high intensity categories. This issue is also compounded by the fact that by far the majority of expressions were neutral – which unfortunately means that in many cases there is little data describing interactions that involve non-neutral expressions. This makes it difficult to draw solid conclusions about how tutors and students interact in the underrepresented categories.

The reasons for the lack of emotional expressions in the data can be attributed to three main reasons:

- Firstly, the domain of counting and addition may not have been the most emotive domain that could have been chosen; as we noted in Section 7.1.1, if a different domain had been chosen there may have been different results.

- Secondly, it is hard to know for sure to what extent students *were* feeling the emotions in the coding scheme, but *not* changing their facial expression. For example, if a student was confused, yet still maintained a neutral expression, a lack of emotion itself would not be the issue, so much as a lack of facial expression.

- Thirdly, the fact that students rarely appeared confused, bored, disgusted, surprised, apprehensive or frustrated may in fact merely be a product of the fact that the human tutors were doing a good job – as all of those could possibly be expressions that a human tutor might (rightly or wrongly) aim to minimise.

In response to these reasons then, ideas to generate more emotional data from students in a future study could include the following three strategies:

- Firstly, the study could be repeated with a different, and possibly more emotive domain, such as perhaps a subject from the humanities. The domain could even be targeted to elicit particular expressions, such as surprise or boredom, to see how human tutors act in response to these expressions.

- Secondly, the study could use media other than facial expressions to measure the affective state of students. This could involve wearable computers, voice analysis, or any of the other ways that emotion can be conveyed that were discussed in Section 2.2.1.

- Thirdly, if it is true that tutors minimise the occurrence of certain emotions, perhaps the ultimate solution will simply be to collect significantly more data than was the case with the study reported in Chapter 3.

Therefore one shortcoming of the observational study was an overall lack of data, and if this research is extended in the future, this is one shortcoming that will need to be addressed. However, it was felt that extending the study with this shortcoming in mind was well beyond the scope of the exploratory research bounded by this thesis, and would need to be left for future work.

7.1.4 Should the data from different tutors be grouped together?

There were three different tutors that took part in the observational study, and the results that were presented in Chapter 3 are a collation of the data from all three of the tutors. However, it is possible that each tutor might act differently in a particular tutoring scenario; this would raise the question of whether or not it is valid to group together the data from the three different tutors, or whether the data from each tutor should only be examined individually. If it were true that individual tutors often interact with students in significantly different ways to other tutors, then it could follow that it was wrong to group the data from the different tutors together in the results of the observational study, and that each tutor's data should only be looked at separately as a individual case-study.

Unfortunately, investigation into this issue was unable to provide solid evidence one way or the other. To test for significant differences between how the three individual tutors interacted with students, an attempt was made to use *t*-tests to compare the data from each of the tutors to the data from the other two tutors. Unfortunately though, this analysis was hampered by the same lack of data that was discussed in the previous section. As can be imagined, if there was already a shortage of data in many categories of turn/expression/intensity as it was, this shortage was made three times worse by the time that each of those categories was divided into three more categories to separate out the data from the three different tutors. Due to this fact, the *t*-tests would not have been valid because the sample sizes were too small, and were thus not able to give a clear result on the issue. As there was no clear result, the data were left as they were, with the data from the three different tutors grouped together.

As it stands then, an underlying assumption behind the results of the observational study is that all three of the tutors responded to students in similar enough ways so that the data from each of their tutoring sessions can justifiably be grouped together. Future research into this issue would need to follow the suggestions raised in the previous section to gather more data more evenly spread across the different turn/expression/intensity categories; this would make the *t*-tests comparing tutors to be valid, and thus give some idea of the level of agreement in tutoring actions between the different tutors.

7.1.5 Improvements to the coding scheme

As the tutoring system Easy with Eve was being developed, it became apparent that the coding scheme that was used to classify the data from the human tutoring videos meant that some of the data that was collected was ambiguous, and also that some potentially noteworthy information about the interactions between tutors and students was left uncoded altogether. These limitations of the coding scheme, as well as corresponding suggestions to improve the coding scheme, are presented in the following paragraphs.

<u>Classifying different types of correct answers.</u> One of the main limitations of the coding scheme was a lack of distinction between whether student answers completed an entire problem, or merely a sub-question within the problem (in Levels 1, 2 and 3 a problem was divided into a set of questions). For instance, if by giving a correct answer a student has just completed an entire problem, then a tutor is likely to respond by giving the student a new problem; however, if by giving a correct answer the student has only completed one question within a wider problem, then a tutor is likely to respond by pumping for additional information, or asking a new question. Unfortunately, this distinction was not made in the coding scheme for the observational study, which has meant that Eve’s responses to student answers fail to distinguish between answers that complete a question within a problem, and answers that complete an entire problem.

Thus, one suggestion to improve the coding scheme in this area would be to distinguish between two types of student answers: student answers that complete a sub-question within a problem, and student answers that complete an entire problem.

<u>Classifying different types of tutor question.</u> Similarly, there was also ambiguity in the coding scheme caused by a lack of distinction between different types of tutor questions. In particular, the content of questions could be completely different depending on whether or not the student has just answered a question correctly. For instance, in the context of a correct student answer, the tutor turn *ask new question* probably means that the tutor should move on to the next question; however, in the context of an incorrect student answer, the tutor turn *ask new question* probably means that the tutor will ask another question related to the same question. However, this distinction was lost within the data.

Thus, one suggestion to improve the coding scheme in this area would be to distinguish between two types of tutor questions: tutor questions that are related to the current question within a problem, and tutor questions that are related to the next question within a problem.

Classifying the strength of hints / discussion. Another limitation of the coding scheme was that no description was given of the strength of a hint, or the level of detail given when discussing or rearticulating a problem, question or solution. This meant that it was not possible to tell if hints became stronger or weaker if a student consistently failed to answer a question correctly.

Thus, one suggestion to improve the coding scheme in this area would be to distinguish between several grades of hints / discussion; for example, there could be three different levels: not very strong, moderately strong, or very strong.

Maintaining expression data in composite turns. A final ambiguity within the coding scheme was related to a loss of expression information within composite turns (when two or more sequential tutor or student turns were grouped together to form a new, single turn). This was because composite turns only had one expression, even though the separate turns that made up the composite turn may have had different expressions. In the data, the expression for a composite turn was made the same as the first expression in the sequence of turns, but unfortunately this meant that any information about subsequent expressions in the sequence of turns was lost.

Thus, one suggestion to improve the coding scheme in this area would be to maintain information about the sequence of expressions in composite turns by allowing each turn to correspond to a sequence of expressions. Ideally, this sequence of expressions would also be time stamped to record exactly when each expression started, and how long it lasted.

Expressions of the non-actor. Finally, it was also realised that potentially noteworthy information about what non-actors do in the interactions between tutors and students was left uncoded altogether. For instance, the coding scheme only recorded information about what the *actor* was doing at any given point in time, not what the *non-actor* was doing. This means that the data fails to give an indication of how the human tutors acted while they were waiting for a student to respond to their question – yet for an animated tutoring character such as Eve to appear realistic, this is in fact a useful piece of knowledge.

Thus, one suggestion to improve the coding scheme in this area would be to also code the behaviour of the non-actor during every turn. This would significantly augment what we could learn about tutor-student interactions from the tutoring videos.

7.2 The case-based tutoring strategies module

In this section we discuss several issues arising from the case-based reasoning tutoring strategies module that was presented in Chapter 4: how the amount of data from the video study affected the tutoring strategies module; and the method used to determine the weights linking similar turns and expressions.

7.2.1 Amount of data from the video study

As we discussed in Section 7.1.3, the video study of human tutors yielded data that was sparse in some areas; obviously this had flow-on effects to the output of the tutoring strategies module, as the tutoring strategies module took the data from the video study as its input. Although the effects of this lack of data were mitigated by the fact that *similar* cases were searched for in the data as well as exact matches, the fact remains that the larger the case-base is, the more useful the output of the module would be.

7.2.2 Selecting weights between similar turns and expressions

The weights that linked similar turns and expressions in the case-based tutoring strategies module were chosen based on familiarity with the data. A better solution would have been to survey teaching faculty on what they considered the closeness of links between particular turns and expressions to be, but as was mentioned in Section 4.4.2, the logistics of carrying out such a survey would not have been easy. For example, to survey faculty on the weights of links between each of the 140 tutor turns with each of the other 139 turns would have resulted in a matrix with almost 20,000 questions to fill out, which was clearly not ideal. However, it would be good if a viable, objective measure of determining the weights of links between similar turns could be devised; this is certainly an avenue for future work.

7.3 The implementation of Easy with Eve

In this section we discuss several issues arising from the implementation of Easy with Eve that was presented in Chapter 5: how the animation of the agent Eve could be improved; and how improvements to the facial expression recognition system could have made Eve more effective.

7.3.1 Animation of Eve

The tutoring character Eve was animated by pre-recording all of her possible tutoring actions; thus, whenever Eve interacted with a student the appropriate tutoring videos were being selected and played. The voice of Eve was also pre-recorded. However, there were three main limitations of this approach: that Eve was confined to a finite number of tutoring responses; that the voice of Eve did not vary according to her facial expression; and that the transitions between Eve's facial expressions were not always smoothly animated. We now discuss each of these issues in turn.

<u>Finite tutoring responses.</u> The first danger of a finite set of tutoring responses is that it raises the possibility that Eve may begin to repeat herself, and thus detract from her overall believability. Approximately 1000 tutoring videos of Eve were pre-recorded for use in the tutoring system; on the surface, that might appear to be sufficient, and one could be forgiven for thinking that 1000 different tutoring actions would comfortably allow Eve to not repeat herself. However, in reality many of these videos had exactly the same dialogue, with Eve's facial expression being the only point of difference, and there were often only 3 or 4 videos with different dialogue that matched a particular tutor turn for a particular question in a problem. It is not hard to imagine that if a student were to repeat a question several times – by failing to answer the question correctly – that it would not take long for the dialogue to begin to repeat itself.

Having said that, the fact that Eve could repeat herself did not detract much from the testing of Easy with Eve that was reported in Chapter 6, as the chances of this happening were significantly lessened by the relatively brief period of time that the students interacted with the system. Nonetheless, a better approach would be to animate the videos in real time, using synthesised speech. This would give Eve a much larger set of possible tutoring responses to choose from, and thus considerably reduce the chance that she might repeat herself.

The second danger of having a finite set of responses was that, once begun, a video of Eve had to be completed before the next video could begin; this made it difficult for Eve to quickly adapt to students. This was because the appearance of seamless continuity between the tutoring videos was based on the fact that each video had the same start and end frame (although sometimes Eve's facial expression could be an exception to this) – which meant that the end of one video had be played before the next video could commence. If a video had been cut short to be interrupted by a different video, then there would have been an awkward jump in the animation of Eve when her body parts suddenly skipped to a slightly different location.

The impact of this limitation was that Eve had to finish what she was saying before she could say anything else. The largest implication of this was that students had to wait for her to finish asking a question before they could enter their answer; the alternative was that Eve continued to ask the question even after the student had already answered it, which would have detracted from her believability even more. Thus, the lesser of the two evils had to be accepted, that students could not answer questions until Eve had finished asking them. This was a weak point in the user interface of Easy with Eve, and led to mild confusion and frustration with several users during the testing at local schools.

Again, the resolution of this issue would be to animate Eve's tutoring in real time instead of using pre-recorded videos; this would enable Eve to react to students immediately, and would further increase her believability. However, as explained in Section 5.3.1, the reason behind not animating Eve in real time was that it was beyond the scope the current research. Therefore, whilst animating Eve in real time would have been the ideal option, the daunting logistics of undertaking that project necessitated that it would have to be left for future work. Thus the use of pre-recorded tutoring videos was considered sufficient for the current research.

Variation of the tone of Eve's voice. As discussed in the previous section, the voice of Eve was pre-recorded rather than synthesised, and these sound clips were fed into the animation tool that was used to create the tutoring videos. However, these sound clips were not recorded separately for the different facial expressions that Eve might be showing when the videos are played; for example, if Eve made a comment with a confused expression, she would sound exactly the same as if she had made the comment with a neutral or a smiling expression. Therefore, it would have added at least a little to the believability of Eve if the sound clips of her dialogue had been recorded separately for each different facial expression that might be used for a video. Alternatively, if Eve were to be animated in real time, then perhaps an artificial voice could be used that included variation based on her projected emotional state.

Transitions between Eve's facial expressions. As has already been noted, the tutoring videos in Easy with Eve were all pre-recorded, and each video featured Eve with a particular facial expression. These facial expressions were built into the animation tool that was used to create the videos, and unfortunately it was not possible to control the “rise and fall” of these expressions, which waxed and waned according to the beat of their own inbuilt drum. This meant that depending on the length of the dialogue, a video might finish with Eve in the middle of showing an expression, as opposed to showing her having just completed showing an expression. Therefore, occasionally the transitions between expressions in consecutive videos were not always smooth. For example, Eve could jump from a smile at the end of one video to being neutral at the beginning of the next – and in that particular example, the sudden change in expression would actually make it seem as if Eve had suddenly been annoyed, which could potentially mislead the student and negatively influence the interaction.

Again, this limitation could be addressed if Eve were animated in real time. This would give the tutoring system much greater control of the expression that Eve showed at any particular time, and would also allow the tutoring system to control the onset and fading of her expressions.

7.3.2 Accuracy of the facial expression recognition

Easy with Eve determined the current affective state of a student on the basis of the output from the facial expression analysis system that was used in the tutoring system. However, there were two main issues relating to the facial expression analysis system: the range of emotions that it was capable of classifying; and the accuracy of recognition when using live subjects as opposed to an artificial data set. These will now be discussed in turn.

<u>Range of emotions classified.</u> One limitation of the facial expression analysis system related to the range of affective states that it was capable of detecting. In particular, the emotions that it was capable of detecting were neutral, smiling, laughing, surprised, angry, and disgusted. However, the reader may realise that this list does not align exactly with the range of states that were used to classify emotion in the observational study of human tutors – these states were neutral, happy, confused, disappointed, bored, surprised, apprehensive and disgusted.

Therefore, the range of emotions that was able to be detected by the facial expression analysis system meant that not all of the data from the observational study could be accessed by the tutoring strategies module. This was because data from the observational study where students displayed confusion, boredom or apprehension would never be searched for by the tutoring strategies module, as these expressions would never be detected by the facial expression analysis.

This issue was somewhat mitigated by the fact that, as shown in Section 3.2.4, only a relatively small minority of student expressions in the observational study (9%) were neither neutral or smiling. Nonetheless, the fact that the facial expression analysis system meant that this 9% of data was not able to be accessed by the tutoring strategies module was a limitation of the study, and could certainly be improved upon in future work.

<u>Accuracy of recognition.</u> As noted in Section 5.2, the facial expression recognition system had only been tested on an artificial data set before it was used in the study of Easy with Eve. This meant that although the accuracy of the system was impressive

with the artificial data set (approximately 90%), the accuracy of the system with real images – such as were used in the study of Easy with Eve – was unknown. This was unfortunate, because the accuracy of the facial expression analysis system was obviously going to influence the degree to which the system could adapt to student emotion – it would only be able to adapt to what it was able to first detect.

As the system was being implemented, it was clear that the accuracy of the facial expression analysis was less than 90%, although it was going to be difficult to estimate a figure for the exact accuracy without carrying out separate study. It was felt that carrying out such a study was beyond the scope of the current research. Nonetheless, this lack of definite knowledge about the accuracy of the facial expression analysis system with real images was certainly a limitation of the study of the effectiveness of Easy with Eve, and is an important topic for future work.

7.4 The study of the effectiveness of Easy with Eve

In this section we discuss the study of the effectiveness of Easy with Eve that was presented in Chapter 6. In particular, we address the results of the study relating to the following issues: the overall effectiveness of Easy with Eve at improving student test scores; the effect of the presence of the animated agent on student test scores; the effect of detecting facial expressions on student test scores; the effects of the presence of the animated agent and detecting facial expressions on student motivation; the effects of the presence of the animated agent and detecting facial expressions on student perceptions of how much they learned; whether or not students could tell if their facial expressions were being detected; and finally the student ratings of the believability of Eve.

