Effective Tutoring with Empathic Embodied Conversational Agents

Thesis

"Effective Tutoring with Empathic Embodied Conversational Agents"

Thesis submitted for the degree of

Doctor of Philosophy

By

Sharon G. Moyo

M.Sc. Computing for Commerce and Industry

B. Business Studies and Computing Science

Centre for Research in Computing

The Open University

September 2013

Abstract

This thesis examines the prospect of using empathy in an Embodied Tutoring System (ETS) that guides students through an online quiz (by providing feedback on student answers and responding to self-reported student emotion). The ETS seeks to imitate human behaviours successfully used in one-to-one human tutorial interactions. The main hypothesis is that the interaction with an empathic ETS results in greater learning gains than a neutral ETS, primarily by encouraging positive and reducing negative student emotions using empathic feedback.

In a preparatory study we investigated different strategies for expressing emotion by the ETS. We established that a multimodal strategy achieves the best results regarding how accurately human participants can recognise the emotions. This approach was used in developing the feedback strategy for our empathic ETS.

The preparatory study was followed by two studies in which we compared a neutral with an empathic ETS. The ETS in the second of these studies was developed using results from the first of these studies. In both studies, we found no statistically significant difference in learning gains between the neutral and empathic ETS. However, we did discover a number of interactions between the ETS system, learning gains and, in particular 1) student scores on an empathic tendency test and 2) student ability. We also analysed the subjective responses and the relation between self-reported emotions during the quiz and student learning gains.

Based on our studies in a formal class room setting, we assess the prospects of using empathic agents in a classroom setting and describe a number of requirements for their effective use.

Declaration

The studies outlined in this thesis were undertaken at the Deanes School, Essex, United Kingdom with the express permission of the Deanes School Governing Committee and the Open University Ethics Committee and supervised by Dr, Paul Piwek, Dr. Neil Smith and Dr. Sandra Williams. I declare that all of the work was done by the author except where otherwise stated. This thesis has not been submitted at any other university.

Sharon Moyo

Milton Keynes, UK, September 2013

Acknowledgements

I would like to express my sincere gratitude to my supervisors: Dr. Paul Piwek, Dr. Neil Smith and Dr. Sandra Williams for their unfailing support, invaluable insight, encouragement and supervision throughout my PhD research. They have made this a wonderful journey of discovery, by challenging me through discussion, suggestions and sharing their knowledge and expertise.

Thanks go to my previous supervisors: Prof. Donia Scott and Prof. Josie Taylor for their supervision and support during the early part of my research. Their assistance was important in defining my research goals.

I am grateful to the Open University for giving me the opportunity to complete my research on a part-time basis. In particular, thanks go to the staff at the research school, the statistical advisory unit, the CRC student forum and the CRC student conferences for providing invaluable support at various key points on my journey to completing my studies.

My thanks go to my colleagues at the Natural Language Generation (NLG) Group for their invaluable comments, suggestions and support. I am grateful for their insightful comments and probing questions which helped me to develop my research.

Special thanks go to the Head teacher Jan Atkinson, staff, students and other members of the Deanes School learning community for their support and involvement in the user evaluations throughout my thesis activities. I am especially grateful to my colleagues in the IT department and IT support for their technical support and patience throughout my research.

I am especially indebted to my parents, Johnson and Sithembile Menziwa-Moyo for their unfailing love and support of every aspect of my life. I thank my siblings, Bianca and Alfred Menziwa-Moyo for keeping me going with their steadfast love and encouragement. Although there are too many to mention, I would also like to extend my sincere thanks to my extended family, friends and wider church community who provided a close support network throughout my studies.

Finally, I would like to thank my husband, Vusumuzi Sibanda for his love, support and motivation throughout my studies.

"I can do all things through Christ who strengthens me." (Philippians 4:13 NKJV)

Contents

List of Tables

List of Figures

Chapter 1 – Introduction

The focus of the current thesis is to improve learning gains in a formal, classroom-based learning environment. Learning is defined by the Oxford dictionary as "to acquire knowledge of or skill in (something) by study, experience, or being taught". Learning *gain* specifically refers to the difference between a student's knowledge before and after some intervention. For this purpose, a tutoring system has been developed that provides affective (or emotional) feedback to students in order to encourage positive emotion and alleviate negative emotion, thereby improving learning gains. The system uses empathic behaviour to model the system's response to student answers and progress. Research into successful human-to-human tutorial strategies shows that tutors provide task based and emotional feedback, suggesting a link between emotion and learning that tutors effectively use to encourage student progress. The challenge for an affective tutoring system is to assess student emotion and respond appropriately to improve learning gains.

Previous research has been building on the success of Intelligent Tutoring Systems (Graesser et al., 1999; Lane and Van Lehn 2005). However, these systems only provide task-based (cognitive) feedback. Therefore these tutorial strategies would need to be successfully combined with an effective emotional strategy such as empathy to assess and respond to student emotion (or affect) in the learning environment. To enhance the expression of empathy, the combined feedback can be presented to the user via an Embodied Conversational Agent (ECA), an animated character with human-like characteristics (Cassell et al., 1994), using natural language generation and artificial intelligence techniques. Therefore, the focus of this thesis is the role of empathy within such an Embodied Tutoring System (ETS) in providing appropriate task-based and emotional feedback, and the impact on student learning gains within formal learning environments such as classrooms.

Before studying learning gains, we investigated the impact of a multimodal strategy in providing recognisable ECA emotions to students. Following on from this we then concentrated on implementing an empathic feedback strategy based on Davis' (1994) theory of empathy within an ETS. The feedback strategy is modelled on one-to-one human tutoring and evaluated with students in a classroom environment to explore the effects on learning gain.

In the remainder of this chapter we proceed as follows: In Section 1.1 we describe in detail the motivation for the thesis activities by discussing the interactions in a formal learning environment. Section 1.2 summarises the main research questions that we are investigating. We then explain the design decisions for developing an empathic ETS in Section 1.3 and give an overview of our empirical studies in Section 1.4. We detail our main contributions in Section 1.5 and finally, Section 1.6 gives an overview of the thesis.

1.1 Motivation

One-to-one human tutoring is the most successful method of improving student learning gains (Wiemer-Hastings et al., 1998). However, in most formal learning environments such as classrooms this successful method is not used due to the cost implications of providing trained tutors for each individual student. Nevertheless, schools continue to seek alternative methods of improving student learning gain because public examination results are an important measure of a learner's attainment and a school's effectiveness in education provision for students. So one compromise in school systems is providing untrained tutors (referred to as teaching and learning assistants) to encourage learning gain as they are more successful as one-to-one or small group tutors for students than a trained classroom teacher with many students (Graesser et al., 1995). However, even with this strategy there may not be enough support for individual children within a classroom throughout a typical lesson. Research into the development of tutoring systems aims to increase the tutor to pupil ratio by providing cost-effective computer-to-human tutorial support within a classroom environment to imitate untrained tutors and match the learning gains they achieve. In the future, sophisticated tutoring systems may be developed that can match the learning gains of trained human tutors in one-to-one tutoring once technological limitations are mitigated.

The development of Intelligent Tutoring Systems (ITS) such as AutoTutor (Wiemer-Hastings et al., 1998; Graesser et al., 1999) has successfully improved student learning gain, by imitating strategies used by untrained human tutors, through the provision of cognitive (task-based) feedback to university students studying Computer Science or Physics. However, these automated systems disregard student emotion even though there appears to be a link between emotion and learning within the human-to-human learning process (Goleman 1995). The human tutor or teacher actively seeks to influence emotion in order to encourage learning.

In the past adding affect to an ITS may not have been explored because emotion expression was difficult to implement. This has changed with the advent of more sophisticated ECA technologies that allow emotion expression through face, gesture and speech. Another factor that has limited the use of affect in recent computer-to-human learning contexts, is that differences in the learner emotions are difficult to detect reliably and respond to within computer-based learning environments, (Beale and Creed 2009).

For the reasons described above, whether there is a link between learning and emotion in computer-to-human learning environments is still very much an open question. The current thesis aims to address this issue and make a contribution to the field by a systematic investigation into emotion and learning in formal classroom settings.

We build on the limited knowledge in this field so far. In particular, there are some identified emotional states within an educational setting such as flow (a state of concentration) (Csikszentmihalyi 1990) and confusion (a student encounters a gap in linking current knowledge to new information) (Kort et al., 2001), that are linked to higher learning gains in comparison to boredom, which has been linked to lower learning gains in computer-based learning environments (Craig et al., 2004). We investigate such emotions and ways in which an artificial tutor may seek to encourage positive emotional states and alleviate a student's negative emotional state.

Recently, research into adding affect (emotion) capabilities to AutoTutor to detect and respond to boredom, confusion and frustration (D'Mello et al., 2009) illustrate the challenges that still remain in affect detection and appropriate affective system response (Picard et al., 2004). There are a number of identified theoretical strategies which have not yet been fully realized in existing implementations (D'Mello et al., 2009). There is a need to implement these strategies and do empirical research into affect detection and response within tutoring systems. This may establish which strategies impact positively on student emotion and consequently improve learning gains. Besides encouraging positive emotions, there is a need to alleviate negative emotions.

Human tutors have been observed to use strategies such as empathy to successfully respond to negative student emotion (Burleson 2006) and consequently reduce negative emotion which can improve learning. An artificial tutor could use a similar strategy within computer-to-human tutorial environments as part of a feedback strategy to encourage higher learning gains.

In this thesis, the concept of empathy plays a central role. We adopt Davis' (1994) definition of empathy as the tendency to vicariously experience other individuals' emotional states. Research on affective systems that use empathy seeks to firmly establish whether it can positively impact learning in computer-to-human tutoring across a range of contexts (Bickmore et al., 2010). In line with this approach, this thesis aims to investigate the impact of an empathic ETS on learning gain to address some of the issues raised. The following section details the four main research questions that are explored in the thesis.

1.2 Research questions

As previously mentioned the overall aim of this thesis is to investigate the impact of an empathic ETS on learning in a formal educational context.

This aim has been divided into the three key areas. Each is associated with a research question: (1) how to facilitate accurate identification of emotion expression by an ECA; (2) whether an ETS using an empathic feedback strategy can improve learning gain and user judgements; and (3) whether there is any link between user empathic tendencies on the one hand and learning gain and user judgements on the other hand. The remainder of this section refines and explains these key research questions.

1.2.1 Expressing ECA emotion accurately

Q1: Which combination of channels used to portray ECA emotions are users able to identify most accurately?

As described earlier, an affective tutoring system must respond appropriately to user answers and emotion, as an inappropriate response can negatively affect the interaction (Bickmore and Schulman 2007). This means that it is of great importance that the empathic ETS' expression of emotion is interpreted accurately by the user.

However, few studies have reliably investigated the optimal combination of modalities with which an ECA can express emotions. We address this issue by measuring whether users are able to accurately identify the ECA emotion expressed in addition to giving confidence and intensity judgements.

1.2.2 Developing an empathic ETS

Q2: Does an empathic ETS feedback strategy positively affect students' learning gains and student subjective user judgements?

So far, few studies have investigated the impact of affective tutoring systems in formal educational settings such as a real classroom. For example, though the work of Arroyo and collaborators (Arroyo et al., 2009b; Arroyo et al., 2009c) was done in a classroom setting, it centres on an embodied empathic learning companion (rather than a tutor). And even though Andallazo and Rodrigo (2013) did work with an embodied affective tutor, their results were inconclusive. In the current studies, we investigate an implemented empathic ETS in an actual classroom by comparing it, on learning gains, with a neutral ETS. We also elicited subjective feedback from the learners on their interactions with the ETS. This way we aim to gain new insights into the effectiveness of empathic ETS in classroom settings.

1.2.3 User empathic tendency

Q3: What is the impact of student empathic tendencies on learning gain and students' subjective judgements whilst interacting with an empathic ETS?

The lack of studies investigating the impact of affective tutoring systems in formal education contributes to the lack of knowledge on the impact of user traits (such as their empathic tendency). Current research has not examined the impact of empathic tendency on interaction with an ETS and on learning gain. The empirical studies in this thesis yielded data that may give an indication about the relationship between learning gain, empathic tendency and type of ETS.

1.2.4 Evaluations in formal environments

Q4: How do results from evaluations with users in a classroom environment compare with results from studies conducted in other settings or different users with affective tutoring systems?

To the best of our knowledge there have been relatively few studies evaluating the impact on learning gain of an embodied affective tutoring system with teenagers within classrooms in secondary schools. Some studies on tutoring systems are based in laboratory settings with university students being tested individually (D'Mello et al., 2010) whilst others have used human judges to observe and evaluate students' emotion (in relation to their behaviour) whilst using Intelligent Tutoring Systems and simulation-based environments in schools (Hershkovitz et al., 2013). Other studies have evaluated different pedagogical affective systems such as learning companions within schools (Arroyo et al., 2009c). The studies in this thesis provide results on the evaluation of an empathic ETS used with teenagers within a classroom setting and can be compared with existing research to establish transfer across domains or learning contexts.

1.3 Design decisions

In this thesis we investigate the impact of an empathic ETS on learning gain within a classroom. To answer the research questions derived from this overall aim and discussed in Section 1.2, we carried out a number of studies where findings from each study informed the design of the empathic ETS and shaped the subsequent studies.

In this section, we explain and justify our design decisions regarding the ETS and the educational context in which it was evaluated. In particular, we discuss the theory-driven approach to empathy and the web-based quiz setting as a teaching activity. The latter needs to be read in the context of Research Question 4 (see Section 1.2.4), where we introduce and justify our focus on the use of an ETS in real secondary school classrooms.

A principal design decision regarding our ETS is the use of a theory-driven approach. In particular Davis' (1994) theoretical concept of empathy informs the ETS's feedback strategy. Davis' organisational model of empathy is important because it unifies both cognitive (understanding emotion) and affective (emotional response) aspects of empathy and describes how the target and observer interact. This is useful, because in our learning context we can define the target as the learner and the observer as the ECA and accurately describe the tutorial interactions between these two parties. Also, the model describes how empathic response can be split between parallel (similar emotion) and reactive (different emotion) empathy. This distinction allows us to group the ETS and ECA emotional responses by whether they mirror the learner's emotion to encourage the continuation of the same emotion or to react with a different emotion to that of the learner's to alleviate it. Sections 2.2 and 2.3 of Chapter 2 detail how parallel empathy can be used to reinforce positive emotion and reactive empathy can reduce negative emotion during a typical interaction between the user and the ETS.

Our second key design decision concerns the type of teaching activity for which the ETS provides support. In research on Intelligent Tutoring Systems, explanation-centred tutorial interaction using full natural language dialogue is an important topic of research. The AutoTutor system (see Section 2.1.2 for details) is the current benchmark in this area of research. The explanation-centred approach is however extremely labour-intensive (involving large project teams to create a full system and appropriate knowledge repositories). Additionally, natural language understanding technology is not yet sufficiently robust for real classroom settings.

For the current research we decided to use a web-based quiz activity as the teaching activity in which the ETS is deployed. Students complete a multiple choice quiz that consists of successive stages of different difficulty. They are coached by the ETS (which provides empathic feedback) as they complete an individual question and are also provided with further support from the ETS in the transition between different stages of the quiz. The choice of a quiz as the central teaching activity is primarily motivated by the real classroom setting in which we decided to evaluate the ETS.

Quizzes are a widely used online teaching activity because they can provide independent learning, and depending on the quiz set-up can be used as a measure of formative and/or summative assessment (McDaniel et al., 2007). Furthermore, quizzes usually require a lower cognitive load as less typing is required from learners and this may provide early learner engagement when compared to explanation-centred online activities with a higher cognitive load.

1.4 Overview of empirical studies

The empirical part of our research is described in Chapters 4, 5 and 6, which each describe a different study conducted in a school environment during normal lessons in a typical IT (Information Technology) classroom.