7.4.1 Overall effect of Easy with Eve on student test scores

The results from the study showed a statistically significant overall increase in student scores from pre-test to post-test (combining all four experimental groups). This result is certainly very encouraging, as it goes some considerable way towards validating the use of the tutoring system as a tool to help students improve their addition.

However, there is one possible reason for caution in interpreting this result, which is related to how familiar students were with the counting and addition exercise that was used as the domain for the tutoring system. In particular, it is possible that an initial lack of familiarity with the exercise caused students to score below their potential in the pre-test; by the time the students reached the post-test, this lack of familiarity would obviously no longer be an issue, which could partly explain why the post-test scores were greater than the pre-test scores. On the other hand though, it was expected that students *would* be familiar with the exercise, as it was taken directly from the New Zealand Numeracy Project curriculum, which is used in New Zealand schools; this issue was also discussed with the teachers at the schools before the testing was carried out. Therefore, even though the study was designed so that a lack of familiarity with the domain would not be an issue, and it is likely that this was indeed the case, it would be best to specifically test for this in any future studies.

In future work, the best way to ensure that students are familiar with the exercise would be to have the students use the tutoring system in a practice session first, or to run the experiment over several different tutoring sessions so that each student uses the system more than once. This would eliminate the potential confounding variable of familiarity when interpreting any increase from pre-test to post-test.

7.4.2 Effect of the presence of the animated agent on student test scores

The results from the study showed no significant effect of the presence of the animated agent on the improvement of student scores from pre-test to post-test. In other words, students in the no-agent groups improved their test scores in similar ways to the students who did interact with Eve.

This result was not unexpected. Although it was anticipated that Eve would create a persona effect, research to date suggests that this relates more to student motivation and perception of learning than it does to actual short-term student performance (van Mulken, André, & Muller, 1998; Prendinger, Mayer, Mori, & Ishizuka, 2003). (However, if a student’s motivation and perception of learning is high, then long-term performance is likely to benefit from the increased effort and attention.) Therefore, this result does not impact on the extent to which Eve carries a persona effect.

However, it should be noted that the relatively small sample size in each group (even though there were 60 students in total) was only big enough to detect large effects caused by variation between the groups. This means that there may in fact have been an effect, though unexpected, of the presence of Eve on student performance, but the small sample size of the experimental groups meant that this effect was not shown in the data.

Future work could confirm whether or not the presence of Eve affects student short-term performance by using a larger sample size in each of the experimental groups. It should also be noted that this issue is also relevant to every point of analysis in this study that involves comparisons among the four different experimental groups.

7.4.3 Effect of detecting facial expressions on student test scores

The results from the study showed no significant effect of detecting facial expressions on the improvement of student scores from pre-test to post-test. In other words, students in groups without the facial expression analysis improved their test scores in similar ways to the students in the groups with facial expression analysis. This result suggests that incorporating the emotional state of the student into the tutoring strategies selection process did not improve the appropriateness of material that the tutoring system presented to the students – or at least not in such a way as to significantly increase a student’s short-term performance.

However, there are three main reasons why this result must only be assessed tentatively.

- Firstly, as discussed in the previous section, the sample size may have been too small to detect the effect caused by detecting facial expressions (if there was one). It is possible that there *was* either a small or medium effect, but the sample size was not large enough to detect it.

- Secondly, the limitations of the facial expression recognition system discussed in Section 7.3.2 would without doubt have affected the quality of data that was input to the tutoring strategies module, resulting in potentially less useful output than the tutoring strategies module was capable of producing.

- Thirdly, the tutoring strategies module itself may have suffered from the lack of data in the observational study, as was discussed in Section 7.2.1; this lack of data would also have affected the output of the tutoring strategies module.

To enhance the certainty with which conclusions could be drawn about the effects of including student emotion in tutoring strategies selection on student short-term performance, future work could pursue the following suggestions.

1. Use a larger sample size that would show small or medium effects as well as large effects. For a 2x2 ANOVA test to achieve 80% power for a medium effect (f=.25), 33 participants would be required in each cell, or 132 in total; to achieve 80% power for a small effect (f=.10), 197 participants would be required in each cell, or 788 in total (Cohen, 1988).

2. Improve the method of emotion detection. This could mean either improving on the facial expression analysis system that was used in the current study, or it could mean using other media for detecting emotions. (Using other media would also require a new study of how human tutors adapt particularly to those media, as the observational study used in the current research was only directly concerned with facial expressions.)

3. Gather more data to be used by the tutoring strategies module, thus improving its output.

The remainder of this chapter now addresses the results from the questionnaire that were described in Section 6.2.3.

7.4.4 Effects of the animated agent and detecting facial expressions on student motivation

The first two questions in the questionnaire (refer to Section 6.1.3 for the list of questions, or to Appendix B for a copy of a questionnaire) gave an insight into how the presence of the animated agent and the detection of facial expressions affected student motivation, as measured by a 5-point Likert scale. In particular, the results from the first question showed that students rated the text-based feedback versions of Easy with Eve to be significantly more enjoyable than the animated agent versions. However, the results of the second question found that although students did not rate the text-based feedback versions of Easy with Eve to be significantly more or less desirable to use again than the animated agent versions, there was a marginally significant interaction between the two variables: student ratings were higher for the animated agent when this was combined with facial expression detection.

These results were interesting because overall they did not support the persona effect. If the persona effect was to have been supported by the results, then we should have seen that the presence of the animated agent both made the tutoring session more enjoyable, and increased students' desire to use the system again. Not only was this not the case, but the results of the first question were exactly the opposite, that students rated the text-based feedback versions to be significantly more enjoyable than the animated agent versions.

But on the other hand, the marginally significant interaction in the second question between the presence of the animated agent and facial expression detection is fascinating too, because this may suggest that there was indeed a slight persona effect, and that it was amplified when the animated agent is enabled to detect and adapt to the affective state of the student. This was one of the main aims of Easy with Eve, so this was certainly an encouraging result that should be explored in future work.

However, there are three main reasons why these results need to be assessed with caution; these reasons are now discussed in turn.

1. Firstly, the limitations of the animation of Eve that were discussed in Section 7.3.1 may have had a direct effect on the believability of Eve, and thus on any potential persona effect. For instance, it seems likely that the less fluent or lifelike that Eve's actions were, the less of a "persona" would be attributed to Eve by students. In other words, for student motivation to be enhanced by interacting with an agent that appears to have a personality, it is clearly important that the agent does appear to have a personality. Therefore, any of the above-mentioned shortcomings in the animation of Eve may have detracted

from her appearance as a believable character, and thus detracted from her ability to convey a persona.

Future work could address this issue by carrying out the suggestions to improve the animation of Eve that were given in Section 7.3.1. Once these improvements were made it would be possible to paint a much clearer picture about how powerful a persona effect Eve is capable of creating. If there was a persona effect, we would expect future studies to show that students rate animated agent versions of Easy with Eve as both more enjoyable and more desirable to use again than versions with simpler, text-based feedback.

2. Secondly, the limitations of the facial expression analysis system that were discussed in Section 7.3.2 may have also had an effect on the believability of Eve. For instance, it could be the case that shortcomings in facial expression recognition could cause Eve (in the animated agent versions) to act in a manner that seems oblivious to the affective state of the student, and thus she might appear less believable to students. This is important for the reason given in the previous paragraphs, that the persona effect may rely largely on the extent to which Eve can portray the appearance of a believable human-like personality.

   Again, future work could address this issue by improving the performance of the facial expression analysis system, which was discussed in Section 7.3.2. If there was a persona effect, we might expect future studies to show that students rate the facial expressions detected and animated agent version of Easy with Eve as both more enjoyable and more desirable to use again than versions of Easy with Eve without the facial expression detection. (Whether or not there can be a persona effect for a text-based feedback version of Easy with Eve is an interesting question that further studies may cast light on.)

3. Finally, the results from the questionnaire may have been affected by a ceiling effect in the data. That a ceiling effect may have affected the results is clear from the data in Appendix D, as many students from all the experimental groups rated their responses for the two questions (and also the third and fourth questions, but not the fifth) as the maximum score of 5 out of 5. This means that it is difficult

to tell the difference between the students that enjoyed using the system a little and the students that enjoyed using the system very much, as they both may have responded with the maximum 5. Thus, the implication of this when interpreting the data is that the real distinctions between the ratings in the different experimental groups may have been blurred because so many students responded with the maximum score. The reality might be that a clearer measure would show that there is less of a difference between the groups, or more of a difference between the groups – either way, as the current results stand, the possibility of a ceiling effect makes it difficult to read too much into the data.

Therefore, future work would need to address the measure that students use to rate their feelings about the tutoring system; in particular, the measure needs to allow for clearer distinctions between student responses, especially at the top end of the scale. Perhaps one way to do this could be to use a continuous scale instead of a discrete scale; students would draw a line on a continuum to indicate their rating rather than choose one of five discrete values. Also, it is possible that the age of the participants may have been a factor in the presence of a possible ceiling effect, as young participants may be more likely to give maximum ratings than older participants; this would also need to be considered when designing a measure for student responses in future work.

7.4.5 Effects of the animated agent and detecting facial expressions on student perceptions of learning

The third question in the questionnaire gave an insight into how the presence of the animated agent and the detection of facial expressions affected how much students believed that they learned during the tutoring session. In particular, the results from the third question showed that students who used the facial expressions detected versions of Easy with Eve rated their amount of learning as marginally higher than the students who used the non-facial expressions detected versions. However, students did not rate their learning using the text-based feedback versions of Easy with Eve to be significantly more or less than their learning with the animated agent versions. Importantly though, there was a significant interaction between the two variables, with facial expression detection increasing student ratings for the animated agent version more than for the text-only version.

These results are interesting because they show a similar interaction between the two variables to that which we saw in the results for Question 2; the results do show support for a possible persona effect. This is because the versions where facial expressions were detected were rated as leading to more learning than the non-facial expressions detected versions, and especially in conjunction with the presence of the animated agent, which is the result that we would expect to see if the persona effect were valid, as was discussed in the previous section. On the other hand, it is noteworthy that this result was not reflected by the presence of the animated agent alone; we would expect a persona effect to cause students who interacted with the animated agent versions of Easy with Eve to rate their learning higher than the students who used the text-based systems, regardless of facial expression detection, but this was not the case. However, it must be noted that the results for this question must also only be assessed cautiously. This is for the same reasons that the results for Questions 1 and 2 in the questionnaire should be assessed cautiously: that the believability of the animation of Eve and the accuracy of the facial expression analysis might affect her ability to convey a persona; and the presence of a ceiling effect may mean that the data is misleading. Future work could address these issues as discussed in the previous section.

7.4.6 Students' ability to tell if their facial expressions were detected

The fourth question in the questionnaire gave an insight into whether or not students could tell if their facial expressions were being detected. In particular, the results from the fourth question showed no significant difference between the responses of the four groups, regardless of whether or not facial expressions were detected or the animated agent was present. There was no interaction between the two variables.

These results seem to reflect the limitations of the facial expression analysis and the animation of Eve that were discussed earlier. This is because if the tutoring system was adapting to the student in a particularly empathetic manner, one might expect the students in the facial expressions detected groups to rate that the web-cam on the monitor was working with a fair degree of certainty; however, the results showed that the means of the four groups were almost identical. Therefore the results could lead to any of the following possible conclusions:

- that facial expression analysis in a tutoring system only makes a noticeable difference to students when coupled with an animated agent, but that the limitations of the animation of Eve counteracted the good work of the facial expression analysis;

- that the facial expression analysis would need to be improved before it would make a noticeable difference to students;

- that there would need to be more data from the observational study of human tutors before a noticeable difference could be made;

- or possibly any combination of the above.

Unfortunately, it is not possible to say with any certainty which of these conclusions is the correct one without running further tests; nonetheless it seems not unlikely that all of the conclusions are in fact correct, and that improvements to the animation of Eve, to the facial expression analysis and to the body of data from the observational study would all help to enhance the impact of the use of detecting facial expressions in a tutoring system. Future work could explore these questions by making the above mentioned improvements, and seeing if this alters the results of similar studies of the effectiveness of Easy with Eve.

7.4.7 Student ratings of the believability of Eve

The fifth and final question in the questionnaire gave an insight into how believable students rated the animated agent Eve to be. Neither the facial expressions detected group or the non-facial expressions detected group rated the believability of Eve to be higher than a mean of 3.40 (midway between *a little* and *quite a bit* on the questionnaire); in fact, the mean ratings between the two groups differed only by 0.02.

This result suggests that whether or not the facial expressions of students were detected made basically no difference to how realistic students considered the agent Eve to be. This was not the result that was hoped for, as it was hoped that the empathetic adaptations of the facial expressions detected version of Eve would make her significantly more realistic to students than the non-facial expressions detected version.

Again however, this result must be taken cautiously; it would be rash to conclude too much about the potential effects of adapting to the emotional state of the student from this one study alone. This is for the same reasons that have been raised in the three previous sections: that limitations in the animation of Eve and the facial expression analysis detract from her ability to adapt in a believably empathetic manner; and that more data in the observational study would also be very likely to improve output of the tutoring strategies module. If these limitations were addressed in future work, it would be possible to conclude with much more certainty the degree to which facial expression detection can improve the believability of an animated agent.

Chapter 8

Conclusions and future work

The strange thing about life is that though the nature of it must have been apparent to everyone for hundreds of years, no one has left any adequate account of it. The streets of London have their map; but our passions are uncharted.

Virginia Woolf, Jacob's Room

8.1 Summary of research

We began this thesis in Chapter 1 by briefly tracing the history of computers in education; we saw that e-learning has grown into a large, booming industry and an area that is receiving significant attention in academic circles. And yet, in spite of all this research, we saw that even the most sophisticated, modern Intelligent Tutoring Systems (ITSs) still struggle to match competent human tutors.

The reason for the disparity in teaching ability between human tutors and ITSs that was focused on in this thesis was the ability of human tutors to *empathise* with their students: to recognise the student's affective state; to know how to adapt to the student's affective state; and then to carry out this adaptation. Therefore, with the successful empathy of human tutors as our blueprint, the overall aim of this research was to develop an Affective Tutoring System (ATS) capable of recognising both the cognitive and affective state of students, of knowing how to adapt to this information, and of usefully carrying out this adaptation.

In particular, the overall research goals in this thesis were stated as such:

1. To gather data about how human tutors adapt to the affective state of students in a one-on-one tutoring scenario.

2. To develop an intelligent emotion-sensitive method for adapting the tutoring strategies of an ATS based on the observational study of human tutors, and to implement this method in an emotion-sensitive tutoring strategies module.

3. To assess the feasibility of an emotion-sensitive tutoring system by applying the emotion-sensitive tutoring strategies module in a functional ATS. The ATS was to be in the domain of primary school addition, and to feature an animated pedagogical agent.

4. To test the effectiveness of the ATS in local schools. This was to gather data about the effectiveness of the emotion-sensitive tutoring strategies approach, and about the impact of agent-based vs. text-based feedback.

In Chapter 2 we reviewed the literature relating to affect and learning, affective computing, and ATSs. To begin, strong links between affect and learning were demonstrated in two main ways: firstly emotions were shown to influence performance on cognitive tasks such as learning; and secondly the fact was noted that human tutors use affective feedback as a basis for the adaptation of their teaching.

Next, the review of affective computing focussed on three main areas: recognising emotion, showing emotion, and having emotion/emotional intelligence. It was seen that of these areas, research on showing emotion in ITSs is well established through the implementation of animated pedagogical agents, and research on recognising emotions is offering promising results through a number of different media, but little work has been done so far to describe emotional intelligence in the context of an artificial affective tutor. It was noted that the challenge of the current research was to combine all three of these areas, and thus to create a system that can recognise emotion, that can show emotion, and that can respond to the student in an emotionally intelligent way.

The review of affective computing was then followed by a review of affect-sensitive tutoring systems in particular. It was seen that research towards a fully functional ATS is still in its infancy, across all of the work that was reviewed; although several groups have had encouraging results in developing affect-detecting ITSs, even the most impressive of these results leave significant room for major improvements. In particular, it was noted that an area of ATSs that has received little attention is the question of *how* to adapt to student emotion once it has been recognised. No system had yet been developed that responds to the affective state of students based on solidly grounded tutoring strategies; investigating a method for how to adapt to the affective state of students was thus one of the major goals of this thesis.

Then in Chapter 3 we presented an observational study of human tutors that was designed to gather data about tutor adaptations to students that could be used in the development of a tutoring strategies module for an ATS. This study involved videoing three human tutors as they interacted one-on-one with a number of children from a local school as they worked through a New Zealand Numeracy Project counting and addition exercise. The data was coded using a scheme that was modified from previous research by Person and Graesser (2003) to include information about the facial expressions of tutors and students; the data was later validated by an inter-rater reliability study. Several interesting findings from the study were as follows: that students, and especially tutors, are predominantly neutral in their facial expressions; apart from smiles, non-neutral expressions were rare for both students and tutors; and tutors were more likely to smile following a student smile, but very unlikely to smile following a student neutral expression.

However, the main outcome of the observational study was simply the collection of the data itself, and this data formed the basis of case-based tutoring strategies approach for adapting to student affect, which was then discussed in Chapter 4. The aim of the case-based method was to use the data from the study of human tutors to recommend tutoring actions based on sequence of interactions between tutors and students, taking both turns and expressions and intensities into account; the case-based method was implemented in an emotion-sensitive tutoring strategies module. The final version of the tutoring strategies module used a fuzzy approach to search for all the sequences in the data that were similar to the current scenario, where each sequence is weighted according to its similarity to the current scenario. The tutoring strategies module was able to run in real time.

The fuzzy tutoring strategies module formed the foundation of the first ever ATS, Easy with Eve, which was then discussed in Chapter 5 as an exploration of the feasibility of an emotion-sensitive tutoring system. The ATS features Eve, a novel emotion-sensitive animated pedagogical agent that can recognise, express and adapt to a set of student emotions; the system uses real time facial expression analysis as a measure of affective state. The chapter began with a discussion of the domain and levels within Easy with Eve; the domain and levels of Easy with Eve were taken from the same New Zealand Numeracy Project counting and addition exercise that was used in the observational study of human tutors in Chapter 3. Emotions were detected in the tutoring system through real time facial expression analysis, and the animation of the agent and counter videos was implemented through creating a set of pre-recorded videos that could be played by Easy with Eve. Also, the interface between all of these components and the tutoring strategies module was explained; the facial expressions analysis helps as input to the tutoring strategies module, whose output determines which agent and counter videos to play. Finally, the method and results of several pilot tests were also described, which proved to be useful in preparation for the formal testing of Easy with Eve in schools.

Chapter 6 presented the methodology and results of a study of the effectiveness of Easy with Eve that was conducted at several local schools in Auckland. A total of 59 participants interacted for around 20 minutes with one of four versions of Easy with Eve, prefaced by a short pre-test and followed by a similar post-test and a questionnaire. The four experimental groups were different combinations of the two independent variables in the study: whether or not the animated agent Eve was present, and whether or not the participant’s facial expressions were detected. These variables were explored across two different measures: the increase in test scores between participants’ pre-tests and post-tests, as well as the results from the questionnaire.

The results from the study of Easy with Eve were then discussed in Chapter 7, as well as issues arising from the video study of human tutors, the tutoring strategies approach, and the implementation of Easy with Eve. Firstly, regarding the video study of human tutors, it was noted that the results of the study could possibly have varied had a different domain been used, or the age of the students had differed; it was also noted that an underlying assumption in the study was the competence of the human tutors that were involved. However, the main issue arising from the video study was the low volume of data, although it was felt that gathering more data would have been beyond the scope of this research. It was also questioned whether or not it was valid to generalise the data from the videos across different tutors, although statistically determining the level of divergence between different tutors was hampered due to the overall lack of data. Improvements to the coding scheme were also suggested.

Regarding the case-based tutoring strategies approach, it was noted that the lack of data from the video study had an effect on the quality of the tutoring suggestions that the module could output. Also, it was discussed that it would also have been good to find a more objective way to weight the similarities between different sequences in the data. Improvements in both of these areas would be required before any final assessment of the case-based approach could be made.

The first main issue arising from the implementation of Easy with Eve was that it would be best to animate Eve in real time, as the more smooth her adaptations are, the more believable she could appear to be. Secondly, the other main issue related to concerns about the accuracy and range of emotions detected by the facial expression analysis component, which meant that the affective state input to the tutoring strategies module could have been improved.

Finally, the results of the study of the effectiveness of Easy with Eve were discussed, based on student scores in the pre- and post-tests and student responses to the questionnaires. The first notable result was that although there was an overall increase from pre-test scores to post-test scores, this was not affected by whether or not the affective state of the student was considered by the tutoring strategies module. This meant that the attempt of Easy with Eve to adapt to the affective state of the student did not improve the performance of students on the post-test. However, it was noted that this result may have been a factor of the limitations of the facial expression analysis or the tutoring strategies module; alternatively, the sample size may simply have been too small to detect the effect, depending on the size of the effect.

The most notable result from the questionnaire responses was that although student responses to the first question, relating to student motivation, did not support a persona effect of Eve, the responses to the second question, also relating to student motivation, and the third question, relating to student perception of learning, did to an extent provide possible evidence in favour of a persona effect. The marginally significant interaction in the second and third questions between the presence of the animated agent and facial expression detection may suggest that there was indeed a persona effect, and that this effect was amplified when the animated agent was enabled to detect and adapt to the affective state of the student. This was one of the main aims of exploring an emotion-sensitive tutoring system, so this was certainly an encouraging result. However, all of these results needed to be assessed cautiously as they may have been affected by limitations in either the animation of Eve or the facial expression analysis, or by a likely ceiling effect in the data. Also, student responses to the fourth question showed that students could not tell whether or not their facial expressions were detected by the system, which would appear to cast doubt on the validity of the animation of Eve, or the facial expression analysis, or the tutoring strategies module, or a combination of all three. This finding seemed to be supported by the responses to the final question on the questionnaire, which showed that students rated the believability of Eve to be almost identical, even across the different experimental groups; adding the student emotion to the tutoring strategies module did not appear to make any difference to how realistic students considered the agent Eve to be. However, this result was also only tentative, as it too was likely to have been affected by limitations of the animation of Eve, the facial expression analysis and the tutoring strategies module.

8.2 Future work

The work presented in this thesis represents a stepping stone towards further research into the effectiveness of adapting to student emotion in ITSs. In fact, there remains ample scope for future research to work towards improving Easy with Eve; this section provides several suggestions for potential future studies.

1. The first avenue for future work would be to improve upon the observational study of human tutors, as discussed in Section 7.1. In particular, the coding scheme could be improved using the suggestions in Section 7.1.5, and the study should be repeated with more students to generate more data than was the case in the current research. As discussed in Section 7.1.4, once enough data is generated it will be possible to compare the similarity between tutoring responses of the different tutors in the study, which will cast light on whether it is appropriate to generalise data across different tutors, or if the data from tutors should only be analysed individually.

2. The observational study could also be repeated in different domains and/or with different aged participants to see what effect this has on the interactions between tutors and students. The similarities between the different sets of results from different domains/ages would help determine to what extent it is appropriate to generalise the results from one domain/age to another.

3. The case-based tutoring strategies module would be automatically improved if suggestion number 1 above was followed and if more data was generated; however, it would also be improved if an objective measure was found to gauge the weight of similarities between the different tutor turns, student turns, facial expressions and expression intensities. This would help to validate the fuzzy element used to evaluate the relevance of similar generated sequences.

4. The appearance of Eve could be improved by animating her in real time, but without sacrificing the quality of the current animations. This would potentially give her a much larger range of possible gestures and facial expressions, and allow her to adapt to students in a more immediate manner; this would help to amplify how believable she appears to students, and aid any investigation into presence or otherwise of a persona effect. Similarly, the sound for Eve’s voice could also be generated in real time, using an automatically generated script, although the quality of sound would be degraded.

5. The facial expression analysis system used to detect emotions in Easy with Eve could be improved by detecting a wider range of expressions, and with a greater accuracy. Alternatively, the use of other media for emotion detection such as physiological measures of affective state could be investigated.

6. Finally, the study of the effectiveness of Easy with Eve from Chapter 6 should be repeated to test the effect of implementing the above suggestions. Ideally, there should be enough participants in the study so that medium or even small sized effects could be detected, although the logistics of this would not be easy, especially for detecting small effects. Further studies could also investigate whether the domain of the ATS or the age and gender of the students have any effect on the impact of Eve.

8.3 Conclusions

This thesis has presented work towards the development of an affect-sensitive animated tutoring character named Eve, as part of a new ATS, Easy with Eve, that helps students with a counting and addition exercise. On the basis of the work presented in this thesis, the following are the conclusions that can be drawn:

1. It is feasible to develop an emotion-sensitive tutoring system, as has been demonstrated by the implementation of Easy with Eve.

2. It is possible to model the interaction with the student and to develop emotion-sensitive tutoring strategies by adopting a fuzzy case-based approach, as has been demonstrated by the implementation of the tutoring strategies module in Easy with Eve.

3. The findings from the study of Easy with Eve indicate that adding the detection of facial expressions to the student model does not improve student short-term performance. However, as has been discussed, these findings are tentative, and there is scope for both the implementation of Easy with Eve and the design of the study that tested it to be improved.

4. The findings from the study of Easy with Eve may tentatively indicate that the animated agent Eve carried a slight persona effect, and that this persona effect was greater when the animated agent was enabled to detect and adapt to the affective state (as evidenced by the facial expressions) of students. Thus it could be tentatively concluded that the detection of affective state in an ATS may lead to an increase in student motivation and perception of learning gains when the emotion detection is used in conjunction with an animated pedagogical agent. However, as has been discussed, this conclusion would need to be verified further by future studies.

As suggested in the previous section, there is ample scope for this work to be extended and further explored; nevertheless, this thesis represents exciting steps in the progress towards ATSs that can truly show empathy towards students. May the days of frustrating, confusing interactions with artificial tutors be drawing swiftly to a close!

Bibliography

Aïmeur, E., Dufort, H., Leibu, D., & Frasson, C. (1997). Some justifications for the learning by disturbing strategy. In *Artificial Intelligence in Education*. Kobe, Japan.

Alexander, S. T. V., Hill, S., & Sarrafzadeh, A. (2005). How do human tutors adapt to affective state?. In *User Modeling*. Edinburgh, Scotland.

Alexander, S. T. V., Sarrafzadeh, A., & Hill, S. (2006). Easy with Eve: A Functional Affective Tutoring System. In *Intelligent Tutoring Systems*. Jhongli, Taiwan.

Alexander, S. T. V., Sarrafzadeh, A., & Fan, C. (2003). Pay Attention! The Computer is Watching: Affective Tutoring Systems. In *E-Learn*. Phoenix, Arizona.

Alexander, S. T. V., Sarrafzadeh, A., & Hill, S. (2007). Foundation of an affective tutoring system: Learning how human tutors adapt to student emotion. *International Journal of Intelligent Systems Technologies and Applications*, in press.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences, 4*(2), 167-207.

André, E., Rist, T., & Müller, J. (1998). Integrating reactive and scripted behaviors in a life-like presentation agent. In *Autonomous Agents*. Minneapolis, Minnesota.

Ashby, F. G., Isen, A. M., & Turken, U. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychological Review*, *106*(3), 529-550.

Bareiss, E. R., Porter, B. W., & Weir, C. C. (1988). Protos: An examplar-based learning apprentice. *International Journal of Man-Machine Studies, 29*, 549-561.

Bartlett, M. S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., & Movellan, J. (2006). Fully automatic facial action recognition in spontaneous behaviour. In *Automatic Face and Gesture Recognition*. Southampton, UK.

Bechara, A., Damasio, H., & Damasio, A. R. (2003). Role of the amygdala in decision-making. *Annals of the New York Academy of Sciences, 985(1)*, 356–369.

Berry, D. C., Butler, L. T., & de Rosis, F. (2005). Evaluating a realistic agent in an advice-giving task. *International Journal of Human-Computer Studies, 63*(3), 304-327.

Beskow, J., & McGlashan, S. (1997). Olga - a conversational agent with gestures. In *Animated Interface Agents - Making them Intelligent*. Nagoya, Japan.

Bickmore, T. (2003). *Relational agents: Effecting change through human-computer relationships*. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4-16.

du Boulay, B., & Luckin, R. (2001). Modelling human teaching tactics and strategies for tutoring systems. *International Journal of Artificial Intelligence in Education, 12*(3), 235-256.

Briton, N. J., & Hall, J. A. (1995). Beliefs about female and male nonverbal communication. *Sex Roles, 32*(1), 79-90.

Brody, L. (1999). *Gender, emotion, and the family*. Cambridge, Massachusetts: Harvard University Press.

Burkitt, E., & Barnett, N. (2006). The effects of brief and elaborate mood induction procedures on the size of young children's drawings. *Educational Psychology, 26*(1), 93-108.

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*. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts.

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

Churchill, E. F., Cook, L., Hodgson, P., Prevost, S., & Sullivan, J. (2000). "May I help you?": Designing embodied conversational agent allies. In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill (Eds.). *Embodied conversational agents*. Cambridge, Massachusetts: MIT Press.

Clark, R. E., & Choi, S. (2005). Five design principles for experiments on the effects of animated pedagogical agents. *Educational Computing Research, 32*(3), 209-225.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, New Jersey: Erlbaum.

Cohen, P. A., Kulik, J. A. & Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal, 19*(2), 237-248.

Conati, C. (2002). Probabilistic assessment of user's emotions in educational games. *Applied Artificial Intelligence, 16*, 555-575.

Conati C., & Mclaren, H. (2005). Data-driven refinement of a probabilistic model of user affect. In *User Modeling*. Edinburgh, Scotland.

Conati, C., & Maclaren, H. (2004). Evaluating a probabilistic model of student affect. In *Intelligent Tutoring Systems*. Maceio, Brazil.

Conati, C., & Merten, C. (to appear). Eye-tracking for user modeling in exploratory learning environments: An empirical evaluation. *Knowledge Based Systems, special issue on Advances in Intelligent User Interfaces*.

Conati, C., Merten, C., Amershi, S., & Muldner, K. (2007). Using eye-tracking data for high-level user modeling in adaptive interfaces. In *Artificial Intelligence*. Vancouver, British Columbia.

Corradini, A., Mehta, M., Bernsen, N., & Charfuelan, M. (2005). Animating an interactive conversational character for an educational game system. In *Intelligent User Interfaces*. San Diego, California.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York, New York: Grosset/Putnam.

Derry, S. J., & Potts, M. K. (1998). How tutors model students: A study of personal constructs in adaptive tutoring. *American Educational Research Journal, 35*(1), 65-99.

D'Mello, S. K., Craig, S. D., Gholson, B., Franklin, S. Picard, R. W., & Graesser, A. C. (2005). Integrating affect sensors in an intelligent tutoring system. In *Affective Interactions: The Computer in the Affective Loop Workshop at 2005 International Conference on Intelligent User Interfaces*. San Diego, California.

D'Mello, S. K., Picard, R. W., & Graesser, A. C. (2007). Toward an affect sensitive autotutor. *IEEE Intelligent Systems,* in press.

Donato, G., Bartlett, M. S., Hager, J. C., Ekman, P., & Sejnowski, T. J. (1999). Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 21*(10), 974-989.

Du, Z., & McCalla, G. I. (1991). A case-based mathematics instructional planner. In *Learning Sciences*. Evanston, Illinois.