The study in Chapter 4 informed the design of the ETS, whereas the studies in Chapters 5 and 6 proceed with a comparative evaluation of the ETS (comparing the empathic ETS with a neutral ETS). In this section, we provide an overview of the three studies.

The principal aim of the preparatory study in Chapter 4 is to determine the optimal combination of modalities (speech, face, gesture) for accurately conveying the ETS emotion to the learner. The empathic strategy that the ETS uses to enhance the learning relies on the learner being able to recognise the emotions that the ETS expresses accurately. Different combinations of modalities were systematically tested with participants. The participants were asked to classify the emotion expressed by the

ETS (together with its intensity and the confidence level of these judgements). The results of this study determined the multimodal expression strategy that was used by the ETS.

Following on from this, the multimodal expression strategy is incorporated within the overall empathic ETS feedback described in Chapter 3 which details the system implementation for the developed ETS used in the studies discussed in Chapters 5 and 6.

The developed empathic ETS is compared to a neutral ETS in two successive real classroom based between-subjects studies. The results from the study described in Chapter 5 inform the improved ETS that was evaluated in Chapter 6.

The final set of studies, detailed in Chapter 6, involved learning in the domain of Information Technology (IT). User empathic tendency is measured using the Interpersonal Reactivity Index (IRI) (Davis 1980), to provide appropriate groupings for each condition. The IRI is a psychological questionnaire that has been widely validated to measure differences in empathic tendency by means of four sub scales in a variety of contexts (Cliffordson 2001; Beven et al., 2004). Empathic tendency measures for learners are used to establish whether learners have high, middle or low empathic tendencies and whether these groupings measure differences in learning gain that can be used to maximise the ETS's effectiveness.

As previously mentioned, users interacted with the ETS by completing a set of quiz questions on subtopics within the domain in the studies described in Chapters 5 and 6. On completing their interaction with the ETS, users were also asked to give their judgements for the system by completing an online questionnaire to get each participant's perspective on their interaction with the ETS in order to gain further understanding on the reasons for their choices. This provided a comparison of their judgements to their learning to establish any patterns that future ETS systems can use to increase learning gain. Throughout this thesis, learning gain was measured by the difference between pre-test and post-test scores, a commonly used measure of learning gain in participants. In addition, we used hand written pre-tests and post-tests as this is comparable to summative assessment tasks that learners are required to undertake for their final examinations.

1.5 Contributions

This thesis contributes to a number of research areas related to developing tools to improve learning in formal education and training environments. Research into the improvement of learning in UK schools through effective provision of technology may find this research useful in extending what is known about deployment issues within classrooms. There are few studies that have conducted any evaluations of this nature in classrooms, in the UK and beyond, and therefore insight may be gained through the thesis findings.

Further to this, our research is relevant to those designing affective systems that use ECAs to implement empathy driven tactics to provide user feedback. In particular, given our focus on the impact of an empathic ETS on learning gain, the results we present can be of value to developers of affective tutoring systems whose aim is to replicate human-to-human tutorial interactions. The thesis contributions are discussed in the following subsections.

1.5.1 Modelling and developing ECA emotion

We have modelled and developed a multimodal strategy for expressing ECA emotion when interacting with users. We evaluated this strategy by combining one or more communication channels in an empirical study with four conditions and users were able to accurately identify the expressed ECA emotion in the multimodal condition that combines speech, facial expression and gesture most accurately when compared to speech alone. These results are important because they determine the best strategy for use by the ETS in accurately communicating with users, therefore increasing the success of such a system in improving user learning gain.

1.5.2 Evaluating the impact of an empathic ETS

We applied the multimodal emotion expression strategy described above to the empathic feedback strategy in the ETS that we developed. Following on from this, we conducted an experiment where the empathic ETS was compared with a neutral ETS. No differences were found in learning gains achieved.

However in a subsequent study, when IT users were grouped by empathic tendency, an interaction effect was identified between the type of ETS and users with the highest and lowest empathic tendencies. Users with the highest empathic tendencies achieved the highest learning gains when compared with users with the lowest empathic tendencies whilst interacting with the neutral ETS. In addition, empathic tendency correlated positively with learning gain in the neutral ETS. Furthermore, empathic tendency correlated negatively with learning gain in the empathic condition when learners with the highest empathic tendencies were ranked by empathic tendency score. These studies provide a step toward identifying the impact of user empathic tendency on learning gain with teenagers, which has not been previously investigated. This may aid identification of user groups to be identified for future studies to provide effective and targeted ETS support to increase learning gain.

1.5.3 Conducting evaluations in a formal educational setting

The results achieved in this study are similar to those achieved by evaluations by Andallaza and Rodrigo (2013) within a similar setting (within a classroom with teenagers) albeit in a different domain and learning context. Therefore, our results extend their findings and are informative to researchers who intend to design and implement an affective tutoring system within a formal setting. In particular, the results in our quiz-based learning environment in the domain of Information Technology are reliable in giving an indication of the impact of an empathic ETS on learning gain with teenagers. These results provide the next step toward understanding the impact of an empathic ETS on teenagers using a quiz-based system.

Further to this, all the evaluations that have been conducted in this thesis have contributed to the understanding of the nature of studies in classrooms and illustrated the challenges of deploying artificial intelligence technology within comprehensive schools in the UK. This work illustrates that an ETS can be viably deployed in a real classroom, given the real-world constraints of hardware, software and network availability.

1.6 Overview of the thesis

The remainder of the thesis is structured as follows:

Chapter 2 reviews relevant literature on the use of affect or emotion in learning contexts and how it can be modelled and implemented in empathic computer-based systems. This chapter also provides a background to the empirical research presented throughout the thesis by explaining the theoretical basis of the techniques used to model the ETS used in this thesis.

Subsequently, Chapter 3 describes the various components of the developed ETS system in detail. This system is used throughout the thesis activities that follow in Chapters 4 to 6.

Chapter 4 seeks to address the first research question: Can a multimodal strategy generate accurate ECA emotional expressions? This chapter begins by describing how the multimodal strategy was modelled and developed. Thereafter, we describe the evaluation setup where users were asked to identify and rate their confidence and the intensity of each expressed ECA emotion.

Chapter 5 seeks to address the second research question: Can an empathic ETS improve learning gains when compared to a neutral ETS? This chapter describes the study conducted with participants in the domain of IT.

Chapter 6 seeks to address the second research question (as stated above) and the third question: Can empathic tendency impact on learning gain? The Chapter describes how the study is set up and explains how learner' empathic tendencies are measured. Then we describe the results of the evaluation carried out with participants in the domain of IT.

Chapter 7 concludes by reviewing the thesis activities and revisiting the original research aims. Following on, we discuss the implications of our research in the wider context of affective tutoring systems. Finally, we recommend future work.

Chapter 2 – Background

In Chapter 1 we discussed related work on affective tutoring systems, stated the research questions and the motivation for this thesis. To address the research questions, we developed an empathic tutoring system using a range of strategies and measured their impact by gathering human user judgements and learning gain. These activities draw from relevant research areas and open issues that are discussed in this chapter.

In Section 2.1 we give an overview of how emotion is used in various learning contexts. We will focus on affective tutoring systems that use an Embodied Conversational Agent (ECA) to interact with learners: ECAs have a computer-animated embodiment. We discuss ECAs that can use speech and non-verbal behaviour to communicate and express affect. We describe the types of affective roles, present the common strategies based on human-to-human tutoring and end with a description of the issues of implementation in ECAs.

In the following sections we look at research linked to empathy which is a type of affect expression used in ECAs. In Section 2.2 we define empathy and discuss how this construct is used and measured in educational settings. We discuss the challenges of modelling human-to-human empathic behaviours using ECAs in Section 2.3 and then summarize recent research on the specific emotions experienced in learning contexts.

In Section 2.4 we return to the broader topic of ECAs implemented as a tutoring system and discuss how learner affect can be detected and the tutorial strategies that can model appropriate responses. We give an overview of implementations of empathy using ECAs in Section 2.5, concentrating on multimodal output, including facial expression, speech and gesture. We propose our theoretical framework in Section 2.6 and situate the current thesis activities in comparison to current research. Finally, we summarise the chapter in Section 2.7.

2.1 Affect in learning contexts

Affect, which can be defined as a feeling or emotion, can play an important part in the instructional process. This is important to understand when learning is being encouraged in any context. Learners who experience negative emotion such as anxiety struggle to take in information (Goleman 1995). Further to this, during the learning process itself, some activities may cause learners to feel negative emotion such as confusion and frustration. As a result they are unable to return to a positive state where learning can be more productive, unless there is direct intervention from a tutor or teacher (Kort et al., 2001).

In section 2.1.1 we describe the key roles that ECAs imitate based on human-to-human communication. We then describe the development of ITS in Section 2.1.2. Finally, in Section 2.1.3 we discuss the impact of human tutoring on artificial tutors.

2.1.1 ECAs in learning contexts

The role of affect in web-based learning can be understood by the pedagogical roles that ECAs are imitating in relation to the stages of teaching and learning as discussed by Payr (2003). Their impact can be measured through the learning that has occurred and the judgements of users. Table 2.1 includes a summary of the learning contexts that use ECAs.

Table 2.1 ECAs in learning contexts

Human to Human Learning Context	ECA Role	Description
Tutor	Tutor	• Observe • Help • Demonstrate
Coach	Coach	• Cooperate • Support
Learning Companion	Learning companion Teachable agent	• Support • Question • Request Help
Learning through Role-Play	ECA as an Actor Educational Games	• Learning tool

Artificial tutors such as AutoTutor (Graesser et al., 1999) aim to provide users with cognitive feedback through an ECA. Tutors are concerned with improving learning gain by imitating one-to-one human tutorial strategies. However, earlier systems did not provide any emotional support to users. As the importance of emotion in learning contexts has been more widely accepted, there has been a shift in such systems providing emotional support. These advances are discussed in greater detail in Section 2.1.2 which looks at Intelligent Tutoring Systems, which includes artificial tutors that do not use an embodiment.

The second role of coach is demonstrated by Coach Mike (Lane et al., 2011; Lane et al., 2013), an ECA implemented within a museum to support users to access a Computer Science exhibit in an informal learning environment. Evaluations showed improved engagement, reduced misuse of the exhibit and higher self-efficacy toward programming; however there was limited impact on cumulative attitudes toward programming.

Learning companions are ECAs that encourage peer learning by imitating the human peer-to-peer social relationship, thus assuming that these social rules apply to media (Reeves and Nass 1996). Arroyo et al. (2009c) use gendered learning companions to help alleviate negative attitudes in maths pupils in a high school (by valuing effort and being empathic) where the perceived difficulty of Maths increases whilst the value and student self-belief in Maths decreases when contrasted with learning English with girls compared to boys. These authors implemented these gendered learning companions in an adaptive multimedia tutoring environment. The learning companion works on each problem separately to the student and prompts the learner on their emotional state every five minutes and uses an empathy-based strategy to mirror learners' self-reported emotions thereby providing emotional support. The learning companion also provides metacognitive and motivational feedback based on Dweck's (Mueller and Dweck 1998; Dweck, 1999) recommendations to reward effort and disregard success after the learner attempts each Maths problem. One limitation of this study was that each learning companion provided parallel empathy by reflecting the learner's emotion and therefore the impact of reactive empathy was not investigated.

Although agents implemented as learning companions do not provide cognitive assistance when compared to agents implemented as tutors, they encourage pupil perseverance and show empathy towards the pupils' emotions. For example, in the previously discussed study, female students appeared to learn better when interacting with the male ECA and vice versa which contradicts an earlier study where the male ECA has a greater impact on learning for both genders (Kim et al., 2007). The process of developing a learning companion should include the consideration of additional ECA characteristics such as ethnicity and evaluations should establish their effect with varied student age groups.

Learning by teaching is a well-established technique in education (Roscoe and Chi 2007). One challenge is evaluating the extent that learning occurs where the tutor learns from the tutee. Roscoe and Chi (2007) have developed SimStudent a teachable agent that learns cognitive skills from worked examples in the domain of algebra. In the current version of the system Lucy the teachable agent is given various example problems to solve in preparation for a quiz. The student (tutor) gives Lucy regular feedback during each step of a given problem by clicking Yes or No. Lucy takes the quiz and, if successful, passes the quiz. Preliminary results in this study suggest that learning by teaching environments have an effect on learning equation solving although further evaluations need to be conducted for other domains. Matsuda et al. (2010) suggest enhancing this environment by getting student tutors to justify their tutorial activities although they have not established the effect on learning gain of this change. This suggests the need to establish cognitive theories on learning by teaching environments that generalize across varied domains.

A further type of ECA is one that acts as the object of practice through interactions in different roles (Rickel et al., 2002). The interaction with the ECA itself is intended to have educative value (Lane 2010). To succeed a participant must apply certain communicative skills correctly. The challenge with this area of research is accurately simulating a social context for the practice and learning of a new communicative skill (Lane 2010). Another limitation is that this type of ECA has little empirical or theoretical grounding (Gratch and Marsella 2005). Current research limits successful application to specific domains such as cultural learning, interpersonal communication, and language learning. This suggests less relevance to domains related to STEM (Science Technology and Mathematics) subjects such as Information Technology. A direct comparison of the effects of ECAs as actors vs. other roles such as a tutor system giving explicit cognitive feedback should be explored in future studies.

ECAs existing in virtual worlds are beginning to provide a realistic and authentic environment for this type of role play to take place (Lane 2010). This is important as ECAs can provide higher consistency in comparison to a human role play situation with a virtual peer or human actor. Research in this area highlights the importance of using complex non-verbal behaviour to convey emotions using facial expressions, gaze, body positioning and gesture. Lane (2010) suggests that the intensity (measured by onset, duration and length) of each of these behaviours can be magnified or dampened to vary feedback and adjust the emotion conveyed. ECAs in tutoring systems can use complex non-verbal behaviour to enhance feedback by varying intensity for different emotions.

Educational games have potential effects on learning and encouraging positive emotional engagement such as motivation (Conati 2002; Ke 2008) but what is less understood is the effects on understanding content and skills in a given domain (Moreno and Mayer 2005). ECA interactions in this environment are governed by the trade-off between user entertainment and learning, highlighting the difficulties in developing appropriate pedagogical moves for the agent to support learning whilst maintaining motivation (Conati 2002).

Intelligent game environments can be used to train users in various skills. Hill et al. (2006) have developed ELECT BiLAT a prototype game based environment that uses guided learning to train solders in bilateral engagements and negotiation skills. The prototype game is structured around key phases of a bilateral agreement that include components of an ITS. For example during meetings an ECA in the role of a coach provides feedback while a second ECA in the role of reflective tutor teams up with the coach to provide feedback during reviews. Although this prototype game has not been evaluated to establish effectiveness, if successful this suggests that the merging of ITS and gaming technologies can improve the effect on learning and motivation of immersive training environments.

2.1.2 Intelligent Tutoring Systems (ITS)

An ITS “aims to support and improve the learning and teaching process in certain domain knowledge, considering individuality of a student like in traditional one-to-one instructional process” (Stankov et al., 2007, p.1). Although one-to-one human tutoring is superior to other forms of learning (Bloom 1984), it is difficult to provide due to practical implications such as cost and availability of human tutors. Consequently, this has led to the development of ITS with an ECA as a tutor (Graesser et al., 1999; Heylen et al., 2005; Prendinger et al., 2005) or without an embodiment (Stankov et al., 2007; Keleş et al., 2008; Rau 2013).

A number of studies put forward theories based on the success of one-to-one human tutoring. The aim of these studies is to create a system that models effective behaviours of human tutors and can improve learning more than a normal classroom situation (Zinn et al., 2005). Students seem to learn best when they are active learners. ITS using ECAs seek to emulate the collaborative nature of a human tutor working with a student to correct errors (Wiemer-Hastings et al., 1998; Graesser et al., 1999). Graesser et al. (1999) focus on developing an artificial tutor that imitates the strategies of untrained human tutors, as trained human tutors use complex strategies developed through extensive training which may not be easily implemented within artificial tutoring systems. The success of the pedagogical role of a tutor with an embodiment is discussed in greater detail in the following paragraphs.

The adaptation of this strategy with computer to human training has been demonstrated by Graesser et al. (2003; 2004) to have a significant effect on learning gain when compared to reading a text book for an equivalent time to a tutor session with AutoTutor an artificial tutor. This has been demonstrated a number of times in varied domains such as computer literacy and conceptual physics. AutoTutor is a system that uses an ECA to engage in an explanation-centred dialogue using natural language with learners in a specific domain. Figure 2.1 contains an extract from a short tutorial dialogue between AutoTutor and a high ability college student in the domain of computer literacy (Graesser et al., 1999).

Figure 2.1 Extract from AutoTutor Dialogue

AUTOTUTOR: When you turn on the computer, how is the operating system first activated and loaded into RAM?

STUDENT: When the computer is turned on a file is automatically booted up.

AUTOTUTOR: Anything else?

STUDENT: The file contains the operating system.

AUTOTUTOR: What else?

STUDENT: The operating system is then read into RAM and activated.

AUTOTUTOR: Right the CPU executes instructions in ROM, which help the operating system to start.

STUDENT: After being read to ROM the OS is activated.

The results achieved from this research suggest that AutoTutor can be as effective as a human tutor communicating with a student using a computer-based medium (Graesser et al., 2004). Limitations of the system are that: firstly AutoTutor provides cognitive feedback alone and disregards learner emotion; secondly even if the system were to provide affective feedback the use of a talking head would limit the effectiveness of implementing non-verbal communication such as gesture and posture and would rely on facial expression alone. An affective tutoring system would require implementation using a half or full bodied embodiment to fully imitate human emotion.

AutoTutor's architecture is based on an explanation-centred tutorial interaction that uses state-of-the-art technology such as Latent Semantic Analysis (LSA) and fuzzy rules. LSA is used to compress the domain corpus and then compare these with student answers or contributions (AutoTutor's effectiveness using LSA is similar to an untrained human tutor and lower when compared to an expert human tutor). Also, the system uses fuzzy rules to choose the next topic. This makes direct comparisons of AutoTutor's architecture (implementing explanation-centred learning) to the current study (implementing quiz-based learning) difficult to make because of the difference in the type of interaction with students. However the strategies developed for the tutorial dialogue moves and productions rules are used in the current study and are discussed in sections 2.4.2 and 3.3.2 in greater detail.

The AutoTutor system architecture includes rules that “specify which dialogue response the tutor will make after a student turn. These are based on the content of the curriculum script, the dialogue history, and the quality of the student’s contribution during the last turn, the cumulative quality of the student’s knowledge, and the cumulative quality of the student-tutor exchange.” (Wiemer-Hastings et al., 1998). For example, the following simple production rule is associated with immediate positive feedback for student contribution C:

IF [scriptComponent = Question(j)

AND max(similarity(C, good-answer(j)) > Threshold]

THEN [produceFeedback: "That’s right."]” (Wiemer-Hastings et al., 1998)

Similarly to AutoTutor, the current study uses an ECA within a tutorial interaction where learning gain is measured. The current approach differs from the work on AutoTutor in that it focuses on tutorial support for students that are completing a multiple choice quiz rather than a fully-fledged explanatory dialogue with learners.

Research in this area is looking at how intelligent tutors can further enhance learning. This includes developing complex feedback strategies, that can easily handle student errors and topic changes to improve learning (Graesser et al., 2005). One strand of research in artificial tutoring builds on the work on ITS, with an ECA that provides cognitive feedback, and seeks to address learner emotion by incorporating affect detection (D'Mello et al., 2006) and affect expression through feedback strategies within dialogue used with learners (Arroyo et al., 2009b). Combined affective and cognitive feedback such as giving hints, summarising learning and corrective and metacognitive tips aim to encourage learners' effort, focus and time spent on a task. These strategies aim to impact on learner states such as motivation (Keller 1987) and self-efficacy consequently improving learning (McQuiggan et al., 2008; Dennis et al., 2011). The success of these strategies has not yet been conclusively proven as these learner emotional states are difficult to measure (Hussain et al., 2011) in order to enhance learning gain.

Recently, researchers are investigating empathic ECAs that can be extended to share physical spaces with users as empathic robots tutors, for example the EMOTE project (Castellano et al., 2013). However this research is in its infancy as evaluations have not been conducted on the impact on learning gain. Section 2.2 describes how empathy, a type of affect expression strategy can be incorporated into artificial tutorial feedback to enhance learning gains by reducing negative emotion such as frustration during interactions with artificial tutors (Burleson 2006).

2.1.3 Human-to-human tutoring

Lepper and Chabay (1998) investigate the motivational goals of human tutors and describe four aspects of tutorial interventions: control, timing, content and style. They suggest that control can be maintained by either a human tutor or the student initiating a tutorial intervention. Ideally, students should control their own learning although some studies indicate that students can make poor choices which may interfere with learning (Atkinson 1974). Therefore we are developing an empathic ECA tutoring system that initiates most tutorial interventions in addition to allowing students an opportunity to request additional assistance.

Curiosity and challenge should be maintained by timing tutorial interventions to maximize these two factors (Lepper and Chabay 1998). This encourages student independence without frustrating students with unwanted interventions that may impact on a learner’s state of “flow” (Csikszentmihalyi 1990). Additionally, the type of learning environment can influence tutorial interventions. Our empathic ETS is based on a web-based quiz environment as opposed to other environments such as explanation-centred and discovery learning environments.

Tutorial interventions can be corrective or meta-cognitive to challenge learners and raise curiosity. Corrective or cognitive feedback is defined in the Oxford Online Dictionary as “information about a person’s performance of a task, etc. which is used as a basis for improvement” (Keir 2012). For example during a multiple choice quiz, an ECA indicates whether a learner’s response or answer is correct or wrong. This feedback can be immediately after a learner’s response.

Metacognitive feedback is described as an “awareness and understanding of one’s own thought processes” (Keir 2012). An example of Metacognitive feedback within a multiple choice quiz would be a suggestion by the ECA that the learner repeat the quiz to try and improve on the low current score. This feedback can be given at the end of a quiz or learning interaction. Feedback that includes a combination of corrective and metacognitive feedback is described as leading to improved learning in the long term (Graesser et al., 2005).

Tutorial content can be used to guide student learning. To determine the most appropriate feedback human tutors may use information about student ability to assess the need to intervene. At other times tutors ask students what they feel is the most appropriate assistance they want to receive (Putnam 1985). The current study will use immediate and corrective feedback to guide student learning. To improve poor learning strategies students will receive meta-cognitive feedback. This is described in detail in Section 3.3.2.

The style of human-to-human tutorial interventions includes both cognitive and affective information, therefore an ETS can implement a similar strategy. Cognitive information may be conveyed using hints questioning or summaries, these strategies are discussed in Section 2.4.2. Affective information can be conveyed using facial expression, gesture and speech. Therefore empathic feedback may be used to motivate and encourage learners to achieve their learning goals. The next section describes the impact of one type of affective information in artificial tutoring systems used to provide feedback.

2.2 Empathy in education

Empathy is defined in the Oxford Dictionary as “the ability to understand and share the feelings of another” (Davis 1994; Keir 2012). Empathy is an important element of human-to-human social interaction (Hoffman 2000). Human tutors can respond to user contributions by assessing the content and the emotion of the learner before responding through appropriate empathic expression. Similarly, Embodied Conversational Agents (ECAs) should assess and respond to users' affective states using empathy within an interactive environment (McQuiggan et al., 2008). Therefore, investigating the theories of empathy and how displays of empathy can be modelled within software agents is an important part of developing an empathic tutor.

Section 2.2.1 describes key theories of empathy whilst the following discussion in Section 2.2.2 describes Davis' definition of the organizational model of empathy and how this impacts on the design of empathic software agents. The final sub section 2.2.3 discusses theoretical perspectives on measuring empathy in humans.

2.2.1 Definitions of empathy

Davis (1994, p.12) defines empathy as a "set of constructs having to do with the responses of one individual to the experiences of another". Additionally Hoffman (1987) describes the affective response as being focused more on another person's situation or emotion than on one's own situation. The emotional response can either be identical (Eisenberg et al., 1994) or different to the other person involved (Stotland, 1969). For example an identical or similar emotional response is when one person feels fear and another feels afraid on witnessing their distress. A different emotional response is when one person feels frustrated and the other feels concern on observing their irritation.

Definitions of empathy can be divided into cognitive and affective empathy. Cognitive empathy is understanding others' feeling rather than sharing them (Kohler 1929), while affective empathy is an observer's affective reaction to another's experience (Stotland 1969). These definitions at times appear to place greater emphasis on either cognitive or affective empathy. For example, Eisenberg et al., (1994) describe empathy as an identical or similar affective response to another person therefore appearing to exclude an affective response that is not similar to another person's emotion such as guilt or concern. Kohler's (1929) cognitive definition places more emphasis on the accuracy of reading physical cues and interpreting these correctly rather than "feeling into" the experiences of others. This is a constraint of dealing with emotions as accuracy may be relative to the specific individuals and their sensitivity to the emotion observed. Within a tutorial context, could the tutor read the learner's physical cues inaccurately and therefore respond inappropriately thus not providing empathic feedback? The following section discusses Davis' (1994) approach to empathy which unites both cognitive and affective aspects of empathy.

2.2.2 Organizational Model of empathy

To establish a more unified approach to define empathy and to distinguish between empathy as a process and empathy as an outcome, Davis (1994) suggests the Organizational Model. The Organizational Model aims to represent empathy by using a more unified approach to the cognitive and affective theories of empathy which were previously investigated separately. One benefit of this approach to empathy is that it allows an inclusive definition which includes the affective responses of one individual (the observer) to the experiences of another (target). An empathic episode could be the observer being exposed to the target resulting in a cognitive, affective and/or behavioural response.

Davis (1994) describes four constructs that can be recognized within an empathic episode:

- Antecedents – Characteristics of the observer, target or situation
- Processes – Mechanisms for producing empathic outcomes including non-cognitive (e.g. mimicry), simple cognitive (e.g. direct association) or advanced cognitive (e.g. role-taking) processes.
- Intrapersonal outcomes – cognitive and affective responses produced in the observer from exposure to the target.
- Interpersonal outcomes – behavioural responses directed toward the target resulting from previous experience to the target.

A limitation of this model could be the focus on the effects on an individual observer. However, this is sufficient within a tutorial context, which we have explored, where the observer is the tutor and the target is the student. Nevertheless, any emergent processes in relation to the interaction *between* observer and target are not considered in this model. A second limitation is the recursive nature of the model which does not take into account the effects of one construct on the next construct for example the interaction of interpersonal and intrapersonal outcomes (defined above) when there are multiple episodes. For instance, the behaviours from the target may have an effect on the observer's subsequent emotions suggesting bi-directional interactions between some constructs.

Modelling empathy successfully within ECAs can be challenging, we can use human-to-human interactions to guide some of this work. The organizational model described above can be used to define how empathy is incorporated into empathic tutorial interventions (McQuiggan and Lester 2006). The following section discusses key methods of measuring empathy in adults and children.

2.2.3 Measuring empathic tendencies

Measures of empathy in learners can be used to estimate the differences in empathic tendencies between individuals. This is important in understanding the impact of empathy on both the observer and the target. Paiva et al., (2005) suggest that shared values or gender are important factors for generating empathic outcomes. Factors such as the gender of an empathic tutor may impact on design guidelines. The following subsection looks at some measures of empathy in children and adults and the impact of gender on empathic outcomes.

The Basic Empathy Scale (BES) has been developed by Jolliffe and Farrington (2006) and later validated by Albiero et al., (2009) with adolescents in a different cultural context (Italy) to establish generalization. The BES intends to measure both domains of cognitive empathy and affective empathy to establish whether there is an association between low empathy and bullying using a self-report questionnaire. The main advantage of this method is that it specifically gives an empathy measure for adolescents aged 15 years which is similar to the age range of learners in the current study. However, the study looks specifically at extraversion (cognitive) and neuroticism (affective) which are specific types of empathy in relation to bullying and therefore may not generalise to empathy experienced by learners in a tutorial or teaching and learning context. In addition, measures of empathy in adolescents are disputed because of the different stages of adolescence (early, mid or late) and their relationship to the method of collecting the data as having an impact on results.

The Interpersonal Reactivity Index or IRI (Davis 1980) is widely accepted as a valid self-report multi-dimensional measure of empathy **(**Eisenberg and Fabes 1990). It is divided into four, seven item sub scales: Perspective taking, Fantasy, Empathic concern and Personal distress (see Appendix 4). Perspective-Taking and Fantasy subscales are related to the cognitive component of empathy while Empathic Concern and Personal Distress subscales relate to the affective component. This multi-dimensional measure can be used to measure participants' empathic tendencies. A benefit of the IRI is that once the internal validity of the results has been established, the scores from a single subscale can be weighted by importance in relation to the empathic setting. For example, in the learning environment, the Empathic Concern sub scale score could be used as the primary measure of learners' empathic tendencies and be applied to group learners accordingly. Although the IRI has been widely validated through test and retest studies in many countries and a variety of contexts, this has primarily been with adult participants.

Role-taking is another method of measuring empathic tendency in adults or children. When measures of empathic concern in adults are distinguished from role-taking, Hoffman (1977), in his review of sixteen studies, discovered some differences in gender scores which were confirmed by Davis (1980). Females scored higher than males for empathic concern and there was little difference in the two genders for role-taking (perspective-taking). Eisenberg and Lennon (1983) investigated affective role-taking measures in children and discovered that there was no reliable difference in gender. They did find a small but significant tendency for girls to display greater affective reactivity using the picture story method to measure affective outcomes.

This discussion demonstrates the difficulty of measuring empathy. However, this can be mitigated by the use of a widely established tool like the IRI to measure learner empathy tendencies. On the other hand, the collected data must be internally validated using expected values from previous literature to ensure that it is accurate.

2.3 Modelling empathic behaviour

In the previous section we described how empathy can be defined and measured. In this section we review the strategies used to model accurate affective behaviour in ECAs. Section 2.3.1 describes how observed empathic interactions in human-to-human interaction impact on models of empathic behaviour. In addition we look at the specific affective or emotional states that occur in learning environments and how these impacts on models of empathic behaviour in Section 2.3.2.

2.3.1 Approaches to modelling empathy in an ETS

We are interested in developing an empathic tutor using ECAs. Using the definitions by Davis, Hoffman and Eisenberg we have established that empathy is an important form of social interaction. We can use empathy to develop an affective ECA to interact with learners in a learning environment. Empathy can be modelled within an ETS using training data (from previous or similar interactions) or through a theoretically based approach. This section compares these two approaches that can inform models of ECA empathic behaviour within an ETS.

Empathy can be modelled using training data obtained from *human-to-human interactions*. McQuiggan and Lester (2006) have developed CARE (Companion-Assisted Reactive Empathizer), a data driven architecture and methodology for learning models of empathy. CARE utilizes empathic assessment (when to be empathic) and empathic interpretation (how to be empathic) to determine from the training data when and how to use parallel and reactive empathy. In a typical CARE training session a human trainer directs the actions of one ECA whilst a second human trainer manipulates the empathic states and behaviours of companion ECAs within a virtual environment. At run time, the developed model is used to derive appropriate empathic responses. The main advantage of this empirical approach is the use of training data drawn from the domain of interest. However, one key consideration of the empirical approach is whether one model can generalize to different user demographics in a similar or different domain. Further to this, the gender of the human trainer may impact directly on the level of empathic response chosen for the agent during a training session. The well-established differences between empathy levels between genders may impact on the accuracy of the developed model.