Duncker, K. (1945). On problem solving. *Psychological Monographs, 58*(5), 270.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*(3-4), 169-200.

Ekman, P. & Friesen, W. V. (1978). Facial Action Coding System. Palo Alto, California: Consulting Psychologists Press.

Elorriaga J. A., & Fernández-Castro, I. (2000). Using case-based reasoning in instructional planning: Towards a hybrid self-improving instructional planner. *Artificial Intelligence in Education, 11,* 416-449.

eMarketer.com (2007). *eLearning 2007 media kit*. Retrieved January 25, 2007, from http://banners.noticiasdot.com/termometro/boletines/docs/consultoras/emarketer/2003/07/emarketer-0703-elearning.pdf

Fan, C., Sarrafzadeh, A., Dadgostar, F., & Gholamhosseini, H. (2005). Facial expression analysis by support vector regression. In *International Image and Vision Computing Conference*. Dunedin, New Zealand.

Fasel, B., & Luettin, J. (2003). Automatic facial expression analysis: A survey. *Pattern Recognition, 36*(1), 259-275.

Fox, B. (1991). Cognitive and interactional aspects of correction in tutoring. In P. Goodyear (Ed.). *Teaching knowledge and intelligent tutoring*. Norwood, New Jersey: Ablex.

Fox, B. (1993). *Human tutorial dialogue*. Hillsdale, New Jersey: Lawrence Erlbaum.

González C., Burguillo, J. C., & Llamas, M. (2000). A qualitative comparison of techniques for student modeling in intelligent tutoring systems. In *Annual Frontiers in Education Conference*. San Diego, California.

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H., Ventura, M., Olney, A. & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavioral Research Methods, Instruments, and Computers, 36*, 180-193.

Graesser, A. C., Person, N. K., Harter, D., & The Tutoring Research Group (2001). Teaching tactics and dialog in AutoTutor. *Artificial Intelligence in Education*, *12*(3), 257-279.

Greene, T. R., & Noice, H. (1988). Influence of positive affect upon creative thinking and problem solving in children. *Psychological Reports, 63*, 895-898.

Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology, 73*, 407-420.

Gulz, A., & Haake, M. (2006). Design of animated pedagogical agents - A look at their look. *Man-Machine Studies, 64*(4), 322-339.

Han, S., Lee, S., & Jo, G. (2005). Case-based tutoring systems for procedural problem solving on the www. *Expert Systems with Applications*, *29*(3), 573-582.

Heffernan, N. T (2001). *Intelligent tutoring systems have forgotten the tutor: Adding a cognitive model of human tutors.* PhD Thesis, Carnegie Mellon University, Pittsburgh, Pennsylvania.

Hill, W. E. (1997). *Learning: A survey of psychological interpretations*. New York, New York: Addison-Wesley Educational Publishers.

Hwang, G. (2003). A conceptual map model for developing intelligent tutoring systems. *Computers and Education, 40*(3), 217-235.

Isen, A. M. (2000). Some perspectives on positive affect and self-regulation. *Psychological Inquiry, 11*(3), 184-187.

Isen, A. M., & Daubman, K. A. (1986). The influence of positive affect on the perceived organization of components of self. Unpublished manuscript, University of Maryland, Catonsville, Maryland.

Isen, A. M., Daubman, K. A., & Nowicki, G. P. (1987). Positive affect facilitates creative problem solving. *Journal of Personality and Social Psychology, 52*, 1122-1131.

Isen, A. M., & Reeve, J. (2005). The influence of positive affect on intrinsic and extrinsic motivation: Facilitating enjoyment of play, responsible work behavior, and self-control. *Motivation and Emotion, 29*(4), 295-323.

Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *Artificial Intelligence in Education, 11*, 47-78.

Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. In *Automatic Face and Gesture Recognition*. Southampton, England.

Kasabov, N. (1996). *Foundations of neural networks, fuzzy systems and knowledge engineering*. Cambridge, Massachusetts: MIT Press.

Khan, T., & Yip, Y. J. (1995). Case-based task management for computer-aided learning. In *Progress in Case-Based Reasoning*. Salford, England.

Kleinginna, P. R. J., & Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion, 5*(4), 345-379.

Kort, B., & Reilly, R. (2002). Analytical models of emotions, learning and relationships: Towards an affect-sensitive cognitive machine. In *Virtual Worlds and Simulation*. San Antonio, Texas.

Kort, B., Reilly, R., & Picard, R.W. (2001). An affective model of interplay between emotions and learning: Reengineering educational pedagogy - building a learning companion. In *Advanced Learning Technologies*. Madison, Wisconsin.

Koton, P. (1988). Reasoning about evidence in causal explanation. In *Artificial Intelligence*. Saint Paul, Minnesota.

Lepper, M. R., Aspinwall, L., Mumme, D., & Chabay, R. W. (1990). Self-perception and social perception processes in tutoring: Subtle social control strategies of expert tutors. In J. M. Olson & M. P. Zanna (Eds.). *Self-inference and social inference: The Ontario symposium*. Hillsdale, New Jersey: Erlbaum.

Lepper, M. R., & Chabay, R. W. (1988). Socializing the intelligent tutor: Bringing empathy to computer tutors. In H. Mandl & A. M. Lesgold (Eds.), *Learning issues for intelligent tutoring systems*. Chicago, Illinois: Springer-Verlag.

Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997). The persona effect: affective impact of animated pedagogical agents. In *Human Factors in Computing Systems*. Atlanta, Georgia.

Lester, J. C., Stone, B. A., & Stellin, G. D. (1999). Lifelike pedagogical agents for mixed-initiative problem solving in constructivist learning environments. *User Modeling and User-Adapted Interaction, 9*, 1-44.

Lester, J. C., Voerman, J. L., Towns, S. G., & Callaway, C. B. (1999). Deictic believability: Coordinating gesture, locomotion, and speech in lifelike pedagogical agents. *Applied Artificial Intelligence, 13*, 383-414.

Lester, J. C., Zettlemoyer, L. S., Gregoire, J., & Bares, W. H. (1999). Explanatory lifelike avatars: Performing user-designed tasks in 3D learning environments. In *Autonomous Agents*. Seattle, Washington.

Lisetti, C. (1999). Modeling cognition-emotion of users for improved interaction with software systems. In *User Modeling*. Banff, Canada.

Litman, D. J., & Forbes-Riley, K. (2006). Recognizing student emotions and attitudes on the basis of utterances in spoken tutoring dialogues with both human and computer tutors. *Speech Communication, 48*(5), 559-590.

Lucey, S., Matthews, I., Hu, C., Ambadar, Z., De la Torre Frade, F., & Cohn, J. (2006). AAM derived face representations for robust facial action recognition. In *Automatic Face and Gesture Recognition*. Southampton, England.

Luger, G. F. (2002). *Artificial intelligence: Structures and strategies for complex problem solving* (4th ed.). Essex, England: Pearson Education Limited.

Massaro, D. W. (2003). A computer-animated tutor for spoken and written language learning. In *Multimodal Interfaces*. Vancouver, Canada.

McCauley, L., Gholson, B., Hu, X., Graesser, A. C., & The Tutoring Research Group (1998). Delivering smooth tutorial dialogue using a talking head. In *Embodied Conversation Characters*. Tahoe City, California.

Mehrabian, A. (1971). *Silent messages*. Belmont, California: Wadsworth.

Merrill, D. C., Reiser, B. J., Ranney, M., & Trafton, J. G. (1992). Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems. *Learning Sciences, 2*(3), 277-30.

Mota, S., & Picard, R. W. (2003). Automated posture analysis for detecting learner's interest level. In *Computer Vision and Pattern Recognition for Human-Computer Interaction*. Madison, Wisconsin.

van Mulken, S., André, E., & Muller, J. (1998). The persona effect: How substantial is it? In *Human Computer Interaction*. Berlin, Germany.

Murray, T. (1999). Authoring intelligent tutoring systems: An analysis of the state of the art. *Artificial Intelligence in Education, 10*, 98-129.

New Zealand Ministry of Education (2003). *Book 1. The Number Framework*. Wellington, New Zealand: Ministry of Education.

Noma, T., & Badler, N. (1997). A virtual human presenter. In *Animated Interface Agents*. Nagoya, Japan.

Ortony, A., Clore, G., & Collins, A. (1998). *Cognitive Structure of Emotions*. Cambridge, England: Cambridge University Press.

Pal, S. K., & Shiu, S. C. K. (2004). *Foundations of soft case-based reasoning*. Hoboken, New Jersey: John Wiley & Sons.

Pantic, M., & Rothkrantz, J. M. (2003). Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE, 91*(9), 1370-1390.

Parkinson, B. (1995). Emotion. In B. Parkinson & A. M. Colman (Eds.). *Emotion and Motivation*. London, England: Longman Essential Psychology.

Person, N. K., Graesser, A. C., Kreuz, R. J., Pomeroy, V., & The Tutoring Research Group (2001). Simulating human tutor dialog moves in AutoTutor. *Artificial Intelligence in Education, 12*, 23-29.

Person, N. K., Graesser, A. C., & The Tutoring Research Group (2003). Fourteen facts about human tutoring: Food for thought for ITS developers. In *Tutorial Dialogue Systems: With a View Toward the Classroom*. Sydney, Australia.

Picard, R. W. (1997). *Affective computing*. Cambridge, Massachusetts: MIT Press.

Picard, R. W. (1998). Towards agents that recognize emotion. In *IMAGINA*, Monaco, Monaco.

Prendinger, H., & Ishizuka, M. (Eds.) (2004). *Life-like characters. Tools, affective functions, and applications*. Cognitive technologies. Heidelberg, Germany: Springer Verlag.

Prendinger, H., & Ishizuka, M. (2007). Symmetric multi-modality revisited: Unveiling users' physiological activity. *Industrial Electronics, 54*(2), 692-698.

Prendinger, H., Mayer, S., Mori, J., & Ishizuka, M. (2003). Persona effect revisited. Using bio-signals to measure and reflect the impact of character-based interfaces. In *Intelligent Virtual Agents*. Irsee, Germany.

Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television and new media like real people and places*. Cambridge, England: Cambridge University Press.

Riesbeck, C. K., & Schank, R. C. (1991). From training to teaching: Techniques for case-based ITS. In H. Burns, J. W. Parlett, & C. L. Redfield, (Eds.). *Intelligent Tutoring Systems: Evolutions in Design*. Hillsdale, New Jersey: Lawrence Erlbaum.

Rotter, N. G., & Rotter, G. S. (1988). Sex differences in the encoding and decoding of negative facial emotions. *Journal of Nonverbal Behavior, 12,* 139-148.

Sarrafzadeh, A., Fan, C., Dadgostar, F., Alexander, S. T. V., & Messom, C. (2004). Frown gives game away: Affect sensitive tutoring systems for elementary mathematics. In *Systems, Man and Cybernetics*. The Hague, Netherlands.

Schank, R. C., & Edelson, D. J. (1989). A role for AI in education: using technology to reshape education. *Artificial Intelligence in Education, 1*(2), 3-20.

Saver, J. L., & Damasio, A. R. (1991). Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage. *Neuropsychologia, 29*(12), 1241-1249.

Scherer, K. R. (1981). Speech and emotional states. In J. K. Larby (Ed.). *Speech evaluation in psychiatry*. New York, New York: Grune and Stratton.

Self, J. (1999). The defining characteristics of intelligent tutoring systems research: ITSs care, precisely. *Artificial Intelligence in Education*, *10*, 350-364.

Shafran, I., Riley, M., & Mohri, M. (2003). Voice signatures. In *Automatic Speech Recognition and Understanding*. Virgin Islands, United States.

Shaw, E., Johnson, W. L. & Ganeshan, R. (1999). Pedagogical agents on the web. In *Autonomous Agents*. Seattle, Washington.

Shiri, M., Aïmeur, E., & Frasson, C. (1998). Case-based student modelling: An accessible solution model. In *Nouvelles Technologies de la Communication et de la Formation*. Rouen, France.

Simonite, T. (2007). *Emotion-aware teaching software tracks student attention*. Retrieved January 9, 2007, from http://www.newscientisttech.com/article.ns?id=dn10894&feedId=online-news_rss20

Sloman, A. (1991). Prolegomena to a theory of communication and affect. In A. Ortony, J. Slack, & O. Stock (Eds.). *AI and cognitive science perspectives on communication*. Heidelberg, Germany: Springer Verlag.

del Soldato, T. (1994). Motivation in tutoring systems. Technical Report CSRP 303, The University of Sussex, England.

Soh, L. (2006). Incorporating an intelligent tutoring system into CS1. In *Technical Symposium on Computer Science Education*. Houston, Texas.

Strongman, K.T. (Ed.) (2003). *The psychology of emotion: From everyday life to theory* (5th ed.). West Sussex, England: John Wiley & Sons Ltd.

Thayer, J., & Johnsen, B. H. (2000). Sex differences in judgement of facial affect: A multivariate analysis of recognition errors. *Scandinavian Journal of Psychology, 41*(3), 243-246.

Tsaganou, G., Grigoriadou, M., & Cavoura, T. (2002). Modelling student's comprehension of historical text using fuzzy case-based reasoning. In *European Conference on Case Based Reasoning*. Aberdeen, Scotland.

Urban-Lurain, M. (2003). *An historic review in the context of the development of artificial intelligence and educational psychology*. Retrieved June 24, 2003, from http://www.cse.msu.edu/rgroups/cse101/ITS/its.htm

VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction, 21*(3), 209-249.

de Vicente, A.(2003). Towards tutoring systems that detect students' motivation: An investigation. Ph.D. Thesis, School of Informatics, University of Edinburgh.

Xu, D., Wang, H., & Su, K. (2002). Intelligent student profiling with fuzzy models. In *Annual Hawaii International Conference on System Sciences*. Big Island, Hawaii.

Zadeh, L. A. (1973). Outline of a new approach to the analysis of complex systems and decision processes. *Systems, Man, and Cybernetics, SMC-3*, 28-44.

Appendix A

Composite turns in the observational study

This appendix gives the definitions and frequencies in the data of the composite tutor and student turns that were used in the coding of the videos of the observational study, as described in Section 3.2.5, as well as the frequencies of single tutor and student turns. There were 112 new kinds of tutor turn, and 6 new kinds of student turn, as shown below in Tables A.1 and A.2 respectively. For example, Table A.1 shows that the new tutor turn, 29, is a composite of the tutor turn 1 followed by the tutor turn 5; the codes for the tutor and student turns that make up the composite turns are given in Section 3.2.4.

Table A.1. Definitions of composite tutor turns, and frequencies of tutor turns.