Alternatively, theoretical research on empathy can be used to derive a model for empathic feedback in software agents. A highly abstracted model of empathy can be used to generalise across a variety of domains (McQuiggan and Lester, 2006). However, different theories emphasize different aspects of empathy (e.g. cognitive role-taking) while omitting others. One alternative is using a unified approach such as the *Organizational Model* proposed by Davis (1994) to model empathy to combine the various aspects of empathy responses.

To address negative emotion in learning situations, empathy is used to alleviate some of the effects of frustration or boredom indicated by learners. In human-to-human interactions social-affective skills such as *active listening* is one method used to provide appropriate empathy and sympathy to alleviate frustration (Klein et al., 1999). Additionally learners can be encouraged by human tutors to *attribute* their failures in achieving their goal to factors that they can control such as effort (Batson et al., 1995). Similarly an empathic embodied agent can give a reactive empathic response after observing negative affect such as frustration in computer based education (Prendinger et al., 2004; Burleson 2006).

Confusion can occur when a learner encounters new concepts or confronts contradictions in their knowledge of a particular domain. One strategy to alleviate the impact of prolonged confusion is a *reactive empathic response* from the tutor acknowledging a learner's attempts to reach a particular goal (D'Mello et al., 2009). This strategy can be implemented by an empathic tutor when responding to confusion in computer based learning. The empathic agent can direct learners to information that may address this state.

Positive emotion in learners suggests that they have succeeded in their goals. Parallel empathy such as the tutor replicating positive emotion in the learner can be used to *acknowledge learner progress* and affect. An empathic tutor can react in a similar fashion by mirroring the learner's affective state to acknowledge their success within the learning environment (Bickmore and Schulman 2007). The specific affective states experienced in artificial tutorial environments are described in the next section.

2.3.2 Approaches to providing accurate empathic feedback

Bickmore and Schulman (2007) suggest restricting user input to ensure that the correct affective state has been identified and therefore allow the tutoring system to provide accurate empathic feedback. To establish the impact of the specific construct of parallel and reactive empathy, as discussed in Section 2.2, the user's affective state should be addressed. If the target is the learner and the observer is the embodied agent in a tutoring system the following rules can be used to decide on empathic interventions:

- Positive emotion from the target would be mirrored by the observer to acknowledge the target's emotion.

- Negative emotion from the target would be addressed through empathic concern by the observer to encourage the target.

Bickmore and Schulman (2007) use a similar strategy in their study where they investigate the benefit of restricting user input to encourage high empathic accuracy in the domain of health care. Users are asked to indicate their affective state from a choice of the following affective statements: great, stressed, anxious, exhausted, disappointed and I’ve been better. Their embodied agent “Louise” responds to positive emotion with a parallel happy facial expression and appropriate feedback such as praise. She responds to all other states with a reactive concerned facial expression and appropriate feedback to that particular state. We intend to investigate whether this strategy can be successful in an educational context such as a classroom.

We are interested in how to induce effects on the emotions of the reader/ hearer using an embodied conversational agent (ECA). This is important because an ECA can use these techniques to encourage the correct emotion for a given application. Two currently used techniques are strategic and tactical. Strategic refers to the varying of the content whilst tactical refers to the varying of the form of content (van der Sluis and Mellish 2009). The success of strategically varying text to induce emotion has been established although tactically varying text has not yet been firmly established. Van der Sluis and Mellish (2009) conduct a study where they establish that the use of emphasis, vague adjectives and adverbs can be successfully used in phrasing positive feedback. Although this is a step forward, further work can investigate how each individual strategy in the use of language (i.e. emphasis) used with positive feedback could impact on learning to inform the tactics of systems providing feedback to users.

Pekrun et al. (2002, p.3) who define student academic emotions as "emotions directly linked to academic learning, classroom instruction and achievement", conducted an investigation on their impact on motivation and academic achievement. The results of their study are important as they propose a taxonomy on which academic emotions affect motivation and learning. This taxonomy can be used to inform the design of affective tutoring systems that include ECAs. In their study, Pekrun et al. (2002) develop a quantitative self-report instrument called the Academic Emotions Questionnaire (AEQ). The emotions measured include students' enjoyment, hope, pride, relief, anger, anxiety, shame, hopelessness, and boredom. This instrument is based on an analysis of emotions experienced by learners at school or university in academic settings such as during class, whilst studying or taking examinations, from a series of qualitative studies. These authors use the AEQ to test Pekrun's (2002) cognitive-motivational model which makes assumptions concerning the effect of emotions on students learning and achievement. These authors suggest that emotions seem to be closely linked to interest, motivation and strategies of learning and in predicting students' academic achievement. Limitations of this study include the use of cross-sectional designs and self-report as this may need to be supported by further evidence on the role of positive as well as negative emotion. Therefore, this discussion emphasizes the need for an artificial tutor to address a variety of positive and negative student emotions; as opposed to D'Mello et al. (2009), who focus on negative emotions and ignore positive emotions.

2.4 Approaches to affective tutoring systems

In Section 2.3, we have discussed modelling techniques for empathic behaviour. This section looks at the tools and techniques used to develop tutoring systems.

Here we consider related methods and techniques to implementing the empathic ETS used throughout the activities in this thesis. In Section 2.4.1 we describe recent approaches to affect detection whilst Section 2.4.2 discusses approaches to affective tutorial feedback.

2.4.1 Detecting learner affect

Affective feedback requires the ECA to identify or detect learner affect in addition to task-related information for the learner. One common method of identifying learner affect is using self-report. Learners can indicate their current affective state via a set of buttons labelled with affective states such as "Excited" or "Confused" at the end of a topic or quiz. However, this method may be unreliable in some participants depending on when it is requested and its impact on subsequent emotion (San Pedro et al., 2013).

Alternatively, student affect can be inferred using software. Machine learning can be used to determine the likely user state through the use of current and or previous interactions. Conati et al. (2002) identify one difficulty with the Andes tutor which adopts this approach. Its processing is slowed down, impacting negatively on the system's response time to learner contributions.

A third method of detecting learner affect includes automatically predicting self-reported user emotion. This can be done through the use of various sensors to detect physiological activity (such as facial expression, gaze, and heart rate) combined with tutor contextual variables (such as number of hints given, time taken to solve the problem and number of incorrect attempts). Arroyo et al. (2009a) demonstrate that facial detection software has a prediction rate of more than 60% of the variance of learner emotional states with tutor contextual variables when compared to the use of tutor contextual variables alone to predict self-reported user emotion. In the same study, sensors such as the pressure seat for measuring confidence, frustration, excitement and interest and the wrist band for measuring user frustration and excitement, are less successful. One issue with the study was the lack of a full data set from every sensor used, due to implementation issues, as this can impact on the accuracy of regression analysis.

Recent research by Hussain et al. (2011) into improving AutoTutor's affect detection using machine learning to detect user emotion from physiological sensors shows moderate success at accurately identifying naturally occurring emotion when compared to emotionally charged photos from the IAPS collection (Lang et al., 1999) to predict retrospective self-reported user emotion. In addition, the intrusive nature of sensors in addition to the cost of deploying these in a classroom may impact on evaluations.

In one recent study using tutor context variables alone to detect student affect, Andallaza and Rodrigo (2013) implemented Grimace, an affective tutoring system using an ECA as an algebra tutor with teenagers with mixed results. These authors did not find significant differences in learning between the two versions of Grimace which varied in how often the system intervened once concentration, boredom or confusion was detected. However, learners preferred Grimace version 2 over version 1, which intervened less often.

This discussion emphasises the need for the tutoring system to accurately identify user emotion and intervene appropriately when negative emotion is identified in learning environments. Therefore, the current study uses self-report as a reliable, cost effective measure to identify learner emotion within an ordinary classroom environment.

2.4.2 Tutorial dialogue

Previously we described the importance of an ECA's pedagogical role in shaping the type of intervention given during learning based on human-to-human tutoring. One-to-one tutoring is a successful strategy in improving learning in computer based learning (Graesser et al., 2005).This section looks at the different strategies used to convey tutorial feedback to users in tutoring systems particularly web-based systems that can be developed rapidly in comparison to other tutoring systems. We look at various theoretical concepts for feedback on a given subject such as hints in Section 2.4.2.1. Finally, we discuss the impact of affective feedback on tutorial dialogue in Section 2.4.2.2.

2.4.2.1 Cognitive feedback

Feedback is used by tutors to give students guidance on the correctness of their answers. Cognitive feedback is related to beliefs, thoughts and rational arguments linked to problem solving (Verplanken, 1998). Tan and Biswas (2006) conducted a study with Betty's Brain a teachable agent system. They demonstrated that task-related or cognitive feedback can be used to improve immediate learning. This is achieved by correcting errors as soon as they occur in computer-based learning. These authors also showed that on-demand guided and meta-cognitive (learning about learning) feedback allows users to understand concepts and transfer knowledge. To support learners, corrective feedback can be used to reduce the cognitive load within an interaction, by immediately addressing correct/incorrect answers. In addition, pupils can receive meta-cognitive feedback to encourage self-reflective learning on completing a topic.

Human tutors use various strategies to intervene when students give incorrect answers. The challenge faced by a tutor is determining when to use a strategy to maximum effect. AutoTutor (Graesser et al., 1999), an ITS, successfully imitates a number of human tutoring strategies. These are based on the observed tactics of untrained tutors, as trained tutors use sophisticated strategies that are difficult to implement with an artificial tutor (Wiemer-Hastings et al., 1998), and correspond to Bloom's taxonomy (1956) of educational goals. This taxonomy follows the thinking process and categorises and orders thinking skills and objectives, these thinking skills are listed below from lower to higher order skills:

- Knowledge
- Comprehension
- Application
- Analysis
- Synthesis
- Evaluation

The current study is developing multiple choice quizzes based on knowledge, comprehension and application thinking skills using a subset of the strategies described in AutoTutor:

1. Immediate feedback – give positive or negative feedback.
2. Summary – Recap or reinforce an answer or solution.

Task related feedback forms an important part of tutorial dialogue and the challenge is to appropriately convey this information to encourage learners to make progress and impact positively on the learning experience using an ETS.

2.5.2.2 Affective feedback

During learning, human tutors respond to user contributions. Their response can target the content, as described in the previous subsection, and the emotion within the contribution. Affective feedback is based on emotions arising while completing an activity (Verplanken et al., 1998). Some affective feedback strategies include politeness (Brown and Levinson 1987), relational dialogue (Bickmore 2003), attribution theory (Weiner 1985) and empathy (Davis 1994). Similarly, in computer-based learning, when negative emotion is detected, tutors can for example express reactive empathic feedback to reduce impact on learning (Burleson 2006). Positive emotion can be mirrored by the ECA to encourage the learner state of flow (concentration) (Csikszentmihalyi 1990).

In an educational or training environment researchers have identified key user affective states based on mixed initiative dialogue interactions with AutoTutor (D'Mello et al., 2006). The aim of the study was to predict how accurately AutoTutor could match student verbal self-reported emotion to context related information from AutoTutor’s logs. D’Mello et al. (2006) identified eureka, confusion and frustration as three of the most common states within natural language tutorial dialogue. These authors defined eureka as “a feeling used to express triumph on a discovery, confusion as “a failure to differentiate similar or related ideas” and frustration as “as making vain or ineffectual efforts however vigorous; a deep chronic sense or state of insecurity and dissatisfaction arising from unresolved problems or unfulfilled needs” (D'Mello et al., 2006). However, one limitation of their research is the low frequency of reported negative emotion using verbal self-report by users. Further to this, AutoTutor itself may inhibit some emotions from occurring because of the *type of interaction* that it imposes on the learner. For example, curiosity had low frequency report by learners possibly due to the lack of choice of tutorial topics for learners**.**

Their findings confirm that negative emotion such as frustration occurs frequently in learning and an ECA tutoring system should adequately address these learner emotions. In contrast, one study suggests that *boredom and frustration* occur frequently within computer-based learning environments such as ITS and as tasks become more challenging (San Pedro et al., 2013). Although the analysis of which emotion occur most frequently is beyond the scope of the current thesis, it is important to note which emotions are most likely to occur and ensure that the developed system can respond appropriately.

2.5 Implementing affective ECA behaviour

Earlier in this chapter we introduced ECAs that can be implemented in the pedagogical role of a tutor, expressing cognitive and empathic feedback as successful strategies for encouraging learning in web-based tutors. We identified techniques for modelling empathic behaviour in Section 2.3 and approaches to affective tutorial feedback in Section 2.4. In this section we look at the specific methods of generating ECA empathic behaviour that is recognisable to humans. We begin with an overview of multimodal output in Section 2.5.1, followed by a more detailed analysis of methods related to generating speech, facial expression and gesture in Sections 2.5.2 to 2.5.4.

2.5.1 Multimodal output

ECAs have been developed to imitate human-to-human conversational properties. Cassell et al. (2000, p.52) describes these as follows:

- "The ability to recognize and respond to verbal and non-verbal input.
- The ability to generate verbal and non-verbal output.
- The ability to deal with conversational functions such as turn taking, feedback, and repair mechanisms.
- The ability to give signals that indicate the state of the conversation, as well as to contribute new propositions to the discourse."

Therefore ECAs could imitate human-to-human conversational properties through the use of more than one communication channel (or multimodal communication) that may include one or more of the following: speech, facial expression, gesture and posture. This is demonstrated by REA an ECA developed with multimodal input and output capabilities in the domain of Real Estate (Cassell et al., 2000) and Greta an ECA developed using communicative acts (Pelachaud 2005). One constraint to consider is that human users can synchronize their non-verbal behaviours to improve conversation smoothness; however an ECA may not have the same ability to improve synchrony and adapt its behaviour to the other party as the conversation continues.

This study uses multimodal output to develop a believable empathic ECA that is able to imitate human conversation through the use of speech, facial expression, gesture and posture. However, the ECA in the current study accepts self-reported emotion in addition to task related information such as the selected answer as input.

2.5.2 Speech

Speech is an important part of human-to-human communication, implementing appropriate speech using ECAs is not a trivial task. A key consideration is whether to use synthesised (machine-generated) speech in comparison to pre-recorded speech in tutoring systems. A number of studies suggest that pre-recorded speech using a human voice may improve student engagement (Baylor et al., 2003), student learning and usability (Atkinson et al., 2005). Another study suggests voice characteristics such as "voice pleasantness" and "listening effort" required is more important than the type of speech (Moller et al., 2006). Pre-recorded speech is less flexible as the speech cannot be adjusted easily whilst synthetic speech can introduce usability issues such as timing although advances in this technology have improved this. Forbes-Riley et al. (2006) suggests little difference in learning, as learners can read the tutor transcript and are not entirely dependent on speech, in their spoken dialogue tutoring system.

Investigations carried out during the study presented in detail in Chapter 4 emphasise the use of redundant channels of communication to ensure that communication goals are met during every interaction. In addition, although pre-recorded speech is less flexible, it is used in the current study to encourage user engagement during interaction with the ETS; based on a pilot study (see Appendix 1) we conducted where learners gave negative reactions to synthetic speech, as described in Section 3.3 and Section 3.4.

2.5.3 Facial expression

Facial expression is related to what is occurring in concurrent conversation (Cassell et al. 1994). The same authors describe facial expression as being related to speech in conversation in the following three ways:

- Syntactic functions - describe facial movements such as raising eyebrows that are synchronized with specific words or accented syllables.
- Semantic functions - emphasize speech by referring to a word or emotion.
- Dialogic functions - control speech flow between two people.

Cassel et al. (1994) identify a number of parameters that impact on the functions above including speaker and listener characteristics such as social identity and the listener’s reaction to the speaker’s utterance.

Pelachaud (2005, p.686) describes the automatic generation of facial expressions for GRETA an embodied agent using temporal parameters. These parameters define attributes such as “sustain” which “is the time during which the expression maintains maximal intensity”. A number of studies build on the work of Eckman and Friesen (1977) on facial expression in realizing affective states in learning environments which is beyond the scope of this thesis; see Poggi and Pelachaud (2000).