Turn	Sub-turns (for composite turns only)	Number of occurrences
1		2
2		11
3		2
4		8
5		80
6		4
7		41
8		176
9		0
10		0
11		23
12		0
13		17

14		2
15		1
16		1
17		2
18		52
19		3
20		9
21		4
22		28
23		6
24		0
25		4
26		3
27		0
28		15
29	1, 5	4
30	11, 8	55
31	11, 5	61
32	13, 22	2
33	13, 5	50
34	11, 21, 5	6
35	11, 4, 5	15
36	21, 5	12
37	13, 8	137
38	8, 8	1
39	23, 8	18
40	11, 23, 5	9
41	11, 2, 5	15
42	13, 23, 8	24
43	23, 5	8
44	11, 2, 21, 5	2
45	11, 4, 21, 5	1
46	22, 19, 8	1
47	18, 22	2
48	11, 7	3
49	11, 22	2
50	11, 17	1
51	13, 23, 5	7
52	13, 2, 5	9
53	11, 4	17
54	13, 2	4
55	2, 5	7
56	8, 11, 4	1
57	5, 23, 8	1
58	23, 11, 5	1
59	11, 23	2
60	23, 11, 4, 5	1
61	13, 4	6
62	23, 11, 4	1
63	25, 8	2
64	8, 11	1
65	7, 8	1

66	11, 19, 11, 4, 5	1
67	25, 2, 5	2
68	25, 5	3
69	7, 5	3
70	11, 22, 5	1
71	13, 4, 5	3
72	13, 21, 5	2
73	11, 23, 8	16
74	8, 23, 8	1
75	13, 23	7
76	19, 23, 8	2
77	19, 2, 17, 5	1
78	11, 2	7
79	28, 5	1
80	8, 18	1
81	11, 23, 11, 28	1
82	28, 28	1
83	28, 4	2
84	23, 2	1
85	11, 25, 3	1
86	24, 4	1
87	15, 19	1
88	15, 22	1
89	11, 3	1
90	19, 8	1
91	23, 11, 3	1
92	13, 7	4
93	13, 22, 8	1
94	13, 23, 11, 2, 5	1
95	11, 8, 8	1
96	18, 5	1
97	17, 22	1
98	13, 11, 4, 5	1
99	13, 21	1
100	22, 18	2
101	11, 1, 5	2
102	11, 5, 11, 23, 4, 5	1
103	5, 11	2
104	11, 23, 25, 4, 5	1
105	11, 4, 11, 4, 5	1
106	13, 23, 21, 5	1
107	11, 23, 4, 5	2
108	7, 11, 23, 11, 4, 5	1
109	18, 11	1
110	21, 11, 21, 5, 11, 23, 8	1
111	23, 4, 5	1
112	21, 8	2
113	23, 12, 5	1
114	7, 18	1
115	13, 18	1
116	13, 6	1
117	11, 8, 11	1

118	18, 22, 18	2
119	5, 2, 13	1
120	25, 21, 5	2
121	22, 5	1
122	11, 5, 3, 13, 7	1
123	8, 5	1
124	13, 19, 11, 4	1
125	20, 22	1
126	11, 21	1
127	5, 17, 5	1
128	13, 17, 8	1
129	19, 26	1
130	19, 23, 4	1
131	28, 4, 5	1
132	23, 4	1
133	2, 26	1
134	13, 26	1
135	18, 8	1
136	23, 8, 19, 8	1
137	13, 23, 11, 21, 5	1
138	11, 21, 8	1
139	4, 5	1
140	11, 23, 2, 5	1

Table A.2. Definitions of composite student turns, and frequencies of student turns.

Turn	Sub-turns (for composite turns only)	Number of occurrences
1		812
2		23
3		105
4		70
5		13
6		0
7		18
8		4
9		0
10		7
11		19
12		0
13		5
14		0
15		16
16	1, 8	1
17	1, 1	2
18	5, 1	1
19	1, 7	1
20	1, 3	1
21	11, 7	1

Appendix B

Questionnaires used in the study of Easy with Eve

The following two pages show the questionnaires that were given to participants in the study of Easy with Eve following their interaction with the tutoring system.

The first questionnaire shown is the questionnaire that was given to students in the two experimental groups where the animated agent was present; the second questionnaire shown is the questionnaire that was given to the two other experimental groups where the animated agent was not present. Question 5 was omitted from the second questionnaire because it was not relevant to participants who had not interacted with the animated agent.

Some questions about Easy with Eve

1. Did you enjoy using Easy with Eve today? (please tick one box)

☐	☐	☐	☐	☐
It was cool!	*It was good*	*It was ok*	*Not really*	*Not at all*

2. Would you like to use Easy with Eve again?

☐	☐	☐	☐	☐
A lot!	*Quite a bit*	*Maybe*	*Not really*	*Not at all*

3. How much do you think you learned today?

☐	☐	☐	☐	☐
Lots!	*Quite a bit*	*A little*	*Not much*	*Nothing*

4. Remember how the camera on the screen only works half of the time – do you think it was working today?

☐	☐	☐	☐	☐
Yes	*I think so*	*Maybe*	*I don't think so*	*No*

5. Did Eve look real?

☐	☐	☐	☐	☐
Lots!	*Quite a bit*	*A little*	*Not really*	*Not at all*

Some questions about Easy with Eve

1. Did you enjoy using Easy with Eve today? (please tick one box)

☐	☐	☐	☐	☐
It was cool!	*It was good*	*It was ok*	*Not really*	*Not at all*

2. Would you like to use Easy with Eve again?

☐	☐	☐	☐	☐
A lot!	*Quite a bit*	*Maybe*	*Not really*	*Not at all*

3. How much do you think you learned today?

☐	☐	☐	☐	☐
Lots!	*Quite a bit*	*A little*	*Not much*	*Nothing*

4. Remember how the camera on the screen only works half of the time – do you think it was working today?

☐	☐	☐	☐	☐
Yes	*I think so*	*Maybe*	*I don't think so*	*No*

Appendix C

Transcript of instructions in the study of Easy with Eve

This appendix gives a transcription of the instructions that were given to participants to help them complete the pre- and post-tests in the study of Easy with Eve. Instructions were given at three stages: before the pre-test; after the pre-test and before the tutoring session; and after the tutoring session and before the post-test. There were two slightly different versions of these instructions, depending on whether spoken (in the animated agent present experimental groups) or text-based instructions were given.

Appendix C.1 Animated agent instructions

<u>Before the pre-test</u>

“Hi there! Welcome to Easy with Eve. My name’s Eve, and today we’re going to be doing some fun counting and adding on the computer with tens frames and counters. In a minute we’re going to start looking at some problems together, and you’re going to have to use the keyboard to type in some number answers, and you’ll need the mouse to click enter. Did you know that I’m special? Because sometimes I can see your facial expressions from the camera on the screen! But half of the time the camera doesn’t work, so we won’t know for sure whether I can see you or not. But just before we get started, we need to do a quick quiz so that we can tell how much learning we do today.

I’ll be back in a minute – don’t worry if the quiz is hard, because we haven’t even started yet! Remember to use the keyboard and the mouse to enter your answers.”

Before the tutoring session

“Cool, well, that’s enough – how did you find that? Now we can get started properly, and work through some problems together. Let’s start with this one.”

Before the post-test

“Oh well, I’m afraid that’s all the time we’ve got for doing problems today – well done, you did really well. But we’re not finished quite yet – first we need you to do just one more short quiz like the one we did at the beginning. Try and use what we’ve learnt today to help you answer the questions. That’s all from me, but the quiz will start in a couple of seconds, and then that’ll be all for now. Hope to see you again soon – bye!”

Appendix C.2 Text-based instructions

Before the pre-test

“Hi there! Welcome to Easy with Eve. Today we’re going to be doing some fun adding with counters. You’re going to use the keyboard to type in number answers, and use the mouse to click enter. Did you know that this program is special? Because sometimes it can see your face from the camera! But half of the time the camera doesn’t work, so we don’t know for sure if it can see you. But before we get started, first we need to do a quick quiz. Don’t worry if the quiz is hard, because it gets easier later. Remember to use the keyboard and the mouse.”

Before the tutoring session

“Cool, now we can start properly, and work through some problems. Let’s start with this one.”

Before the post-test

"Oh well, that's all the time we've got for doing problems today – well done, you did really well. But we're not finished yet – first we need you to do one more quiz. Try and use what you learnt to help answer the questions. The quiz will start soon – bye!"

Appendix D

Raw data from the study of Easy with Eve

This appendix gives the raw data for the study of Easy with Eve that was described in Chapter 6. Table D.1 below shows the gender, pre-test score, post-test score and responses to the five questionnaire questions for each of the participants in the four experimental groups. As explained in Section 6.1.1, these four experimental groups were as follows:

– in Group 1, the animated agent was present and facial expressions were detected (as a measure of student emotion to be used by the tutoring strategies module);

– in Group 2, the animated agent was not present (text-based feedback was used) but facial expressions were detected;

– in Group 3, the animated agent was present but facial expressions were not detected; and

– in Group 4, neither the animated agent was present or facial expressions detected.

As explained in Section 6.2, results for three of the participants were invalid.

Table D.1. The raw data from the study of Easy with Eve.

ID	Group	Gender	Pretest / 15	Posttest / 15	Q1 / 5	Q2 / 5	Q3 / 5	Q4 / 5	Q5 / 5
0	2	F	9	7	4	4	4	5	
1	2	M	9	11	5	5	3	4	
2	2	M	8	7	5	5	4	4	
3	2	F	5	8	4	5	4	5	
4	2	F	10	11	5	5	5	4	
5	2	F	6	7	5	5	4	5	
6	2	F	11	13	4	3	3	5	
7	2	M	9	7	5	5	5	3	
8	2	F	9	12	5	3	5	4	
9	2	M	11	12	5	5	4	5	
10	2	F	6	12	5	5	5	4	
11	2	M	10	12	5	4	5	4	
12	2	F	7	12	5	4	3	2	
13	2	M	10	9	5	4	5	3	
14	2	M	11	10	4	3	5	3	
15	2	F	5	9	5	3	5	4	
16	4	M	11	12	5	5	4	5	
17	4	F	6	6	5	5	5	4	
18	4	M	7	8	5	5	5	5	
19	4	M	8	10	5	5	5	5	
20	4	F	8	7	5	5	5	5	
21	4	F	8	10	5	5	5	5	
22	4	M	10	12	5	3	4	4	
23	4	M	8	9	4	3	3	1	
24	4	F	12	14	5	5	5	3	
25	4	F	13	8	4	5	4	2	
26	4	F	7	9	4	4	4	3	
27	4	F	9	9	5	5	5	5	
28	4	M	10	10	5	4	4	5	
29	4	M	7	6	5	5	5	5	
30	4	F	8	12	4	3	4	3	
31	3	F	9	9	3	4	4	3	3
32	3	M	6	9	4	3	3	4	4
33	3	M	6	8	4	5	5	4	3
34	3	F	8	6	5	5	3	5	5
35	3	F	5	8	5	4	5	4	3
36	3	M	10	7	3	3	2	3	1
37	3	M	invalid						
38	3	M	10	9	5	4	4	4	4
39	3	F	10	8	4	4	3	4	4
40	3	F	7	9	3	4	4	4	3
41	3	F	7	7	4	5	4	3	4
42	3	M	10	11	3	2	1	4	1
43	3	F	7	9	5	5	5	4	5
44	3	M	10	15	5	4	4	4	4
45	1	F	7	8	5	5	4	4	3

46	1	M	11	10	5	3	5	5	5
47	1	M	invalid						
48	1	M	7	6	5	5	5	5	4
49	1	F	8	7	5	4	4	4	4
50	1	M	4	12	4	4	5	4	2
51	1	F	5	10	4	5	4	5	3
52	1	F	8	7	5	5	5	5	5
53	1	M	9	10	5	5	4	4	3
54	1	F	11	12	5	5	5	5	4
55	1	M	10	12	3	3	4	1	1
56	1	F	11	8	5	5	5	5	3
57	1	M	9	10	3	4	5	4	3
58	1	F	4	5	5	5	5	5	5
59	1	M	15	10	5	5	3	3	1
60	1	M	invalid						
61	1	F	5	8	5	5	5	5	5

Appendix E

Extensible structure of Easy with Eve

This appendix details how the modular structure of Easy with Eve facilitates future extension. As shown below in Figure E.1, the ATS can be extended in three possible areas: detecting student emotions, domain content, and data for the tutoring strategies module.

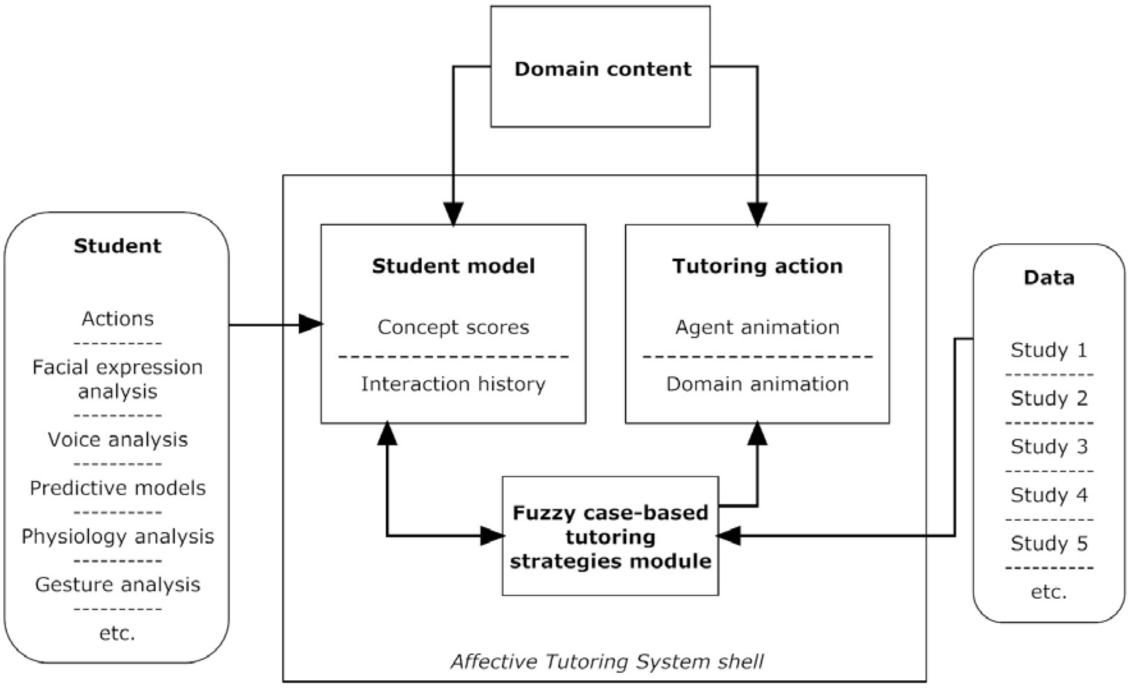


Figure E.1. Extensible structure of Easy with Eve, showing how new emotion detection methods, content domain and data sets for the tutoring strategies module can be added to the ATS.

Firstly, any new methods for detecting student emotion – such as voice analysis, predictive models, physiology analysis or gesture analysis – will be able to add their output as a plug-in to the student model. This information will be added to the existing capability of the system to detect student facial expressions, thus allowing the student model to make a better informed decision about the affective state of the student. A new method would need to be developed in the student model for weighing the evidence generated by the different (and potentially contradictory) streams of affective input.

Secondly, new domain content will allow the ATS to be tailored to different subjects; this will inform both the agent animations and the domain animations of the system, as well as inform the student model about the accuracy of student answers to questions. For instance, the tutoring strategies module merely outputs a set of tutoring actions to be carried out; what these tutoring actions are actually comprised of is defined by the domain content module, which can be replaced or adapted as required for new ATSs in different domains.

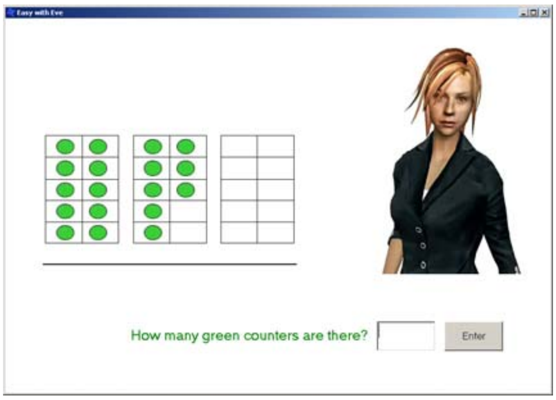
Finally, new data from future studies of human tutors (similar to the observational study that was discussed in Chapter 3) can also be added to the case-base for use by the fuzzy case-based tutoring strategies module. A blackboard architecture would be used whereby the data from the most relevant study would be given precedence, yet all of the data could be accessed by the tutoring strategies module.

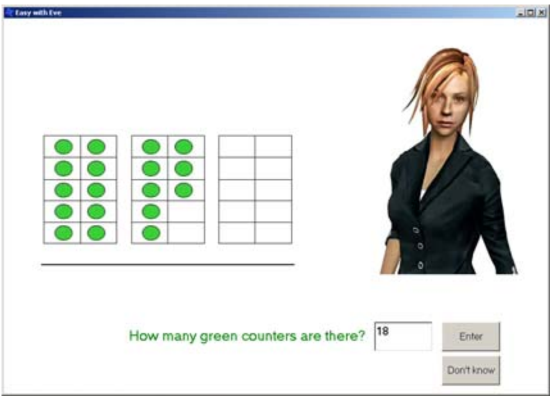
The rest of the ATS shell (the student model, tutoring actions and tutoring strategies module) can remain more or less unchanged, regardless of any new emotion detection methods, domain content or data sets for the case-base that may or may not be plugged in. The design of the student model, the range of actions of Eve, and the structure of the tutoring strategies module would remain consistent.

Appendix F

User interaction with Easy with Eve

This appendix demonstrates a typical user interaction with Easy with Eve as a student follows through one of the problems at Level 1. In this problem the student will add 18 green counters to 9 yellow counters using the tens frames shown below. The dialogue of Eve, as well as a brief description, is given at each step.



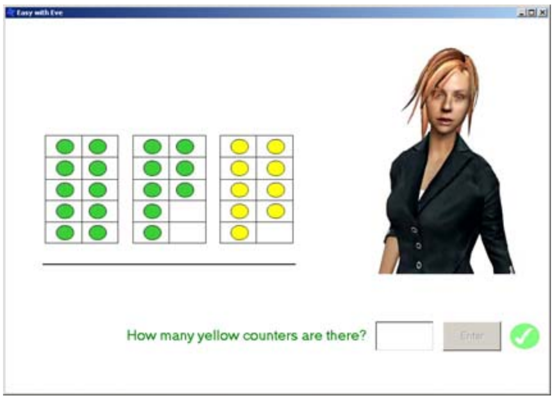


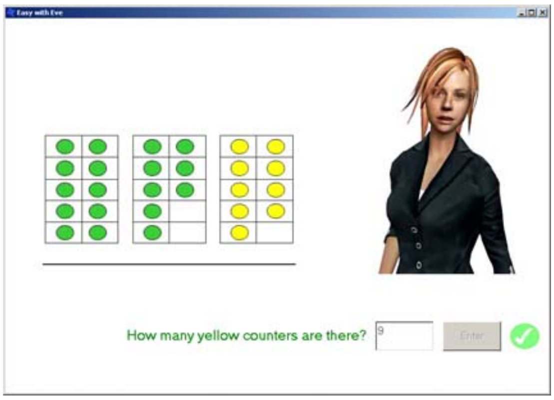
Eve: “How many green counters are there?”

Student: 18

Eve: “Good work”

At this stage, the student is asked to count the number of green counters. Note that after a period of time, a “Don’t know” button appears to below the “Enter” button. The student answers correctly, and so Eve moves on to the next question.



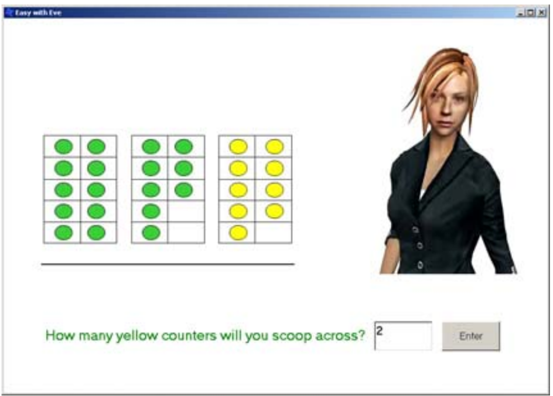


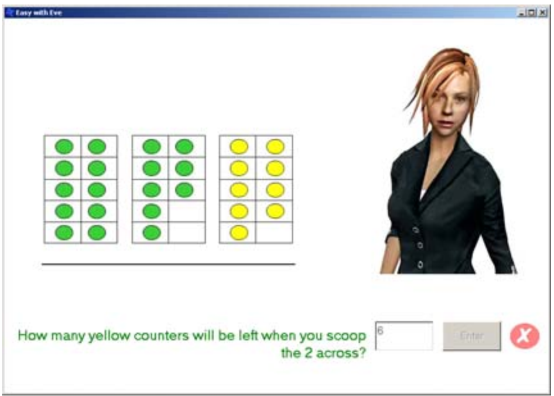
Eve: “And how many yellow counters can you see?”

Student: 9

Eve: “That’s good”

A green tick temporarily appears to indicate that the student answered the previous question correctly, and then Eve asks the second question. The student answers this question correctly also.





Eve: “How many yellow counters do you think you'll scoop across to the greens?”

Student: 2

Eve: “Yup…”

Eve: “And how many yellow counters will be left when you scoop the yellows across?”

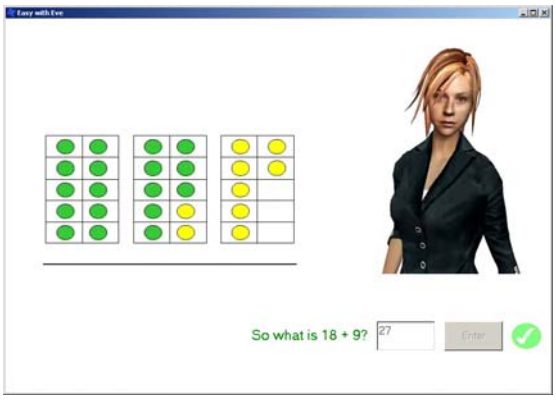
Student: 6

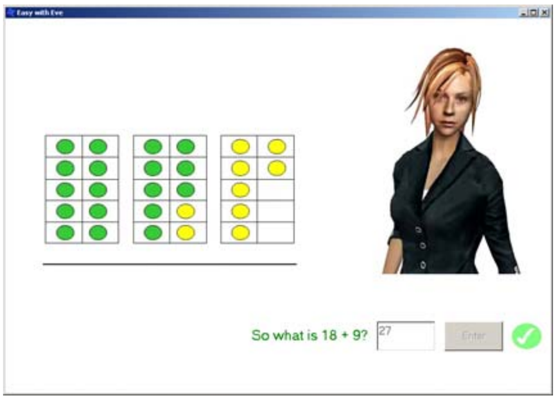
Eve: “What I mean is, how many yellow counters will that leave when you scoop the yellows across?”

Student: 7

Eve: “Hmm...”

Now the student is asked two more questions: the student initially answers the second question incorrectly, and so a red cross appears as Eve rearticulates the question. The student then answers correctly.





Eve: “So what is 18 + 9?”

Student: 27

Eve: “OK...”

Eve: “Let's go on to another problem”

Eve then asks the student to add the two numbers together (with the counters as a visual aid). The student answers correctly, and thus completes the problem; because the student’s score is not sufficient to move on to the next level, another problem at Level 1 will be selected and the cycle will repeat.

Appendix G

Ethics documentation

This appendix gives the ethics documentation for the the observational study of human tutors described in Chapter 3 and the study of the effectiveness of Easy with Eve described in Chapter 6. The following documents are presented:

- Massey University Human Ethics Committee application form for the observational study of human tutors.

- Supporting documentation for the observational study application: parent and student information and consent forms.

- Letter of ethical approval for the observational study of human tutors.

- Massey University Human Ethics Committee low risk application form for the study of the effectiveness of Easy with Eve.

- Parent and student information and consent forms for the study of the effectiveness of Easy with Eve.

- Letter of ethical approval for the study of the effectiveness of Easy with Eve.

Massey University
Te Kunenga ki Pūrehuroa

Application No: _____ / __________
This number is assigned when your application is accepted.
Quote on all documentation to participants and the Committee.

Human Ethics Committee

APPLICATION FOR APPROVAL OF PROPOSED RESEARCH/TEACHING/EVALUATION INVOLVING HUMAN PARTICIPANTS

(All applications are to be typed and presented using language that is free from jargon and comprehensible to lay people)

SECTION A

1. **Project Title** Human Tutor Adaptation to Non-verbal Behaviour of Students

   Projected start date 10/11/04 **Projected end date** 30/11/04

2. **Applicant Details** *(Select the appropriate box and complete details)*

STUDENT APPLICATION

Full Name of Student Applicant	Samuel Thomas Vaughan Alexander
Employer (if applicable)	N/A
Telephone	ext 9257
Email Address	S.T.Alexander@massey.ac.nz
Postal Address	IIMS, Albany Campus
Full Name of Supervisor 1	Abdolhossein Sarrafzadeh
School/Department/Institute	IIMS
Region *(mark one only)*	**Albany** [X] **Palmerston North** [ ] **Wellington** [ ]
Telephone	9549
Email Address	H.A.Sarrafzadeh@massey.ac.nz
Full Name of Supervisor 2	Stephen Hill
School/Department/Institute	School of Psychology
Region *(mark one only)*	**Albany** [ ] **Palmerston North** [X] **Wellington** [ ]
Telephone	7566
Email Address	S.R.Hill@massey.ac.nz

3. **Type of Project** *(mark one only)*

Staff Research [ ]

Student Research:
- **PhD Research** [X]
- **Master's Research** [ ]
- **Honours Research** [ ]
- **Undergraduate Research** *(individual project)* [ ]

- **Evaluation Programme** [ ]
- **Undergraduate Teaching Programme** [ ]
- **Other** [ ]

If Other, specify:

4. **Summary of Project**
Please outline in no more than 200 words in lay language why you have chosen this project, what you intend to do and the methods you will use.
(Note: all the information provided in the application is potentially available if a request is made under the Official Information Act. In the event that a request is made, the University, in the first instance, would endeavour to satisfy that request by providing this summary. Please ensure that the language used is comprehensible to all)

The overall aim of my PhD is to develop an animated computer tutor that can adapt to the non-verbal behaviour of students.

The aim of this particular study is to find out how competent human tutors adapt to the non-verbal behaviour of students in one-to-one tutoring situations. The results of this study will be used as a basis for the design of the animated computer tutor.

My PhD is part of a wider research group, Next Generation Intelligent Tutoring Systems (NGITS – http://ngits.massey.ac.nz), that is aimed at developing a new generation of human like intelligent teaching machines. The group includes two other PhD students, Chao Fan and Farhad Dadgostar, who are developing systems that detect facial expressions and gestures in images taken by a web-cam. These systems will be used by my animated computer tutor to interpret the non-verbal behaviour of students in real time.

Ways in which a competent human tutor may adapt to non-verbal behaviour of students will be identified in this study by videoing a professional tutor(s) tutoring primary school students in a one-on-one situation. A coding scheme will be used to collect data from the videos, which will be used to draw conclusions about how human tutors adapt to the non-verbal behaviour of students.

5. **List of Attachments** (tick boxes)

Completed "Screening Questionnaire to Determine the Approval Procedure" (compulsory)	X	Advertisement	
Information Sheet/s *(indicate how many)*	3	Health Checklist	
Translated copies of Information Sheet/s		Questionnaire	
Consent Form/s *(indicate how many)*	4	Interview Schedule	
Translated copies of Consent Form/s		Evidence of Consultation	
Transcriber Confidentiality Agreement		Letter requesting access to an institution	X
Confidentiality Agreement *(for persons other than the researcher / participants who have access to project data)*		Letter requesting approval for use of database	
Authority for Release of Tape Transcripts			

Applications that are incomplete or lacking the appropriate signatures will be returned to the applicant for completion. This could mean delays for the project.

Please refer to the Human Ethics website (http://humanethics.massey.ac.nz) for details of where to submit your application and the number of copies required.

SECTION B: PROJECT INFORMATION

General

6 **I/we wish the protocol to be heard in a closed meeting (Part II).** **Yes** [ ] **No** [X]

(If yes, state the reason in a covering letter)

7 **Does this project have any links to other approved Massey University Human Ethics Committee application/s?** **Yes** [ ] **No** [X]

If yes, list HEC protocol number/s and relationship/s.

8 **Is approval from other Ethics Committees being sought for the project?** **Yes** [ ] **No** [X]

If yes, list the other Ethics Committees.

9 **For staff research, is the applicant the only researcher?** **Yes** [ ] **No** [ ]

If no, list the names and addresses of all members of the research team.

N/A

Project Details

10 **State concisely the aims of the project.**

To find out how a competent human tutor(s) adapts to the non-verbal behaviour of students in real time.

11 **Give a brief background to the project to place it in perspective and to allow the project's significance to be assessed.** *(No more than 200 words in lay language)*

Non-verbal behaviour is a significant part of human interaction, thus students convey rich information to tutors about their cognitive and motivational state through their facial expressions and gestures. Competent human tutors use this information to adapt their tutoring to the precise needs of individual students.

Computer tutoring systems aim to adapt to students as effectively as human tutors. However, no tutoring system has yet been developed that adapts to the non-verbal behaviour of students. Given that adapting to non-verbal behaviour is an important part of human tutoring, then perhaps a tutoring system that can adapt to non-verbal behaviour will be more effective than existing artificial tutoring systems.

Finding out ways in which a human tutor adapts to the non-verbal behaviour of students will provide a foundation needed to develop artificial tutors that can adapt in the same ways. The results of this study will be used to help develop an animated tutor that will be able to adapt its tutoring to the non-verbal behaviour of students, as detected by real time image processing systems already under development at Massey.

12 **Outline the research procedures to be used, including approach/procedures for collecting data. Use a flow chart if necessary.**

1. At least 1 (up to 3) professional tutors, headed by Carol Mayers of *TopNotch Maths Ltd*, will tutor participants one-on-one for about 20 minutes.
2. The subject of tuition will be a maths exercise already used in schools, designed by the New Zealand Numeracy Project (part of the Ministry of Education).
3. Both the participants and the tutor will be videoed.
4. A coding scheme will be used to draw data from the videos about how the tutor(s) adapts to the non-verbal behaviour of participants.

13 **Where will the project be conducted? Include information about the physical location/setting. If the study is based overseas, specify which countries are involved.**

The project will be conducted at a local primary school, during school hours.

14 **What experience does the researcher/s have in this type of project activity?**

I have no experience in this type of research. I will draw upon the experience of my colleagues, especially my psychology supervisor, Stephen Hill.

Participants

15 **Describe the intended participants.**

8-9 year old students at a local primary school.

16 **How many participants will be involved?**

About 10.

What is the reason for selecting this number?

(Where relevant, attach a copy of the Statistical Justification to the application form)

10 students will guarantee as much data as I will need – this should provide over 3 hours of video. From viewing existing New Zealand Numeracy Project videos we estimate that students will be reasonably expressive – 3 hours of video should comfortably provide as much data as I'll need.

17 **Describe how potential participants will be identified and recruited?**

Teachers at the local school will be asked to identify students that are at the right level for the tutoring exercise to be effective. These students will then be recruited by sending information sheets and consent forms home to parents. Students will receive separate information sheets and consent forms.

Consent will be required from both the parents/guardians of the students, and the students themselves. Of consenting students at the right level, whose parents also consent, about 10 students will be randomly chosen.

18 **Does the project involve recruitment through advertising?** **Yes** [ ] **No** [X]

(If yes, attach a copy of the advertisement to the application form)

19 **Does the project require permission of an organisation (e.g. a school or a business) to access participants or information?** **Yes** [X] **No** [ ]

(If yes, attach a copy of the request letter/s, e.g. letter to Board of Trustees/Principal, CEO etc to the application form. Note that some educational institutions may require the researcher to submit a Police Security Clearance)

20 **Who will make the initial approach to potential participants?**

Teachers at the local school. A separate information sheet for teachers is attached.

21 **Describe criteria (if used) to select participants from the pool of potential participants.**

Participants will have to be at the right level of maths ability for the tutoring exercise to be effective.

22 **How much time will participants have to give to the project?**

About half an hour. The exact timing of the study will be negotiated with teachers.

Data Collection

23	**Does the project include the use of participant questionnaire/s?** *(If yes, attach a copy of the Questionnaire/s to the application form)*	**Yes** [ ]	**No** [X]
	If yes: i) will the participants be anonymous?	**Yes** [ ]	**No** [ ]

ii) describe how the questionnaire will be distributed and collected.

(If distributing electronically through Massey IT, attach a copy of the request letter to the Director, Information Technology Services to the application form)

....