2.5.4 Gesture and posture

Non-verbal behaviour such as gesture provides various functionalities in communication including signalling affirmation or rejection (McClave 2000; Kendon 2002). Gesture or gaze can be incorporated into ECA speech using nonverbal behaviour generation rules that are described by Lee and Marsella (2006). These rules are based on analysing surface text, after which relevant non-verbal behaviours are added to the interaction. Each nonverbal behaviour rule has a priority and a set of associated key words that occur in close proximity to the rule. The challenge during implementation of any generated behaviour, for example a pointing gesture, is maintaining the synchrony of the communication act with the speech and facial expression to ensure that the meaning is clearly conveyed.

Further non-verbal communication such as gaze and posture are beyond the scope of this study, although the current study uses automatically generated off-task non-verbal behaviour to imitate human behaviour such as breathing, blinking and direction of gaze. Furthermore, the current study uses automatically generated posture for specific affective states such as concern where the ECA leans forward; see details of implementation in Chapter 3 (Section 3.4.1).

2.6 Implications for the design of the ETS mind

In Sections 2.1 to 2.5 of this chapter, we have provided an overview of key research on Embodied Conversation Agents (ECAs), Intelligent Tutoring Systems (ITSs) including how they are combined with ECAs, and theoretical work on both tutoring and empathy.

In the current section we describe the implications of some of the research that we have described for the design of the "mind" of the current ETS. The ETS “mind” is a term that can be used to describe the tutorial strategies and empathic states which form the theoretical basis for the ETS system implemented in this thesis. These strategies specify how the ETS is to appropriately respond to user input and progress during interactions with the user.

2.6.1 User emotional states

A number of groups have attempted to identify a set of student emotions that are relevant to education (Kort et al., 2001; D’Mello et al., 2006 and D’Mello et al., 2009). These include boredom, flow, confusion and frustration. D’Mello et al. (2006) identified eureka as a frequently occurring state, whilst Craig et al. (2004) identified delight (which is similar) less frequently in their educational context. We used these core affective states for student self-report within the ETS.

Previous research including D'Mello et al. (2010) concentrated on identifying *university student* emotions (2010). We extend this research with our current studies on *high school student’s* emotions.

Other researchers have investigated emotions with Intelligent Tutoring Systems *without an embodiment* (Doddannara et al., 2011; Hershkovitz et al., 2013). In contrast, our system uses an *embodied agent*.

2.6.2 ECA empathic states

Having reviewed the possible user emotional states in the previous sub-section we outlined a set of synthetic emotions that can be portrayed by embodied agents in response to user emotion. This includes excitement, motivation, satisfaction, concern, and surprise. In our opinion, the inclusion of surprise is particularly important because it is based on intuitive professional judgment from a human teacher and tutor in a formal learning environment with over 10 years of experience. We used Davis's theory on empathy (Davis 1994) to determine which affective response the ECA should portray at each point in an interaction. Affective states are generated using relevant facial expressions specific to the Cantoche ECA, described in Section 3.3. In Chapter 4, we describe a study where we evaluate how accurately learners can recognise these ECA affective states *prior* to implementation within the ETS as suggested by Beale and Creed (2009) and the impact of these results on subsequent studies.

2.6.3 Measuring empathic tendency

In the first study of its kind, we utilize the Interpersonal Reactivity Index to measure empathic tendency in learners in the current thesis activities. This is because it is the most reliable method available within the field of Psychology. This tool will be used to investigate the impact of learner empathic tendency on learning gain with users interacting with the developed ETS in Chapter 6. As we are not aware of any previous research that has measured learner empathic tendency, this study will contribute towards understanding the effects of learner empathic tendency on learning gain within an ETS.

2.6.4 Detecting student affect

Julie, the ECA used throughout the thesis, just like a human tutor, should be able to regularly and accurately identify user emotion and judge learner progress. However, detecting student affect remains a key challenge when developing an ETS. Self-report is a reliable method of identifying learner emotion when compared to teacher judgements (Woolf et al., 2008). Although self-report is identified as possibly having an effect on consequent emotion, this method can be reliably used when interpreted to relate to a previous problem (Arroyo et al., 2009a). In addition, when automatically predicting self-reported emotion, Hussain et al. (2011) suggest that physiological sensors detecting confusion, frustration, boredom and flow are only moderately successful in identifying naturally occurring affect. Another strand of affect detection looks at context related variables such as number of errors and time taken to complete a step within a problem or question (Arroyo et al., 2009b; Andallaza and Rodrigo 2013). However, this method relies on the tutorial learning context (or previous interactions) providing rich interactional data that allows metrics to be calculated. This would not be suitable in a quiz-based learning environment.

Therefore, because of the prohibitive cost of biometric sensors such as wrist bands, web cams and pressure seats and their intrusive nature in an ordinary classroom situation these factors were a limitation to the current thesis activities. However, as most studies still compare self-reported emotion with data from biometric sensors (Hussain et al., 2011) or tutor contextual variables (Arroyo et al., 2009a), we feel that self-reported emotion remains a cost-effective and valid measure of affect in learning in a classroom environment. Consequently, we will complement current research on evaluating an affective tutoring system that uses task-related information to detect affect within a classroom with teenagers in the domain of Mathematics (Andallaza and Rodrigo 2013), by evaluating our ETS that uses *self-reported emotion* within a classroom based environment *on Information Technology in a quiz-based environment.*

Therefore, Julie can respond to *self-reported user emotion* based on data collected from appropriately labelled *self-report buttons* (Kapoor et al., 2007) to allow users to indicate their affective state on request. Kort et al. (2001) in their discussion identify the key affective states that users experience during learning informing the interface button labels in this study (See Section 2.4 for detail). Users will press one button on request by the ETS to indicate their current affective state from the following list of corresponding labels:

1. Excited
2. Motivated
3. Satisfied
4. Frustrated
5. Bored

Then, the ETS responds to self-reported learner emotion with the appropriate response as described in Section 2.6.6.

2.6.5 Tracking and responding to student progress

The ETS implements the cognitive theory (Demetriou et al., 1993) to make a weighted judgement on a student's progress by comparing the test result to a concept based on pre-determined values to establish whether inadequate, adequate or good progress is being made. In addition, this strategy is effectively used in various existing affective tutoring systems to provide meta-cognitive feedback on the task being completed, in order to encourage student learning (Aleven and Koedinger 2002).

Graesser et al. (2005) use metacognitive feedback to encourage learning in an explanation-centred environment with ITS systems using ECAs implemented as tutors whilst Tan and Biswas (2006) use metacognitive feedback to encourage learning with Betty's Brain, an ECA implemented as a teachable agent; both in the domain of Science. Therefore, we can extend research on learning by using the *cognitive theory* in a *quiz-based learning environment in Information Technology*, to evaluate whether the developed ETS encourages users to improve their progress through metacognitive feedback that includes both affective and cognitive aspects at the end of each section.

2.6.6 ETS tutorial feedback strategies

2.6.6.1 Cognitive feedback

The ECA tutor moves or interactions are based on the correctness of each student answer during each question of the multiple choice quiz (cognitive feedback). These interventions are based on observed intervention tactics of untrained tutors (Wiemer-Hastings et al., 1998) and correspond to Bloom's (1956) taxonomy of educational goals. This taxonomy describes increasing levels of cognitive difficulty with lower levels of Knowledge, Comprehension and Application and higher levels of Analysis, Synthesis and Evaluation. The higher levels are used in AutoTutor (Graesser et al., 2004) to encourage users to give more detailed responses that included responses to "why" and "how" questions with an emphasis on detailed examples and extended answers from the users within an explanation centred learning environment. The educational goals and corresponding tactics are highly dependent on the learning environment being implemented.

The current study seeks to complement previous research on the impact of the untrained tutorial tactics on high level educational goals in an explanation centred learning environment in physics and computer literacy. We implement the tutorial tactic of *immediate feedback* with an ETS that implements *the lower level educational goals that include knowledge, comprehension and application within Information Technology in relation to a quiz-based learning environment.*

2.6.6.2 Affective feedback

We adopt an analytical approach based on the Organizational Model which unifies theoretical concepts of empathy within a set of four constructs as discussed in Section 2.2.2.

We apply the Organisational Model within a one-to-one tutorial context which we are investigating in the current thesis. As stated previously, the specific situation we have explored is a computer-to-human tutorial interaction within a classroom setting. The *observer* is the ECA implemented within the ETS whilst the learner is the *target*. The ETS models an *advanced cognitive process* to interpret learner self-reported emotions and give a suitable response by applying a specific set of rules based on human-to-human tutorial interactions. The intrapersonal construct looks at the affective and non-affective outcomes of the ECA's exposure to the learner's emotion (Davis 1994). An example of an affective outcome is the *emotional reaction of the ECA*, whilst an example of a non-affective outcome can be the *correct interpretation of a learner's emotion* based on cognitive role-taking; "the attempts of one to understand another by imagining the other's perspective" (Davis 1994, p.17). This is important in modelling how software agents such as ECAs can express empathic behaviour within an interactive tutorial environment. In a learning environment, an interpersonal outcome can be an empathic tutor that responds affectively to a learner's perceived emotion using either *parallel or reactive interventions* in order to *help* learners make progress in the learning process. A parallel intervention includes the imitation in an ECA of the learner's feelings such as *excitement or satisfaction*, whilst a reactive intervention describes an affective reaction in the ECA which differs from the observed affect (Davis 1994). For example, when observing distress or *frustration* by the learner, an ECA may experience *concern* for the target individual.

Burleson (2006) shows that empathy can alleviate negative student emotion for example frustration. Similarly, other researchers focus only on addressing *negative emotion* and ignore student concentration or any other positive affect detected (D'Mello et al., 2009).

However, affective feedback that accompanies the interactions described in the previous subsection should additionally aim to *enhance student motivation* (Schunk 1991). Lepper (1993) describes four main goals of motivating learners:

1. Challenge
2. Give them confidence
3. Raise their curiosity
4. Make them feel in control

In addition “positive politeness” and “attending to the hearer” are two strategies described by Wang et al. (2008) as an extension of the Politeness Theory (Brown and Levinson 1987).

Therefore the current study will extend previous research that uses empathy to respond to *negative user emotion only* by implementing an affective strategy based on *Davis’ theory of empathy* to respond to *positive user emotion with parallel empathy and to negative user emotion with reactive empathy*.

This will address three of the motivational goals of challenge, confidence and control by attending to the hearer and incorporating positive politeness. These goals will be achieved by various tutorial tactics to give appropriate feedback (Heylen et al., 2004):

1. Challenge – selecting appropriately difficult questions by including low and middle ability content accessible to all students.
2. Confidence – maximize by praising directly or indirectly.

3. Control – Suggesting that students make a choice by offering the option for multiple attempts on each quiz section.

2.7 Summary

This chapter has critically evaluated existing systems, techniques and methods in developing affective tutoring systems implementing ECAs. The implications of previous work on the current thesis activities and their impact on the ETS design are discussed in detail in Section 2.6.

Section 2.1 illustrates that the basis for developing tutoring systems is founded on the success of one-to-one human tutoring. Within the field of Intelligent Tutoring Systems (ITS), work on AutoTutor (Graesser et al., 1999) has played an influential role. AutoTutor is *explanation-centred*. Subsequent, recent work on affective ETS has also taken this explanation-centred approach (Andallaza and Rodrigo 2013). Apart from the fact that studies into affective ETS have only recently been carried out, the focus of this work on explanation-centred interaction has been at the detriment of the exploration of other forms of interaction. In particular the use of an empathic ETS with *quiz-based* interactions, which are widely used in education, has not yet been investigated. Our work adds to the recent studies into empathic ETS by exploring their use in *quiz-based* interaction.

Current studies have conducted evaluations of affective tutoring systems within *controlled laboratory experiments with adults*, whilst other studies have evaluated ECAs implemented in classrooms in other affective pedagogical roles such as a *learning companion* (Arroyo et al., 2009c). In contrast to the aforementioned, we have conducted our study of an *empathic ETS* (whose implementation is detailed in Section 3.2) within an actual *classroom* environment with *teenagers.* These evaluations are described in Chapters 4, 5 and 6.

In Section 2.2 we looked at the role of empathy in computer-based learning, and in Section 2.3 how empathy can be modelled and implemented in human-computer tutorial environments. Davis's theory of parallel and reactive empathy has so far only been explored with multiple ECAs that engaged with learners as actors in an *interactive role-playing* learning environment (McQuiggan et al., 2008). From an educational perspective, this approach is most similar to that of *learning companions*, rather than tutors. So far, to our knowledge the work of Davis has not been implemented as an ETS system in a *tutorial* context. Our work addresses this gap. We describe the evaluations of our empathic ETS in Chapters 5 and 6.

Furthermore, we reviewed various measurements of empathy and use the *Inter-Reactivity Index* to measure learner empathic tendency in Chapter 6. This is to investigate the impact on learning gain within our evaluations with the ETS, which has not been previously done in this context.

Following on from this, Section 2.4 describes the theoretical grounding for the ETS we have developed. The ETS identifies learner emotion using self-report and provides cognitive and affective empathic feedback to respond to learner emotion. In addition, we provide metacognitive feedback based on learner emotion and progress. We provide a detailed description of the ETS implementation in Sections 3.3 and 3.4.

We have reviewed the strategies used to implement ECA behaviours including speech, facial expression and gesture in Section 2.5. Previous research has not investigated the optimal combination of modalities used within an ETS. Therefore, to address this gap in knowledge we have carried out an empirical study in which different modalities including speech, facial expression and gesture were compared in Chapter 4. In addition, previous studies have not evaluated the ECA with users *prior* to implementation within an ETS; for example Andallaza and Rodrigo (2013) therefore to address this gap we have conducted our study prior to use within an ETS to inform the subsequent studies in Chapters 5 ad 6.

The following chapter describes the implementation of our empathic ETS.

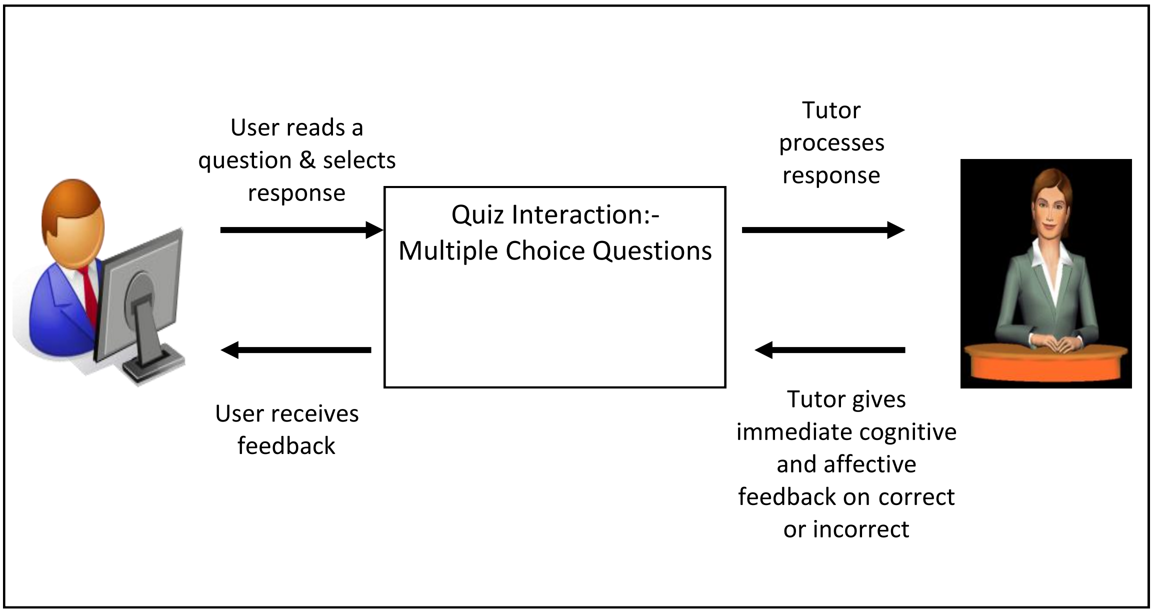
Chapter 3 – Design of an empathic ETS

In Chapter 2, we reviewed existing literature related to the development of empathic tutoring systems implemented using ECAs. This chapter introduces the methodology and focuses on the theoretical background of ECAs that are used in empathic tutoring systems. Section 3.1 describes the quiz-based learning context used in the evaluations in this thesis. The theories relevant to developing an ECA "body" are discussed in relation to developing an empathic ECA tutor that uses speech and non-verbal behaviour in Section 3.2. Finally, Sections 3.3 and 3.4 detail the system implementation and the chapter is summarised in Section 3.5.