24 Does the project include the use of focus group/s? **Yes** [ ] **No** [X]

(If yes, attach a copy of the Confidentiality Agreement for the focus group to the application form)

25 Does the project include the use of participant interview/s? **Yes** [ ] **No** [X]

(If yes, attach a copy of the Interview Questions/Schedule to the application form)

26 Does the project involve audiotaping? **Yes** [ ] **No** [X]

27 Does the project involve videotaping? **Yes** [X] **No** [ ]

(If agreement for taping is optional for participation, ensure there is explicit consent on the Consent Form)

If yes, state what will happen to the tapes at the completion of the project.

(e.g .destroyed, returned, stored by the researcher, archived in an official archive)

The tapes will be securely stored by my supervisor for 5 years.

28 If audiotaping is used, will the tape be transcribed? **Yes** [ ] **No** [X]

If yes, state who will do the transcribing.

(If not the researcher, a Transcriber's Confidentiality Agreement is required – attach a copy to the application form. Normally, transcripts of interviews should be provided to participants for editing, therefore an Authority For Use Of Participants' Tape is required – attach a copy to the application form. However, if the researcher considers that the right of the participant to edit is inappropriate, a justification should be provided below)

....

29 Does the project require permission to access databases? **Yes** [ ] **No** [X]

(If yes, attach a copy of the request letter/s to the application form)

30 Who will carry out the data collection?

Just me, although I may seek assistance from my main supervisor, Abdolhossein Sarrafzadeh, or my co-supervisor, Stephen Hill.

SECTION C: BENEFITS / RISK OF HARM TO PARTICIPANTS

31 What are the possible benefits (if any) of the project to the participants?

Each participant will receive a free 20 minute maths lesson with a professional tutor, covering material that is part of the New Zealand Numeracy Project.

32 **What discomfort (physical, psychological, social), incapacity or other risk of harm are participants likely to experience as a result of participation?**

(Consider the risk of harm to individuals and also to groups/communities and institutions to which they belong)

None. Students will only miss about half an hour of class time, and this will be negotiated with teachers to ensure that participants are not disadvantaged. Also, there will be no coercion for parents to enrol their children in an external tutoring company.

33 **Describe the strategies the researcher will use to deal with any of the situations identified in Q32.**

N/A

34 **What is the risk of harm (if any) of the project to:**

i) **Researcher/s**

None.

ii) **Any other persons/groups/organisations affected by the research.**

None.

35 **How do you propose to manage the risk of harm for points i) and ii) above?**

N/A

36 **Is ethnicity data being collected as part of the project?** **Yes** [ ] **No** [X]

(Note that harm can be done through an analysis based on insufficient numbers)

If yes: i) will the data be used as a basis for analysis? **Yes** [ ] **No** [ ]

ii) justify this use in terms of the number of participants.

37 **If participants are children/students in a pre-school/school/tertiary setting, describe the arrangements you will make for children/students who are not taking part in the research.**

(Note that no child/student should be disadvantaged through the research)

Participants will be called one at a time from class – all other students will continue as usual. The exact timing of the study will be negotiated with teachers.

SECTION D: INFORMED AND VOLUNTARY CONSENT

38 **By whom and how, will information about the research be given to participants?**

An information sheet will be sent home to the parents of students by teachers at the local school. Potential participants will receive a separate information sheet from their teacher. Teachers will also receive a separate information sheet.

39 **Will consent to participate be given in writing?** **Yes** [X] **No** [ ]

(Attach copies of Consent Form/s to the application form)

If no, justify the use of oral consent.

40 **Will participants include persons under the age of 16?** **Yes** [X] **No** [ ]

If yes, indicate the age group and competency for giving consent.

(Note that parental/caregiver consent for school-based research may be required by the school even when children are competent. Ensure Information Sheets and Consent Forms are in a style and language appropriate for the age group)

The participants will be aged 8 to 9. They and their parents will be clearly informed by the information sheets and well able to give consent.

41 **Will participants include persons who are vulnerable or whose capacity to give informed consent may be compromised?** **Yes** [ ] **No** [X]

If yes, describe the consent process you will use.

42 **Will the participants be proficient in English?** **Yes** [X] **No** [ ]

If no, all documentation for participants (Information Sheets/Consent Forms/Questionnaire etc) must be translated into the participants' first-language

(Attach copies of the translated Information Sheet/Consent Form etc to the application form)

SECTION E: PRIVACY/CONFIDENTIALITY ISSUES

43 **Will information about participants be obtained from third parties?** **Yes** [ ] **No** [X]

If yes, describe how and from whom.

44 **Will any identifiable information on the participants be given to third parties?** **Yes** [ ] **No** [X]

If yes, describe how.

45 **Will the participants be anonymous (i.e. their identity unknown to the researcher?)** **Yes** [ ] **No** [X]

If no: i) will the participants be given a unique identifier? **Yes** [X] **No** [ ]

ii) will the participants' identity be disclosed in publication of the research? **Yes** [ ] **No** [X]

46 **Will an institution (e.g. school) to which participants belong be named or be able to be identified?** **Yes** [X] **No** [ ]

(Ensure that institutions have been informed of this in your request to access them)

47 **Outline how and where the data (including tapes/transcripts) and Consent Forms will be stored.**

(Note that Consent Forms should be stored separately from data)

The video tapes and consent forms will be kept in separate lockable drawers in my supervisor Hossein Sarrafzadeh's office (currently QA2.41).

48 i) **Who will have access to the data/Consent Forms?**

Only my supervisor Hossein Sarrafzadeh and I will have access to the consent forms.

The following will have access to the videos:
Hossein Sarrafzadeh – my supervisor
Stephen Hill and Tanja Mitrovic – my co-supervisors
Farhad Dadgostar and Chao Fan – PhD students in NGITS (and other postgraduate students that may join the NGITS research group)
Peter Hughes – research associate from the New Zealand Numeracy Project
Carol Mayers – head tutor involved in the study (and any other tutors that are part of the immediate study)

My supervisors will have access to the videos to help me with gathering data from the videos.

Peter Hughes and Carol Mayers will only be able to watch the videos with me, to offer expert feedback on the tutoring strategies used by tutors in the videos.

Farhad Dadgostar and Chao Fan will also have access to the videos to test their facial expression and gesture analysis systems.

ii) **How will the data/Consent Forms be protected from unauthorised access?**

The data and consent forms will be kept secure in my supervisor's office.

49 **Who will be responsible for disposal of the data/Consent Forms when the five-year storage period is up?**

(The Massey University HOD Institute/School/Section / Supervisor / or nominee should be responsible for the eventual disposal of data)

My supervisor.

50 **Will participants be given the option of having the data (particularly tapes) transferred to an official archive?** *(This option may apply when data collected is of historical significance)* **Yes** [ ] **No** [X]

(If yes, include this option in the Consent Form)

51 **Will participants be given the option of having their tapes returned to them?** *(If yes, include this option in the Consent Form)* **Yes** [ ] **No** [X]

SECTION F: DECEPTION

52 **Is deception involved at any stage of the project?** **Yes** [ ] **No** [X]

If yes, justify its use and describe the debriefing procedures.

SECTION G: CONFLICT OF INTEREST

53 **Is the project to be funded in any way from sources external to Massey University?** **Yes** [ ] **No** [X]

If yes: i) state the source.

(The tutors will be paid by an NGITS research group fund)

ii) **does the source of the funding present any conflict of interest with regard to the research topic?**

54 **Does the researcher/s have a financial interest in the outcome of the project?** Yes [ ] No [X]

If yes, explain how the conflict of interest situation will be dealt with.

55 **Is there any professional or other relationship (e.g. employer/employee, lecturer/student, practitioner/patient, researcher/family member) to the researcher?** Yes [ ] No [X]

If yes, describe the relationship and indicate how the resulting conflict of interest situation will be dealt with.

Carol Mayers works for a tutoring company, TopNotch Maths Ltd. However, there will be no coercion for parents to enrol their children in TopNotch Maths, or any other tutoring company. The name of her company will not even be given to parents or students (unless they ask for it).

Also, Carol Mayers is not part of the NGITS research team – she is just a tutor employed for the purposes of this study only.

SECTION H: COMPENSATION TO PARTICIPANTS

56 **Will any payments or other compensation be given to participants?** Yes [ ] No [X]

If yes, describe what, how and why.

(Note that compensation (if provided) should be given to all participants and not constitute an inducement. Details of any compensation provided must be included in the Information Sheet)

SECTION I: TREATY OF WAITANGI

57 **Does the proposed research impact on Maori persons as Maori?** Yes [ ] No [X]

If yes describe how.

58 **Are Maori the primary focus of the project?** Yes [ ] No [X]

(If yes, complete Section I, otherwise proceed to Question 63)

63 **If Maori are not the focus of the project, outline what Maori involvement there may be and how this will be managed.**

I am aware of the need to uphold the Treaty of Waitangi.

I am aware that some of the participants may be of Maori descent, or belong to other ethnic groups from all over the world. In such an eventuality, I will seek advice from the school concerning any cultural issues that may potentially arise.

No ethnicity data will be collected.

SECTION J: OTHER CULTURAL ISSUES

64 **Are there any aspects of the project that might raise specific cultural issues, other than those covered in Section I?** **Yes** [ ] **No** [X]

If yes, explain. Otherwise, proceed to Section K.

Please see Question 63.

SECTION K: SHARING RESEARCH FINDINGS

72 **Describe how information resulting from the project will be shared with participants.**

Participants and their parent/guardians will be able to access the overall results of the project on the research group website (http://ngits.massey.ac.nz). The details of this website will be given on all of the information sheets and consent forms.

Participants and their parent/guardians will also be given the contact details of myself and my supervisor on the information sheets, and invited to contact us if they wish to learn anything more about the project.

SECTION L: INVASIVE PROCEDURES/PHYSIOLOGICAL TESTS

73 **Does the project involve the collection of tissues, blood, other body fluids or physiological tests?** **Yes** [ ] **No** [X]

(If yes, complete Section L, otherwise proceed to Section M)

SECTION M: DECLARATION *(Complete appropriate box)*

STUDENT RESEARCH

Declaration for Student Applicant

I have read the Code of Ethical Conduct for Research, Teaching and Evaluations involving Human Participants and discussed the ethical analysis with my Supervisor. I understand my obligations and the rights of the participants. I agree to undertake the research as set out in the Code of Ethical Conduct for Research, Teaching and Evaluations involving Human Participants.
The information contained in this application is to the very best of my knowledge accurate and not misleading.

Student Applicant's Signature .... Date: ....

Declaration for Supervisor

I have assisted the student in the ethical analysis of this project. As supervisor of this research I will ensure that the research is carried out according to the Code of Ethical Conduct for Research, Teaching and Evaluations involving Human Participants.

Supervisor's Signature .... Date: ....

Print Name ....

Declaration for Head of Department/School/Institute

I declare that to the best of my knowledge, this application complies with the Code of Ethical Conduct for Research, Teaching and Evaluations involving Human Participants and that I have approved its content and agreed that it can be submitted.

Head of Dept/School/Inst Signature .... Date: ....

Print Name ....

Massey University
COLLEGE OF SCIENCES

INSTITUTE OF INFORMATION & MATHEMATICS SCIENCES
Private Bag 102 904
NSMC
Auckland
New Zealand
T 64 9 414 0800
F 64 9 441 8136
www.massey.ac.nz

Human Tutor Adaptation to Non-verbal Behaviour of Students

PARENT/GUARDIAN INFORMATION SHEET

My name is Sam Alexander, I am a computer science PhD student at Massey University studying intelligent computer tutoring systems. My supervisor is Dr Hossein Sarrafzadeh, also of Massey University. You can contact either of us by writing to:

> Institute of Information and Mathematical Sciences
> Private Bag 102 904
> NSMC
> Auckland

Alternatively, you can call me at (09) 414 0800 ext. 9257, or Hossein at (09) 414 0800 ext. 9549, or email S.T.Alexander@massey.ac.nz or H.A.Sarrafzadeh@massey.ac.nz.

The aim of this information sheet is to inform you of a study soon to be undertaken at Murrays Bay School, in which your child is invited to participate.

I am a member of the Next Generation Intelligent Tutoring Systems research group based at Massey University in Albany. Our aim is to develop a new generation of human like intelligent teaching machines. More information is available at our website: http://ngits.massey.ac.nz.

The overall aim of my research is to develop an animated tutor for a tutoring system that can adapt to the non-verbal behaviour (i.e. facial expressions and gestures) of students, as detected by real time image processing software already under development at Massey.

However, to know how to adapt to non-verbal behaviour, I first need to study how human tutors adapt to the non-verbal behaviour of students, which is where I need your child's help.

With your permission, I hope that your child will be able to participate in a study investigating the adaptation of human tutors to non-verbal behaviour. The details of this study are as follows:

- A team of professional tutors will tutor about 10 participants one-on-one for about 20 minutes.
- The subject of tuition will be a maths exercise already used in schools, as it was designed by the New Zealand Numeracy Project (part of the Ministry of Education), with whom I collaborate.
- Both the participants and the tutor will be videoed.
- I will then draw data from the videos about how tutors adapt to the non-verbal behaviour of participants.
- The project will be conducted on school grounds, during school hours.
- And your child gets a free maths lesson with a professional tutor!

Potential participants have been selected solely on the basis of their maths ability, as they have to be at just the right level for the tutoring exercise to be most relevant. Your child has also been given a separate information sheet and consent form. Of the consenting potential participants whose parent/guardians also give consent, ten students will be randomly chosen to take part in the study. Ten participants should be more than enough to generate the data that I need from the current study.

The privacy of participants will be closely protected. The name of your child will not be published in any form. No image of your child will be published in any form unless specific consent is given by both you and your child. This consent would not be sought until a later date, after the videos have been analysed.

The consent forms and video tapes will be stored in a locked drawer in my supervisor's office for a period of five (5) years. At the end of this period, both the consent forms and the video tapes will be destroyed.

The results of this study will be used by the Next Generation Intelligent Tutoring Systems research group to help develop an exciting new kind of computer tutoring system, that can adapt to the non-verbal behaviour of students. This system should be ready sometime around Term 2, 2005, and you and your child will be most welcome to trial the system.

You will be able to access the results of this study at any time by contacting either myself or Dr Hossein Sarrafzadeh, or by visiting our website: http://ngits.massey.ac.nz.

You are under no obligation to accept this invitation. If you decide to allow your child to participate, you and your child have the right to:

- decline to answer any particular question;
- withdraw from the study within one week following the tutoring session;
- ask any questions about the study at any time during participation;
- provide information on the understanding that your child's name will not be used unless you give permission to the researcher;
- be given access to a summary of the project findings when it is concluded.

Your child also has the right to ask for the video tape to be turned off at any time during the tutoring session.

Please feel free to contact either myself or my supervisor if you have any questions at all about the project.

This project has been reviewed and approved by the Massey University Human Ethics Committee, ALB Application 04/082. If you have any concerns about the conduct of this research, please contact Associate Professor Kerry Chamberlain, the Chair of the Massey University Campus Human Ethics Committee at Albany. You can call him at (09) 414 0800 ext. 9078, or email humanethicsalb@massey.ac.nz.

Massey University
COLLEGE OF SCIENCES

Institute of Information and Mathematical Sciences
Te Kura Pūtaiao o Mōhiohio me Pāngarau

Private Bag 102 904,
North Shore MSC,
Auckland,
New Zealand
Telephone: 64 9 441 8160
Facsimile: 64 9 441 8136

Human Tutor Adaptation to Non-verbal Behaviour of Students

PARENT/GUARDIAN CONSENT FORM

This consent form will be held for a period of five (5) years

I have read the Information Sheet and have had the details of the study explained to me. My questions have been answered to my satisfaction, and I understand that I may ask further questions at any time.

I agree to the tutoring session being video taped.

I agree to allow my child ....
to participate in this study under the conditions set out in the Information Sheet.

Signature: .... **Date:** ....

Full Name - printed ....

If you would like to receive a summary of the findings of this study, please indicate in the space below how you would like to receive the summary (e.g. email, postal address).