3.1 Quiz-based learning environment

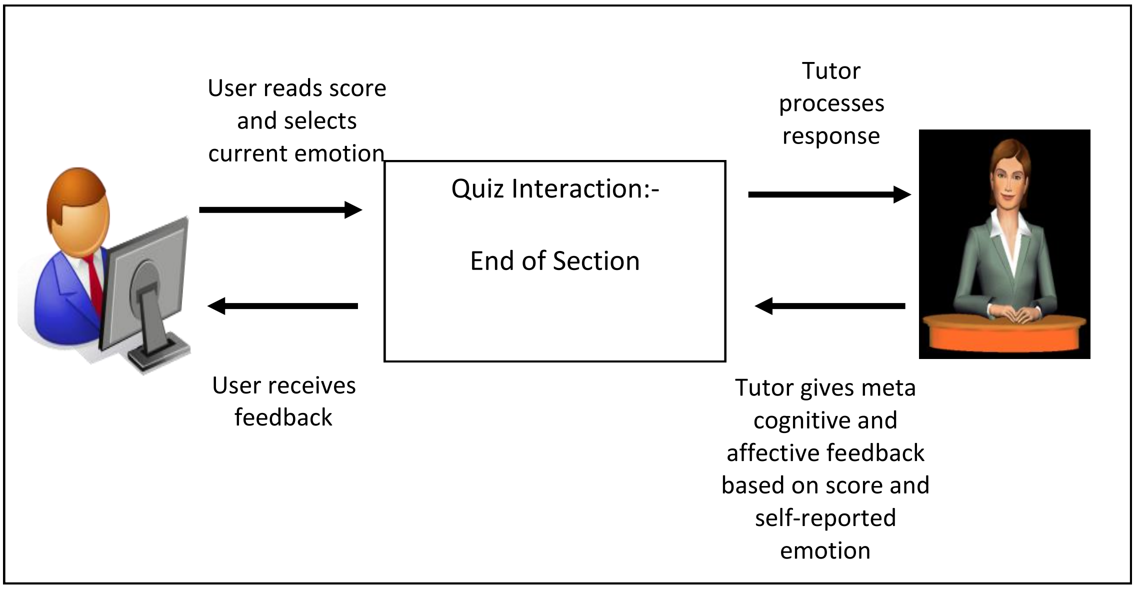
The ETS that we are implementing is set in a quiz-based learning environment as shown in Figure 3.1. The student reads each multiple choice quiz question and then selects a response independently. Then, the ETS provides immediate affective and cognitive feedback in response to a correct or incorrect answer using speech and non-verbal communication as described in detail in Sections 3.3 & 3.4. The student continues until a section is completed. Whilst the student has not selected a response, the ETS is idle exhibiting nonverbal behaviour imitating breathing and blinking.

Figure 3.1 Quiz-based Interaction: questions



In addition as shown in Figure 3.2, at the end of each section, users are given their current section score and asked to self-report on their emotion (see also Figure 3.8). The ETS provides affective and metacognitive feedback based on this response to improve their learning strategies as they progress through the interaction. Learners can select to repeat a section at this stage or continue on to the next section until the quiz has been completed. Finally, learners are prompted to submit their scores by the ETS and thanked for completing the quiz interaction.

Figure 3.2 Quiz-based interaction: end of section



3.2 ETS Body

As stated earlier, the ETS aims to imitate human to human conversation (Cassell et al., 2000). The ETS implemented throughout the thesis activities uses Julie a female ECA developed by Cantoche. Julie includes the following parts: a face and half body, and Julie uses speech and non-verbal behaviour such as facial expression and gesture to communicate; see Figure 3.3 below. Julie was intuitively chosen from the ECAs available to maximise non-verbal communication (specifically facial expression and gesture) with learners and therefore increase the accuracy of emotional communication with the learner. A half-bodied agent was chosen over a full-bodied agent to magnify gestures and facial expressions. Although research by Arroyo et al. (2009c) suggest that male ECAs implemented in the role of learning companion encourage learning more than female ECAs whilst Kim et al. (2007) suggest that for both genders male ECAs encourage learning, it is not clear how the gender of ECAs implemented in the role of tutor impacts on learning gain. Since this is beyond the activities of the current thesis, we discuss the implications of this constraint and future studies in Chapter 7 (Sections 7.2 and 7.3). The following sub sections describe the theories related to implementing the ECA "body" as an empathic ETS.

Figure 3.3 Julie the ECA

3.2.1 ECA overview

The embodied conversational agent (ECA) architecture in the current study will be required to implement the following criteria based on human conversation (Cassell 2000):

- Using different modalities
- Synchronize conversational behaviours

3.2.2.1 Affective speech

Pre-recorded speech has been described as improving engagement, subjective system judgements in addition to sounding more human-like in studies comparing a human voice to a machine-generated voice (Baylor et al., 2003; Atkinson et al., 2005).

Atkinson et al. (2005) achieve similar results and further to this, suggest that pre-recorded speech can improve learning when problem solving in the domain of Mathematics. However, limitations of pre-recorded speech include cost and limitation as the speech cannot be dynamically generated. Appendix 1 describes the results of a pilot study which confirm some of the findings above, as machine-generated speech, which is more flexible, was used in this study but received low subjective judgements from users.

We use pre-recorded speech that includes affective inflections on key words based on techniques used in Drama to express emotion. A trained Drama teacher recorded the speech used by the system. The affective states include speech to reflect excitement, satisfaction and concern. Table 3.1 includes examples of the affective speech for each emotion.

Table 3.1 Example ECA affective speech

ECA Empathic State	**ECA Affective Speech**
Excitement	"Well done!"
Satisfied	"Very good work, you have scored well."
Concerned	"I understand that you may be struggling. It is still important to try your best and get a good score at the very least. Try this quiz again."

3.2.2.2 Nonverbal behaviour

Facial expression

Facial expression can be generated using automatic generation methods which allow fine control over individual ECA facial features such as gaze, eyebrow and mouth (Graesser et al., 1999; De Carolis 2005); however synchronization between speech and nonverbal behaviour can be a limitation of this method.

In addition there have been few, if any, evaluations on the believability of facial expressions by prospective users, prior to system implementation, even though this can affect system outcomes. Conversely, facial expression can be manually generated using pre-defined facial expressions at the cost of the designer having reduced control of facial features on the implementation of appropriate facial expressions; although duration and intensity of facial expressions can be manipulated.

For the current thesis activities, we use predefined expression and animations as detailed in Section 3.3. We extend current research on the believability of affective systems that implement automatically generated facial expression by *indirectly evaluating the impact on the accuracy of students identifying emotions generated by an ECA that uses predefined facial expressions*. Additionally we extend previous research by providing *subjective user judgements on the believability of an empathic ETS implementing predefined facial expressions in a classroom environment*.

Incorporating gesture

Gesture is incorporated into speech using a subsection of relevant nonverbal behaviour generation rules that are described by Lee and Marsella (2006). These rules are based on analysing surface text, after which relevant non-verbal behaviours are added to the interaction. Each nonverbal behaviour rule has a priority and a set of associated key words that occur in close proximity to the rule (See Table 3.2). Gesture is used to enhance the multimodal effect of our feedback in the developed ETS to increase believability and consequently positively impact on learning outcomes.

As a result, we extend the previous authors' research by indirectly evaluating the *impact of a subset of the nonverbal generation rules,* as shown in Table 3.2, *on learning gain and user judgements in our empathic ETS*. We have included key text from these rules in the affective speech used in this study as an illustration. Culture, age and gender have not been taken into account when developing these nonverbal behaviour generation rules and this could be an area for future studies, as it is beyond the scope of the activities in this thesis.

Whilst Lee and Marsella (2006) use *automatic generation* using a non-verbal behaviour generator, we have *manually inserted* the relevant behaviours as part of our implementation of our ECA and empathic ETS. This is described in detail in Section 3.3.1 and this technique is evaluated in human evaluations described in chapters 4, 5 & 6.

Table 3.2 Non-verbal generation rules

Priority	Nonverbal Behaviour Generation Rule	Key Words	Actions
1	NEGATION	No	Head shakes and brow frown (throughout sentence or phrase)
2	AFFIRMATION:	Yes, yeah, I do, I am, We have, We do, You have, true, OK	Head nods and brow raise (throughout sentence or Phrase)
3	ASSUMPTION / POSSIBILITY	I guess, I suppose, I think, maybe, perhaps, could, probably	Head nods (throughout sentence or phrase) and brow frown (where words occur)
4	CONTRAST	But, however	Head moved to the side (lateral movement) and brow raise (co–occurring)
4	INCLUSIVITY	Everything, all, whole, several, plenty, full	Lateral head sweep (co–occurring)

3.2.2 ETS overview

Since the ETS is imitating a human-to-human tutorial conversation within a quiz-based learning environment, we have identified a subset of key requirements for a successful artificial tutor from those described by Zinn et al. (2005):

- Process student affective input
- Process student answers
- Provide multimodal feedback

The ETS architecture implements these identified requirements of an affective tutoring system within a *quiz-based learning environment.* The architecture is partially based on the work on Easy with Eve by Sarrafzadeh et al. (2007); a *Maths tutor in a problem solving learning environment with primary students* with the following common modules:

1. Tutoring Module – Gives appropriate positive or negative feedback based on student progress and affect
2. Student Module – Stores student progress and self-reported emotion
3. Domain Module - Presents the domain content in a structured manner

In addition to the modules described above the current study extends this architecture by including the *following modules* which are specific to implementing the ETS within the specific quiz-based context described in the following sub sections and their success is *evaluated* with students in Chapters 5 and 6:

1. *Manager Module*: This module controls data flow within the system as described in the Black Board architecture for problem solving (Nii 1986) and has the following functions:

- Invokes the tutor module to give individualized instructions and suggestions during each student's learning process.
- Maintain student information, learning status and testing results in the student module.
- Deliver the course content via the interface module.

2. *Interface module* – The quiz interface shows multiple choice questions and written system feedback; whilst the ECA interface displays speech and non-verbal communication.

Other differences between our system and Easy with Eve by Sarrafzadeh et al. (2007), include the *strategies* used within each module which are specific to the type of interaction, for example we are implementing a set of *multiple choice quizzes* which impacts on the presentation of the domain information. In addition, our student module will be using *self-report* as opposed to *physiological sensors*. Nonetheless, the architecture does have the flexibility for modifications and added complexity for similar studies. Moreover, the *human evaluations* we have conducted inform future studies on the success of an ECA based on this *architecture* for tutoring systems.

3.3 Behaviour realization strategies

The following sections describe the system implementation of the empathic ETS. The ECA implementation is described first followed by the ETS overview.

3.3.1 ECA implementation

Julie, the female ECA used throughout the thesis activities is an embodied agent developed by the company Cantoche. Julie appears on the screen as a half-bodied agent sitting at a table as shown in Table 3.3. Julie is able to imitate human behaviours in conversation, such as speech and non-verbal behaviour. Julie has been implemented as a half body agent to ensure that she can clearly communicate using non-verbal behaviour whilst being integrated into the quiz interface.

Julie communicates with users through multimodal output that combines three channels (speech, facial expression and gesture) as discussed in Section 3.2. Table 3.3 summarises this information by describing the link between learner emotion, the ECA empathic state and the corresponding ECA speech, facial expression and gesture required for the thesis activities. Please see Appendix 2 for full details of the set of emotional states and animations for Julie.

For every interaction, Julie's expression is set and the appropriate animation is played before the speech is uttered and thereafter, an appropriate gesture is played. For example Figure 3.4 shows the ECA expression being set to "Happy" whilst the animation "Happy" is played. This supports the required synchronization between each modality necessary for believability.

Table 3.3 Learner emotion compared to ECA empathic state

Learner Emotional State (polarity)	Agent Empathic State (type)	ECA Speech	Cantoche ECA Expression	Cantoche ECA Animation	Graphic
Excited (positive)	Excitement (parallel)	Positive Meta-cognitive Explanation	Happy	Happy Explain	
Motivated (positive)	Motivated (Parallel)	Positive Meta-cognitive Explanation	Smile_02	Acknowledge	
Satisfied (positive)	Satisfied (parallel)	Positive Meta-cognitive Explanation	Basic	Speak	
Frustrated (negative)	Concern (reactive)	Concern Meta-cognitive Explanation	Concerned	Speak	
Bored (negative)	Concern (reactive)	Concern Meta-cognitive Explanation	Concerned	Speak	

In addition, Julie includes off-task non-verbal behaviour whilst the interface is waiting for the next instruction. This includes blinking, head movement and upper body movement to imitate breathing in and out. These behaviours allow the ECA to imitate human off task behaviour in addition to indicating that the system is waiting to respond to user input. In this implementation, the ECA speech bubble has been turned off to maintain the user's primary focus on the quiz questions being completed.

One limitation of implementing Julie is that she can only adopt one of approximately 16 different expressions, for example “Happy” or “Concerned”. In addition, Julie’s expression can be combined with approximately 48 different animations such as “Wave” or “Gesture Left”. These predefined actions limit the author’s fine control of Julie’s features, such as the mouth or left hand, making the implementation less flexible if changes are required. Appendix 2 includes a detailed list of possible expressions and animations.

3.3.2 ETS architecture

The empathic tutoring system architecture supports the presentation of quizzes in the domain of Information Technology (IT).

The system has five main modules which are discussed in the following subsections and illustrated in Figure 3.4. The manager module communicates with the interface module that presents each question and receives user input in addition to maintaining the learner profile. The user input received is passed onto either the tutoring module where feedback is given on the correctness of each response, or at the end of a section, to the student module where feedback is given on current emotion and progress. Feedback is via multimodal output from the ECA.

3.3.2.1 Manager module

This module controls data flow within the system. Instructions are simultaneously passed to the ECA interface to instantiate Julie onto the webpage and to the flash video to load the initial screen on the quiz interface. Julie is therefore visually integrated onto the quiz interface; however, users can move Julie around the browser window using the mouse.

3.3.2.2 The interface module

The user interface is divided into two sections, the quiz interface and the ECA interface. Users interact by using appropriately labelled buttons to acknowledge instructions, indicate their choice of answer or to self-report an affective state. Each system request is presented on the left hand side of the interface.

The interface communicates with the user via Julie the ECA, which is realized through Cantoche software. Julie appears on the right hand side of the webpage as a female half-bodied embodied agent. The half-bodied agent can be incorporated into the system interface more easily than a full-bodied agent to keep a clear focus on the quiz questions. The agent interface conveys every tutor action through speech, gesture, body movement and facial expressions.

The user interface is realized through the use of the ECA interface and a video that has been generated using Flash MX and CS3 Professional as shown in Figure 3.5. The video communicates with the ECA interface through a set of flash commands that correspond to those defined on the webpage. For instance, Appendix 2 includes pseudo code illustrating the “welcome” and “instructions” command. These are sent from the flash video to the ECA interface to launch the ETS and give the user important information on how to use the interface.