Contact details: ....

....

....

Massey University
COLLEGE OF SCIENCES

INSTITUTE OF INFORMATION
& MATHEMATICS SCIENCES
Private Bag 102 904
NSMC
Auckland
New Zealand
T 64 9 414 0800
F 64 9 441 8136
www.massey.ac.nz

How do teachers teach?

Hi there - my name is Sam Alexander, and I'm a computer science PhD student at Massey University.

I'd like to learn about how teachers do their job, so I can try and make a teacher on a computer do the same things as real teachers.

To do this I need to see how teachers help you learn. I'd like to video a special teacher helping you by yourself with your maths, and then I'll be able to see what makes teachers good at their job. It will take about 30 minutes of your normal school day, and you won't have to miss lunch or stay after school.

Only 10 students can be chosen for the videoing, so please don't be sad if you miss out.

If you want to know more, you can ask your teacher or your parents/guardian. Also, why not look at our website? It's http://ngits.massey.ac.nz.

If you want to be videoed, don't forget to fill out the permission slip!

Also, please don't forget to give your parents/guardian the other sheets that your teacher gave you.

Thanks for your help!

Massey University
COLLEGE OF SCIENCES

Institute of Information and Mathematical Sciences
Te Kura Pūtaiao o Mōhiohio me Pāngarau

Private Bag 102 904,
North Shore MSC,
Auckland,
New Zealand
Telephone: 64 9 441 8160
Facsimile: 64 9 441 8136

Human Tutor Adaptation to Non-verbal Behaviour of Students

STUDENT CONSENT FORM

This consent form will be held for a period of five (5) years

I have read the Information Sheet and have had the details of the study explained to me. My questions have been answered well, and I know that I can ask more questions at any time.

I agree to the tutoring session being video taped.

I agree to take part in this study.

Signature: .... **Date:** ....

Full Name - printed ....

Massey University
AUCKLAND

OFFICE OF THE
DEPUTY VICE-CHANCELLOR - AUCKLAND
Private Bag 102 904
North Shore MSC
Auckland
New Zealand
T Deputy Vice-Chancellor - Auckland
64 9 414 0800 extn 9517
Regional Registrar - Auckland
64 9 414 0800 extn 9516
F 64 9 414 0814
www.massey.ac.nz

29 October 2004

Sam Alexander
C/- Dr H Sarrafzadeh
College of Science
Massey University
Albany

Dear Sam

HUMAN ETHICS APPROVAL APPLICATION – MUAHEC 04/082
"Human Tutor Adaption to Non-Verbal Behaviour of Students"

Thank you for your application. It has been fully considered and approved by the Massey University, Albany Campus, Human Ethics Committee.

If you make any significant departure from the Application as approved then you should return this project to the Human Ethics Committee, Albany Campus, for further consideration and approval.

Approval is for three years. If this project has not been completed within three years from the date of this letter, a new application must be submitted at that time.

Yours sincerely

Associate-Professor Kerry Chamberlain
Chairperson,
Human Ethics Committee
Albany Campus

cc Dr H Sarrafzadeh
College of Science

Massey University
Te Kunenga ki Pūrehuroa

NOTIFICATION OF LOW RISK RESEARCH/EVALUATION INVOLVING HUMAN PARTICIPANTS

(All notifications are to be typed)

SECTION A:

1. Project Title Testing an Affective Tutoring System for Addition in Schools

Projected start date for data collection 1 December 2006 **Projected end date** 15 December 2006

2. Applicant Details *(Select the appropriate box and complete details)*

STUDENT NOTIFICATION

Full Name of Student Applicant Samuel T V Alexander

Postal Address IIMS, Albany

Telephone Ext. 9257 **Email Address** s.t.alexander@massey.ac.nz

Employer (if applicable) N/A

Full Name of Supervisor(s) Hossein Sarrafzadeh

School/Department/Institute Computer Science, IIMS

Region *(mark one only)* **Albany** [X] **Palmerston North** [ ] **Wellington** [ ]

Telephone 9549 **Email Address** h.a.sarrafzadeh@massey.ac.nz

3. Type of Project *(mark one only)*

Staff Research/Evaluation:		**Student Research:**	X	**If other, please specify:**
Academic Staff		**Qualification**	X	PhD
General Staff		**Points Value of Research**		

4. Describe the peer review process used in assessing the ethical issues present in this project.

This project follows on from a previous study for which ethical approval was gained: Massey University Human Ethics Committee, ALB Application 04/082.

The project has been discussed with my supervisors.

5. **Summary of Project**
Please outline in no more than 200 words in lay language why you have chosen this project, what you intend to do and the methods you will use.
(Note: all the information provided in the notification is potentially available if a request is made under the Official Information Act. In the event that a request is made, the University, in the first instance, would endeavour to satisfy that request by providing this summary. Please ensure that the language used is comprehensible to all)

The aim of this project is to test an Affective Tutoring System (teaching software) for addition in local schools. I would like to test 60 eight/nine year old students, so will go to as many schools as this takes.

Students will use the system one-on-one for about 20 minutes, as well as completing a pre- and post-test. The system involves answering questions about counting and addition. The system also uses a web-cam that can detect facial expressions of students; this will be explained to parents and students, and the video from the web-cam will not be stored without consent.

There will be 4 different experimental groups, based on whether the system features an animated tutor and whether or not the information from the web-cam is being accessed.

Parental and student consent will be asked for before carrying out the study. Parents and students will be asked for extra consent so that the tutoring session can be videoed.

Please submit this Low Risk Notification (with the completed Screening Questionnaire) to:

The Ethics Administrator
Research Ethics Office
Old Main Building, PN221
Massey University
Private Bag 11 222
Palmerston North

SECTION B: DECLARATION *(Complete appropriate box)*

STUDENT RESEARCH
Declaration for Student Applicant
I have read the Code of Ethical Conduct for Research, Teaching and Evaluations involving Human Participants and discussed the ethical analysis with my Supervisor. I understand my obligations and the rights of the participants. I agree to undertake the research as set out in the Code of Ethical Conduct for Research, Teaching and Evaluations involving Human Participants. The information contained in this notification is to the very best of my knowledge accurate and not misleading.

Student Applicant's Signature .... Date: ....

Declaration for Supervisor
I have assisted the student in the ethical analysis of this project. As supervisor of this research I will ensure that the research is carried out according to the Code of Ethical Conduct for Research, Teaching and Evaluations involving Human Participants.

Supervisor's Signature .... Date: ....

Print Name ....

Massey University
COLLEGE OF SCIENCES

INSTITUTE OF INFORMATION & MATHEMATICAL SCIENCES
Te Kura Pūtaiao o Mōhiohio me Pāngarau
Private Bag 102 904
NSMC
Auckland
New Zealand
T 64 9 441 8160
F 64 9 441 8136
http://iims.massey.ac.nz

Testing a Tutoring System that can Adapt to Facial Expressions

PARENT/GUARDIAN INFORMATION SHEET

Your child is invited to participate in a study at Torbay Primary School to test a world-first in tutoring systems technology: a tutoring system can that adapt in real time to the facial expressions of students. The aim of this information sheet is to inform you of this study, and to seek permission for your child to take part.

My name is Sam Alexander: I am a computer science PhD student at Massey University studying intelligent computer tutoring systems. My supervisor is Dr Hossein Sarrafzadeh, also at Massey University. You can contact either of us by writing to:

Institute of Information and Mathematical Sciences
Private Bag 102 904
NSMC
Auckland

Alternatively, you can call me at (09) 414 0800 ext. 9257, or Hossein at (09) 414 0800 ext. 9549, or email S.T.Alexander@massey.ac.nz or H.A.Sarrafzadeh@massey.ac.nz.

Our research team, Next Generation Intelligent Tutoring System, has developed an animated tutoring character called Eve for a maths system, *Easy with Eve*. Eve can adapt to the facial expressions of students, as detected by real time image processing software developed at Massey. The domain of the tutoring system is counting and addition using tens frames and counters; this is one of the exercises that has been developed by the New Zealand Numeracy Project. More information is available at: http://ngits.massey.ac.nz.

With your permission, I hope that your child will be able to participate in a study investigating the effectiveness of this system. The details of this study are as follows:

- The study will be conducted on school grounds, during school hours.
- Your child will work with the system for a total of 30 minutes, including a 5 minute pre-test and a 5 minute post-test.
- The subject of tuition will be a New Zealand Numeracy Project maths exercise.
- Students will be randomly divided into 4 different test groups, based on whether or not facial expressions are detected and whether or not feedback is text-based or via Eve.
- The facial expressions of half of the students will be analysed by taking images from a web-cam, but these images are not stored. Any data recorded by the system is completely anonymous.
- If you consent, your child will also be videoed (separately to the web-cam). However, this is completely optional, and there is a box for this on the attached consent form.
- And your child gets a free maths lesson with exciting new software!

Your child has also been given a separate information sheet and consent form; your child must also give their consent to participate in the study.

The privacy of participants will be closely protected. The name of your child will not be published in any form. No image of your child will be published in any form unless specific consent is given by both you and your child. This consent would not be sought until a later date, after the videos have been analysed.

The consent forms and video tapes will be stored in a locked drawer in my supervisor's office for a period of five (5) years. At the end of this period, both the consent forms and the video tapes will be destroyed.

The results of this study will be used by the Next Generation Intelligent Tutoring Systems research group to help evaluate the effectiveness of the tutoring system, and to pinpoint particular areas for improvement. You will be able to access the results of this study at any time by contacting either myself or Dr Hossein Sarrafzadeh, or by visiting our website: http://ngits.massey.ac.nz.

You are under no obligation to accept this invitation. If you decide to allow your child to participate, you and your child have the right to:

- decline to answer any particular question;
- withdraw from the study within one week following the tutoring session;
- ask any questions about the study at any time during participation;
- provide information on the understanding that your child's name will not be used unless you give permission to the researcher;
- be given access to a summary of the project findings when it is concluded.

If you consent to your child being videoed, your child also has the right to ask for the video tape to be turned off at any time during the tutoring session.

Please feel free to contact either myself or my supervisor if you have any questions at all about the project.

This project has been evaluated by peer review and judged to be low risk. Consequently, it has not been reviewed by one of the University's Human Ethics Committees. The researchers named above are responsible for the ethical conduct of this research.

If you have any concerns about the conduct of this research that you wish to raise with someone other than the researchers, please contact Professor Sylvia Rumball, Assistant to the Vice-Chancellor (Ethics & Equity), telephone 06 350 5249, email humanethics@massey.ac.nz.

Massey University
COLLEGE OF SCIENCES

INSTITUTE OF INFORMATION
& MATHEMATICAL SCIENCES
Te Kura Pūtaiao o
Mōhiohio me Pāngarau
Private Bag 102 904
NSMC
Auckland
New Zealand
T 64 9 441 8160
F 64 9 441 8136
http://iims.massey.ac.nz

Testing a Tutoring System that can Adapt to Facial Expressions

PARENT/GUARDIAN CONSENT FORM

This consent form will be held for a period of five (5) years

I have read the Information Sheet and have had the details of the study explained to me. My questions have been answered to my satisfaction, and I understand that I may ask further questions at any time.

I agree to allow my child .... to participate in this study under the conditions set out in the Information Sheet.

I agree to the tutoring session being video taped (optional): ☐ YES ☐ NO

Signature: .... **Date:** ....

Full Name - printed ....

If you would like to receive a summary of the findings of this study, please indicate in the space below how you would like to receive the summary (e.g. email, postal address).

Contact details: ....

....

....

Massey University
COLLEGE OF SCIENCES

INSTITUTE OF INFORMATION
& MATHEMATICAL SCIENCES
Te Kura Pūtaiao o
Mōhiohio me Pāngarau
Private Bag 102 904
NSMC
Auckland
New Zealand
T 64 9 441 8160
F 64 9 441 8136
http://iims.massey.ac.nz

How do teachers teach?

Hi there - my name is Sam Alexander, and I'm a computer science PhD student at Massey University.

I've made a fun computer program called 'Easy with Eve' that helps you learn how to add. What's special about Eve the tutor is that sometimes she can *see* how you're feeling!

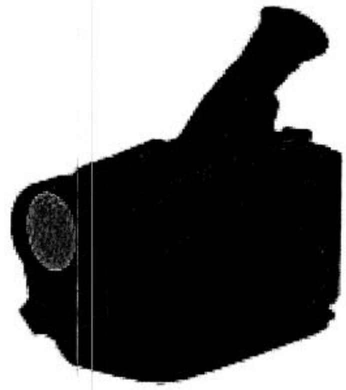
To see how good Eve is at teaching, I'd like you to have a go with the computer program. It will take about 30 minutes of your normal school day, and you won't have to miss lunch or stay after school. If it's ok with you I'd like to video you using the program, but you don't have to be videoed if you don't want to.

If you want to know more, you can ask your teacher or your parents/guardian. Also, why not look at our website? It's http://ngits.massey.ac.nz.

If you want to have a go with 'Easy with Eve', don't forget to fill out the permission slip!

Also, please don't forget to give your parents/guardian the other sheets that your teacher gave you.

Thanks for your help!



Massey University
COLLEGE OF SCIENCES

INSTITUTE OF INFORMATION
& MATHEMATICAL SCIENCES
Te Kura Pūtaiao o
Mōhiohio me Pāngarau
Private Bag 102 904
NSMC
Auckland
New Zealand
T 64 9 441 8160
F 64 9 441 8136
http://iims.massey.ac.nz

Testing a Tutoring System that can Adapt to Facial Expressions

STUDENT CONSENT FORM

This consent form will be held for a period of five (5) years

I have read the Information Sheet and have had the details of the study explained to me. My questions have been answered well, and I know that I can ask more questions at any time.

I agree to take part in this study.

I agree to the tutoring session being video taped (optional): ☐ YES ☐ NO

Signature: .... **Date:** ....

Full Name - printed ....

Massey University

OFFICE OF THE ASSISTANT TO THE VICE-CHANCELLOR
(Ethics & Equity)
Private Bag 11 222
Palmerston North
New Zealand
T 64 6 350 5573/350 5575
F 64 6 350 5622
humanethics@massey.ac.nz
animalethics@massey.ac.nz
gtc@massey.ac.nz
www.massey.ac.nz

8th December 2006

Samuel Alexander
Institute of Information and Mathematical Sciences
Albany

Dear Samuel

Re: Testing an Affective Tutoring System for Addition in Schools

Thank you for the documents received on 1 December 2006. These have been assessed as meeting the requirements of a Low Risk Notification. For future reference, please note that the full MUHEC application does not have to be provided for a Low Risk Notification. The relevant documentation is available for downloading from the MUHEC website.

Your project has been recorded on the Low Risk Database which is reported in the Annual Report of the Massey University Human Ethics Committees.

Please notify me if situations subsequently occur which cause you to reconsider your initial ethical analysis that it is safe to proceed without approval by one of the University's Human Ethics Committees.

A reminder to include the following statement on all public documents:

> *"This project has been evaluated by peer review and judged to be low risk. Consequently, it has not been reviewed by one of the University's Human Ethics Committees. The researcher(s) named above are responsible for the ethical conduct of this research.*
>
> *If you have any concerns about the conduct of this research that you wish to raise with someone other than the researcher(s), please contact Professor Sylvia Rumball, Assistant to the Vice-Chancellor (Ethics & Equity), telephone 06 350 5249, e-mail humanethics@massey.ac.nz".*

Please note that if a sponsoring organisation, funding authority or a journal in which you wish to publish requires evidence of committee approval (with an approval number), you will have to provide a full application to one of the University's Human Ethics Committees. You should also note that such an approval can only be provided prior to the commencement of the research.

Yours sincerely

Sylvia Rumball

Sylvia V Rumball (Professor)
Chair, Human Ethics Chairs' Committee and
Assistant to the Vice-Chancellor (Ethics & Equity)

cc HoS Prof Robert McKibbin
Institute of Information and Mathematical Sciences, Albany