Figure 3.4 System architecture

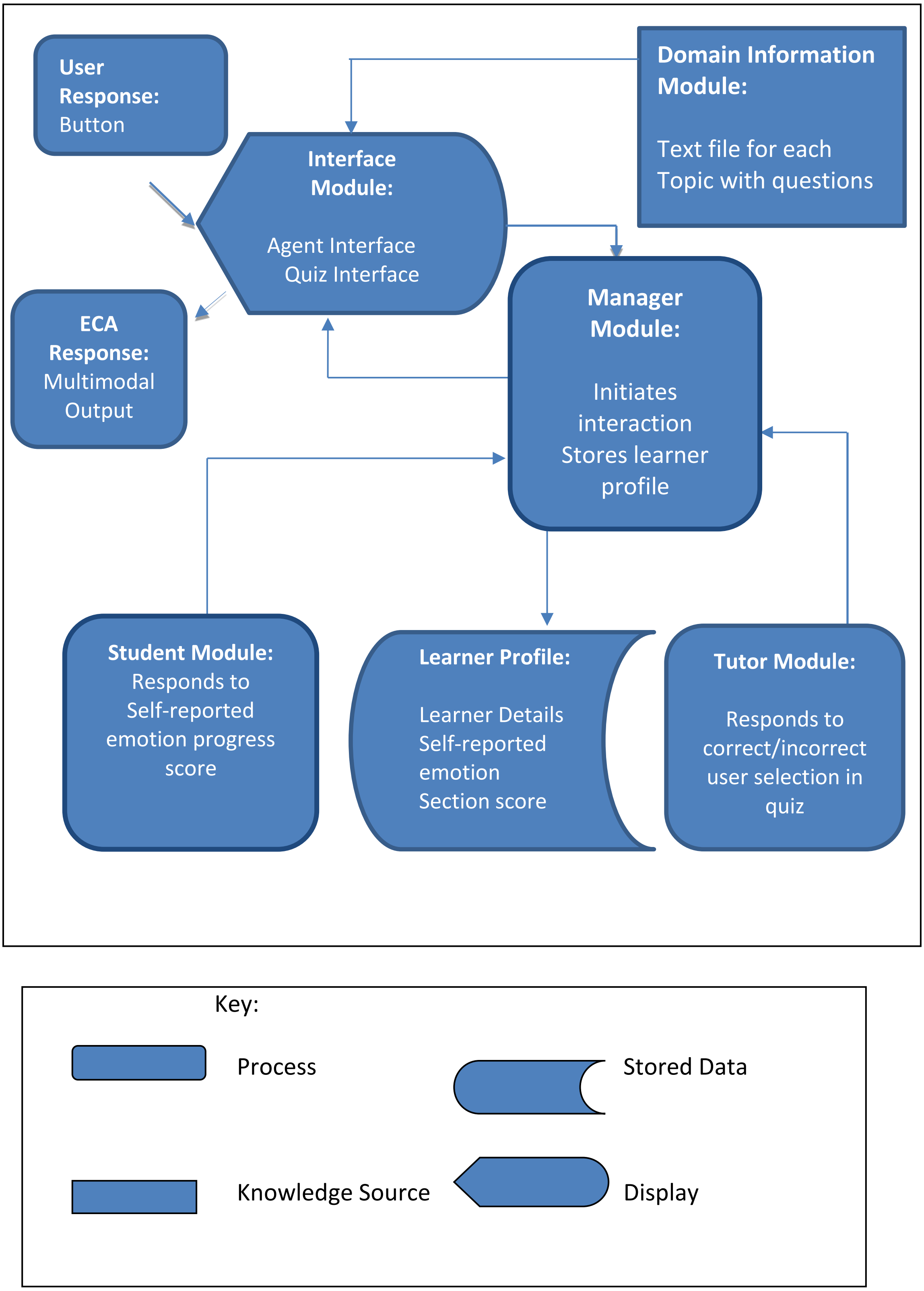
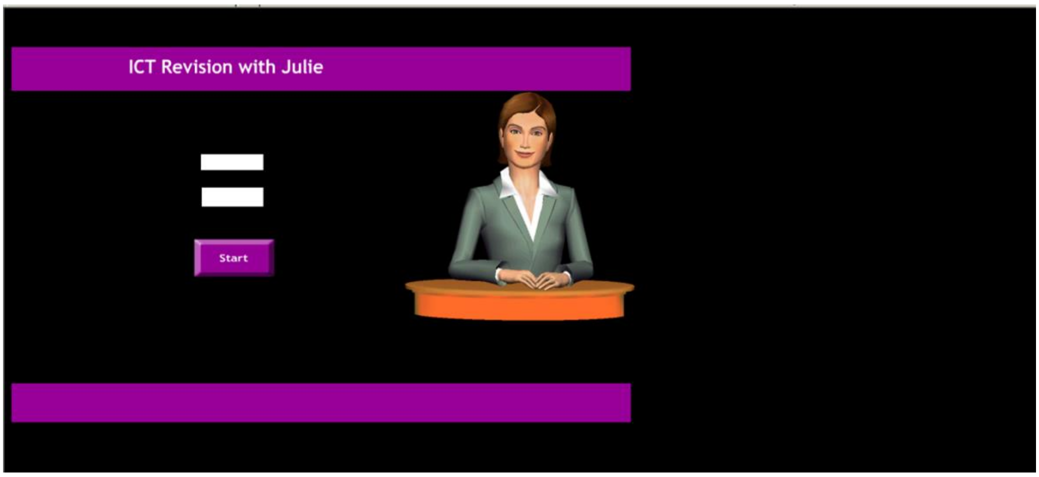


Figure 3.5 Empathic tutor interface



The Flash video includes four scenes for each topic of the quiz that in turn is split into a variety of sections. The first screen is the welcome view where learners are welcomed and requested to log in. User information such as log-ins and passwords are used as input to the Flash video. The information is in a text file that is ordered as follows:

Topic=user log in;user password;user name:

These log-in details are initialised into the Flash video and can then be used to establish the accuracy of log in details for every user of the system using two functions initialise users and then find user details (See Appendix 2 for further details).

Data is passed from the Flash video to the webpage using commands that include appropriate variables, for example: “fscommand (“Score”, 5)” refers to a user section score of 5. This data can be used by other modules within the tutoring system.

The second screen gives instructions to learners on how to use the system. The third screen is used to display each question and possible answers whilst the fourth screen prompts users to self-report emotion from a choice of four (excited, satisfied, frustrated or bored).

The final section for the empathic ETS prompts learners to select finish to save their scores. A "summary" command is invoked which saves the scores in a temporary form on the webpage and the submit button triggers the transfer to a permanent text file using PHP as detailed in Appendix 2.

3.3.2.3 Tutor module

The tutor module aims to provide appropriate feedback based on the correctness of the user's response to each multiple choice quiz question presented by the tutoring system. Each quiz question includes five possible answers with one correct answer.

Once a learner makes their selection and if their answer is correct the "right" command is invoked and this randomly selects the ECA response from a variety of positive verbal responses that are coupled with nonverbal behaviour that includes a "happy" facial expression and a "nod" acknowledging positive learner progress within the feedback as shown in Figure 3.6.

Figure 3.6 System response to correct answer - interface

When the user selects an incorrect answer the “wrong” command is invoked and this provides a variety of verbal cognitive responses indicating that the answer is incorrect. In addition the correct answer is given whilst the ECA’s non-verbal behaviour is that of concern at the lack of progress for the current question as described in Figure 3.7.

The system includes a number of variations when responding to a positive response for example: “Good Answer!” or “That is right!” or “You are right!” or “Correct!” Examples of a response to a wrong answer include: “That’s incorrect” or “Wrong answer” or “Not quite” or “That answer is incorrect”.

Figure 3.7 System response to incorrect answer – interface

3.3.2.4 Student module

The student module analyses the student's self-reported emotion in relation to their progress in a particular section. Section 3.2 describes in detail key theories that influence the implementation described in this subsection.

As previously mentioned, data is collected from self-report buttons to allow users to indicate their affective state at the end of a section after reading their section score. Users will press one button on request to indicate their current affective state from the following:

1. Excited
2. Motivated
3. Satisfied
4. Frustrated
5. Bored

Table 3.4 Pupil scores and metacognitive feedback

Learner Test Score (out of 5)	ECA Speech
High (>=4)	Confirmation
Middle (3)	Confirmation
Low (<3)	Suggest: revision increase effort

The ETS system affective response is primarily based on the emotion of the learner and this is coupled with metacognitive information based on learner progress. A weighted judgement on a student's progress is made by comparing the test result to pre-determined values to establish whether inadequate, adequate or good progress is being made as shown in Table 3.4 and illustrated in Figure 3.8. See Appendix 2 for detailed implementation of each emotion.

Figure 3.8 System responses to self-reported emotion and current score – interface

The user is given feedback that includes both affective and cognitive feedback at the end of each section, whilst the learning profile of the student is updated at the end of the interaction; the most recent test result and emotional state from each section is included in the learning profile.

3.3.2.5 Domain knowledge module

The material in the domain knowledge is divided into topics on Information Technology for pupils aged 12-16 years in line with the UK Key Stage 3 & 4 curriculum for IT for 2008-2011. These subjects were chosen because they are the subject areas taught by the thesis author. In addition, the theoretical aspects of IT are the focus of the studies conducted in this thesis as pupils in classrooms find it difficult to engage with these aspects of the course when compared to practical skills.

Every student completes the same topics using the same sequence. Each topic has subtopics which include multiple choice questions where only one answer is correct from a possible five answers. For example, users complete quizzes on “Types of Computers” an important topic within ICT. Chapters 5 & 6 include further details on the human evaluations with the developed system.

The quiz content is presented through the use of Flash MX and CS3 Professional. The input for the video is a plain text file that includes all relevant questions, answers and the position of the correct answer for a particular section. The text file data is ordered as follows:

Topic=question:answer;answer;answer:position of correct answer:

This data is loaded into the Flash video and split accordingly. When the quiz questions are initialised, each question is presented sequentially to the user who must choose the appropriate answer. Commands are triggered to the ECA depending on the correctness of each question and the tutor module responds appropriately. The user test score is stored at the end of each section and sent to the webpage for use by the student module to establish progress.

3.4 Comparison of ETS versions

The first version of the system (ETS v1) is evaluated in Chapter 5 whilst the second version of the system (ETS v2) includes a change within the Tutoring Module that is motivated by the results and discussion in Sections 5.6 and 5.7. The ECA affective speech is varied from five to fifteen possible responses for positive and negative feedback, as shown below in Table 3.5. Additionally, due to the length of time between the studies and availability, different female voices were used to record pre-recorded speech for each version of the system.

Table 3.5 ECA response to correct answer

Previous ECA Speech ETS Version 1	**Current ECA Speech in ETS Version 2**
That is a very good answer!	Good answer! Perfect!
Well done! You have got it right	That's right! Well done! It's right! You're right!
"Great! That is right",	Great answer!
"Yes! You are right",	Yes!
"Good! That is correct"	That's good! Good! Correct!
	Exactly! True! That's Spot-on! Absolutely right!

3.5 Summary

In this chapter, we have described the theory relating to Julie the ECA and the empathic ETS (version 1 and 2) used in subsequent chapters to complete human evaluations. We have described the two main sections of the system; the ECA "mind" that describes Julie's tutorial strategies and the ECA "body" describing the theories of implementing multimodal communication and The ETS architecture, which is similar to existing affective tutoring systems implementing multimodal ECA output.

Following on from this, we have described the implementation of the empathic ETS used in subsequent chapters to complete human evaluations. We have described the two main sections of the system; the ECA implementation and the tutoring system implementation.

The developed system is relevant to studies that implement affective tutorials systems that incorporate ECAs in formal learning environments such as classrooms. In Section 7.2 we will discuss the implications for other systems in greater detail.

Chapter 4 – Expressing emotion with ECAs

Our overall aim is to develop an effective empathic tutoring system. To achieve this, we need the tutoring system to accurately convey emotion. We have developed and implemented a strategy for expressing emotions with specific agent behaviours: speech, facial expression and gesture. In this chapter we report on an empirical study in which we investigate how successful this strategy is. In particular, we collected data on whether human users can accurately recognise the emotion that the ECA tries to convey. We generated four ECA emotions namely: excitement, satisfaction, concern and surprise. We compared a number of alternative strategies for each emotion (speech; speech and gesture; speech and facial expression; speech, facial expression and gesture).

The human evaluations required users to identify each generated expression and give subjective judgements on confidence of their emotion classification and emotion intensity in a within-subjects study. The results of this evaluation have been used in subsequent studies in this thesis as a basis for developing an accurate method of modelling ECA expression to learners using multimodal output that includes speech, facial expression and gesture.

This chapter is structured as follows. In Section 4.1 we begin by describing the implementation for empathic expressions using an ECA. Subsequently, in Section 4.2, we outline the evaluation study designed to assess the accuracy of varied combinations of communication channels that generate multimodal empathic expressions. Finally, Section 4.3 reviews the results of this evaluation study in the wider context of the thesis.

4.1 Implementing Emotions with ECAs

This section describes the implementation of multimodal output for the evaluation detailed in this chapter.

4.1.1 Speech condition

The first modality that has been developed for the ECA is affective speech; this is delivered by the ECA with a neutral facial expression with no gesture using pre-recorded speech that includes affective inflections on key words. The affective states include speech to reflect excitement, motivation, satisfaction, surprise and concern as described in Section 3.2.1.1. Table 4.1 includes examples of the affective speech for each emotion and the corresponding emotions in columns A & B.

4.1.2 Speech and facial expression condition

In this condition affective speech is combined with facial expression. Table 4.1 includes the set of facial expressions that are relevant to the empathic emotional states that we intend to use in subsequent studies and shows the ECA implementation of these emotional states in columns A, B and D.

4.1.3 Speech and gesture condition

The third condition of this study combines affective speech and gesture to establish the effectiveness of combining these two modalities. Gesture is incorporated into speech using nonverbal behaviour generation rules that are discussed in Section 3.2.1. Table 4.1 columns A, B and E illustrate this implementation whilst facial expression is neutral.

4.1.4 Speech, facial expression and gesture condition

Speech, facial expression and gesture are combined to develop the ECA's expression in the fourth condition. We have developed the three modalities discussed and matched each ECA affective state to affective speech, facial expression and relevant gestures in Table 4.1.

Table 4.1 Speech, facial expression & gesture condition

A. ECA Empathic State	B. ECA Affective Speech	C. ECA NVBG Rule	D. Facial Expression	E. ECA Software NVB/ Gesture	F.ECA
Excitement	We will cover many new and thrilling topics over a couple of sessions.	Inclusivity	Happy	Head tilt Explain	
Motivated	I hope to encourage all pupils to link practical skills in class to key terms in IT.	Possibility	Smile	Explain	

Table 4.1 Speech, facial expression & gesture condition (continued)

A. ECA Empathic State	B. ECA Affective Speech	C. ECA NVBG Rule	D. Facial Expression	E. ECA Software NVB/ Gesture	F.ECA
Satisfied	Regular revision helps to improve on how you apply your theory to practice.	Possibility	Neutral	Explain	
Concerned	Theory is a difficult concept for students, so I am here to help you	Affirmation	Concerned	Lean forward	
Surprised	It is amazing how many key words are known to students, but not fully understood.	Contrast	Surprised	Lean backward	

4.2 Evaluation of ECA emotion

This section describes the first study which measured the accuracy of expressed ECA emotions by having participants observe and try to identify this emotion. In addition participants gave subjective judgements on confidence in their judgement and perceived intensity of the emotion.

This study is similar to one conducted by Bickmore and Schulman (2007) to evaluate the accuracy of empathic feedback used to comfort users as discussed in Section 2.3.2. However there are some differences with the current study. Firstly, we are evaluating empathic feedback for a tutoring environment. Secondly, the comforting agent interacts with the participants whilst this study requires participants to identify the ECA expressed emotion only.

The results of the previous evaluation by Bickmore and Schulman (2007) suggest that an ECA that gives accurate empathic feedback when user emotion is restricted to a specific set of emotions, as compared to free speech input, leads to better outcomes. During an interaction accurate responses are more effective at comforting users than an ECA that gives inaccurate responses when user emotion is not restricted.

Results from a study conducted by Tolksdorf et al. (2008) who evaluated how recognisable facial expressions of the ECA Max are, showed that primary (basic) emotions such as happy and sad were identified more accurately than secondary (complex) emotions such as jealous and hopeful because the latter require subtle interpretation. Therefore secondary emotion is more difficult to express and recognise. They also showed that participants tended to use the “happy” label to identify the first positive stimulus presented. Their study suggested that facial expression alone was not sufficient to identify secondary emotions. Therefore this study will extend their findings.

Redundant communication channels may increase the likelihood of message comprehension during human to human conversation (Bickmore and Schulman 2007). Therefore, the most expressive condition should include the highest number of modalities to convey each of the empathic emotional states.

Therefore, participants in this study should identify the ECA's expression more accurately using speech, facial expression and gesture. Consequently, confirming the success of using an ECA communication strategy modelled against human-to-human communication in web-based tutoring. Following on from the results of the studies described above, we developed the following hypotheses for this study:

1. An ECA implementing multimodal behaviour using three channels (combining speech, facial expression and gesture) will generate expressions that users recognise better in comparison with an ECA that uses one or two channels to generate multimodal behaviour.
2. An ECA implementing multimodal agent behaviour using three channels will elicit higher judgements of confidence and intensity than an ECA that uses one or two channels to generate agent behaviour.

4.2.1 Pilot study

We carried out a small pilot study with four male colleagues who are teaching staff in the information technology department[1]. Each participant completed 20 interactions with the agent by identifying the agent emotion and giving judgements for confidence and intensity. The key finding from the pilot study is the participants' confusion in identifying "motivation" and "satisfaction".

[1] The Deanes School, a comprehensive secondary school in Essex, United Kingdom.

Therefore "motivation" was removed from the study described in the following subsections and four ECA emotional states were evaluated (excitement, satisfaction, concern and surprise).

4.2.2 Participants

Participants for this study were taken from pupils in a comprehensive school in Essex, UK. As part of their Information Technology (IT) lesson, these pupils were asked to interact with the developed system. The study was conducted with 39 students (22 males and 17 females) aged 12-16 years old. Participants were taken from Years 8 - 11. The study was approved by the ethics committee at the Open University and the board of Governors at the school.

4.2.3 Methodology

The study was run using a web-based interface. Participants were asked to open the appropriate resource and the ECA appeared to deliver an introductory speech before prompting participants to enter their log in details. A set of instructions appeared on the next screen. After reading theses instructions the participant pressed a button to continue.

During the study, participants were presented with pairs of screens for each emotion. On the first screen, the participant was asked to listen to and to observe the ECA. Participants could click a button to repeat the ECA actions before continuing to the next screen where they were prompted to identify the ECA's currently expressed emotional state from a choice of four states.

In addition each participant was asked to rate their confidence and the intensity of the emotion observed using a Likert scale by selecting an option from one to five (least to most). See Figure 4.1 and 4.2. These judgements were collected in order to provide a meaningful comparison using a quantitative statistical summary of subjective judgements to identified emotion, and establish the most successful experiment condition. One limitation of this method was that learners were unable to explain their choices, through open ended questioning.

The study was a within subjects' evaluation where all the participants observed each of the four emotions in each of the four conditions in the same order: speech vs. speech and facial expression vs. speech and gesture vs. speech, facial expression and gesture. Each participant viewed each emotion in the order shown in Table 4.2. The within subjects design was chosen for this experiment in order to provide a direct comparison between the experiment conditions for each emotion. In designing the experiment, it was important for users to identify the most accurate method of conveying an emotion in each condition to inform subsequent ETS studies on the best strategy to convey ECA emotion. This increased the number of observations for each experiment and increased the validity of the results in the study as the subjects were the same in every condition. However, a learning effect is a possible confounding variable of this type of evaluation. This may be mitigated by the fact that although learners saw the same emotions in each of the four conditions, emotions were not shown in the same sequence within each condition (see Table 4.2). A second limitation of the study is that the content of the speech was varied in each condition to convey the emotion being expressed.

Table 4.2 Order of interaction within each experiment condition

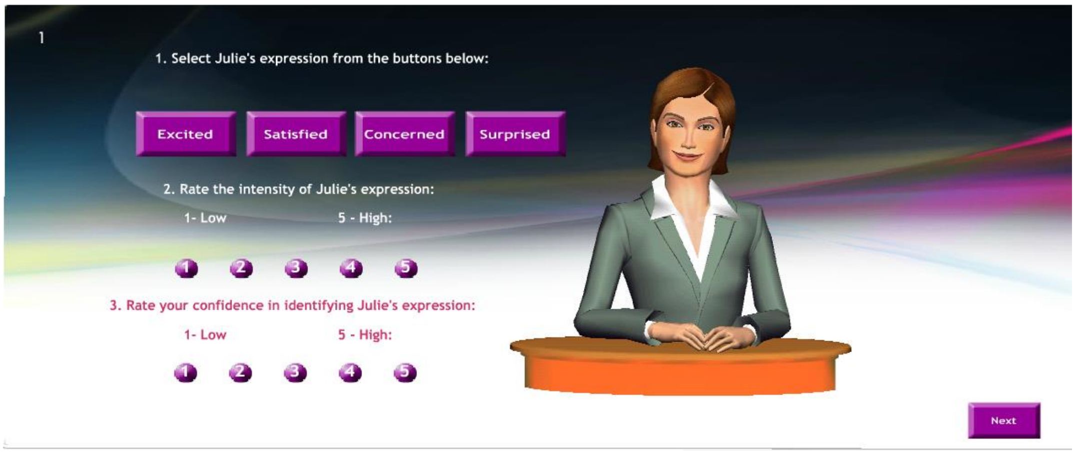
Study Condition			
Speech	**Speech &Facial Expression**	**Speech &Gesture**	**Speech, Facial Expression & Gesture**
1.Concerned	5. Satisfied	9. Satisfied	13. Surprised
2. Surprised	6. Excited	10. Concerned	14. Excited
3. Satisfied	7. Surprised	11. Surprised	15. Concerned
4. Excited	8. Concerned	12. Excited	16. Satisfied

4.2.4 Materials

We developed a series of ECA empathic emotional expressions that imitate typical human tutor responses to learner emotion. The interaction was developed using JavaScript and Flash MX to generate a simple interface that included two main screens. The first screen included a clear screen as a background to the ECA, with two buttons “next” and “repeat” whilst the second screen included a prompt for the three questions: see Fig 4.1 and 4.2.

Figure 4.1 ECA expressing emotion

Figure 4.2 Participant prompted for selections



4.2.5 Results

The following section describes the results and analysis of the first study conducted with Information Technology pupils. This study evaluates how accurately pupils can identify the ECA emotion expression.

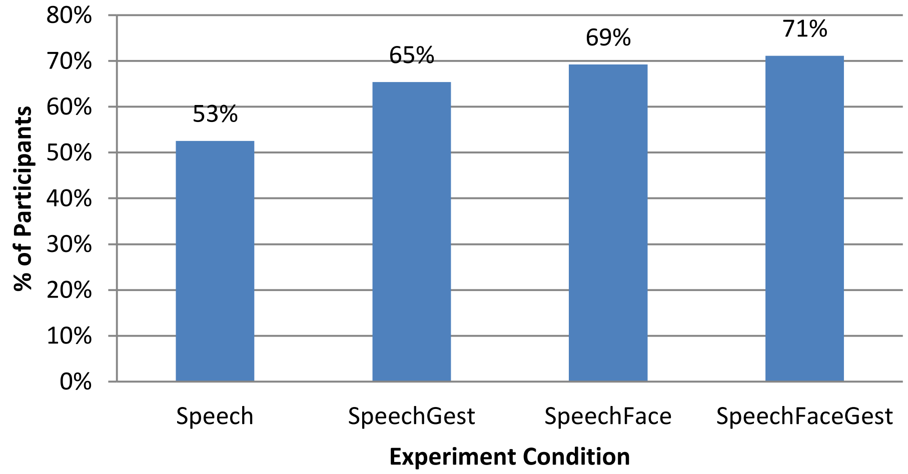
4.2.5.1 Hypothesis results

Hypothesis 1: An ECA implementing multimodal behaviour using three channels (combining speech, facial expression and gesture) will generate expressions that users recognise better in comparison with an ECA that uses one or two channels to generate multimodal behaviour.

Hypothesis 1 result: Confirmed

The first hypothesis investigated the accuracy of the multimodal algorithm we had developed that used three channels. The chi-square statistic, X^2 (3, N =624) = 14.31, p = .03, indicates that each study condition has a significant effect on correctly identified emotions when compared to incorrectly identified emotions. Figure 4.3 illustrates that the speech, facial expression and gesture condition has the highest accuracy at 71% whilst the speech condition has the lowest accuracy on identified emotions at 53%.

Figure 4.3 Correctly identified ECA emotions

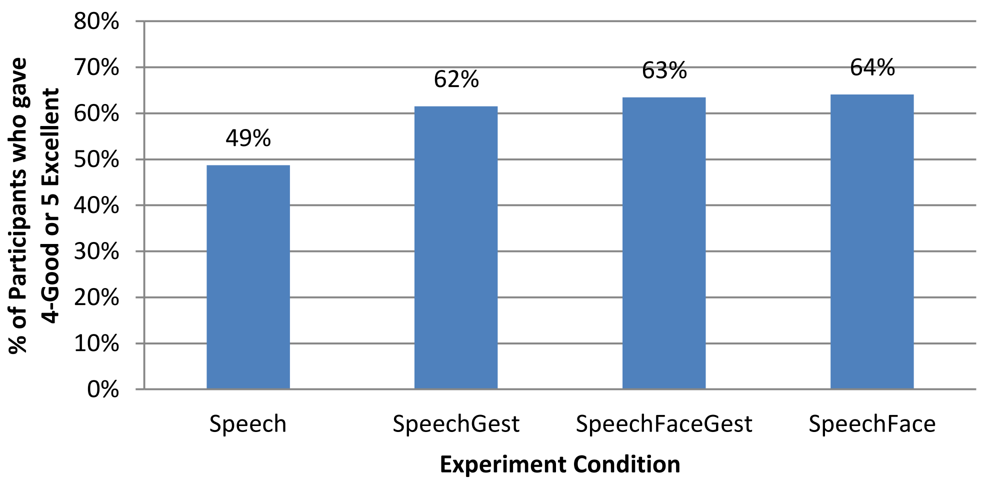


Hypothesis 2:- An ECA implementing multimodal agent behaviour will elicit higher judgements of confidence and intensity than an ECA that uses one channel to generate agent behaviour.

Hypothesis 2 result: Not confirmed

Our second hypothesis investigated the confidence and intensity judgements of a multimodal agent when compared to an ECA using one channel. Figure 4.4 shows the confidence judgements from learners for the four emotions in each condition. The speech and facial expression condition has the highest average confidence judgements (users who gave a judgment of 4 - Good or 5 - Excellent) at 64% of participants whilst the speech condition has the lowest average confidence judgements of 49%. Whilst the differences shown are not statistical significant, the Wilcoxon Signed-Rank Test did identify significantly higher confidence judgements for satisfied when comparing speech vs. speech, facial expression and gesture ($z=-2.094$, $p<0.05$).

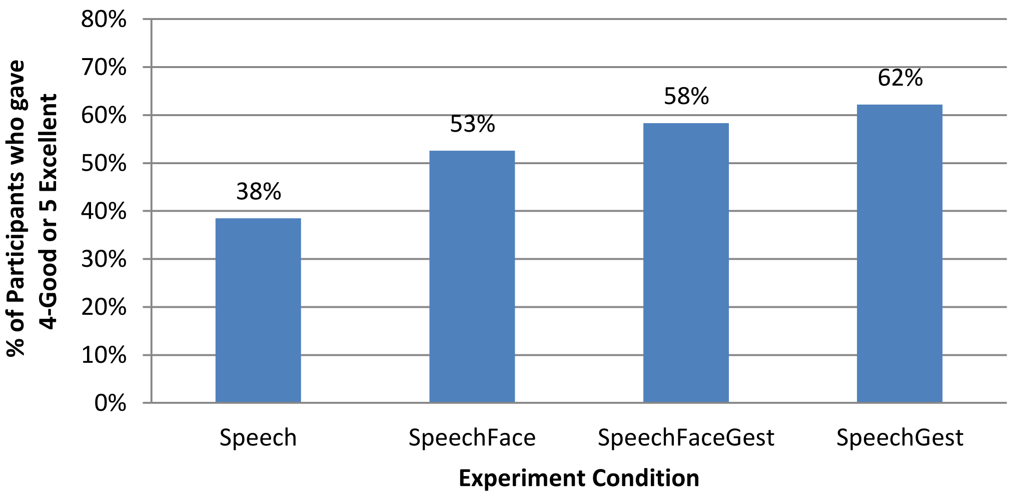
Figure 4.4 Average confidence judgements



The average intensity judgements (for users who gave a judgement of 4 - Good or 5 - Excellent) for this evaluation are shown in Figure 4.5 where the speech and gesture condition has the highest average intensity judgements of 62% closely followed by the speech, facial expression and gesture condition at 58% whilst the lowest is speech at 38%.

The speech, facial expression and gesture condition has a significantly higher intensity rating for excitement, satisfaction and surprise (z=-2.423, -2.473. -2.074, $p<0.05$) in comparison to the speech condition as measured by the Wilcoxon Signed-Rank Test. Other results show that differences in concern are not statistically significant.

Figure 4.5 Average intensity judgements



4.2.5.2 Other results

Pupils were least likely to use the repeat button in the speech, facial expression gesture condition (at 4%) whilst pupils were most likely to use the repeat button in the speech condition at 19% as shown in Figure 4.6.

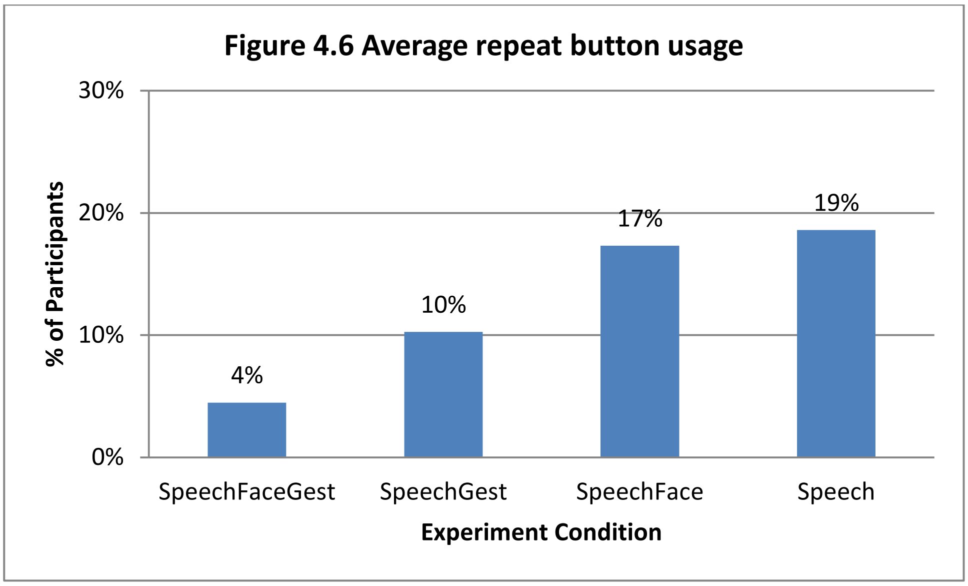
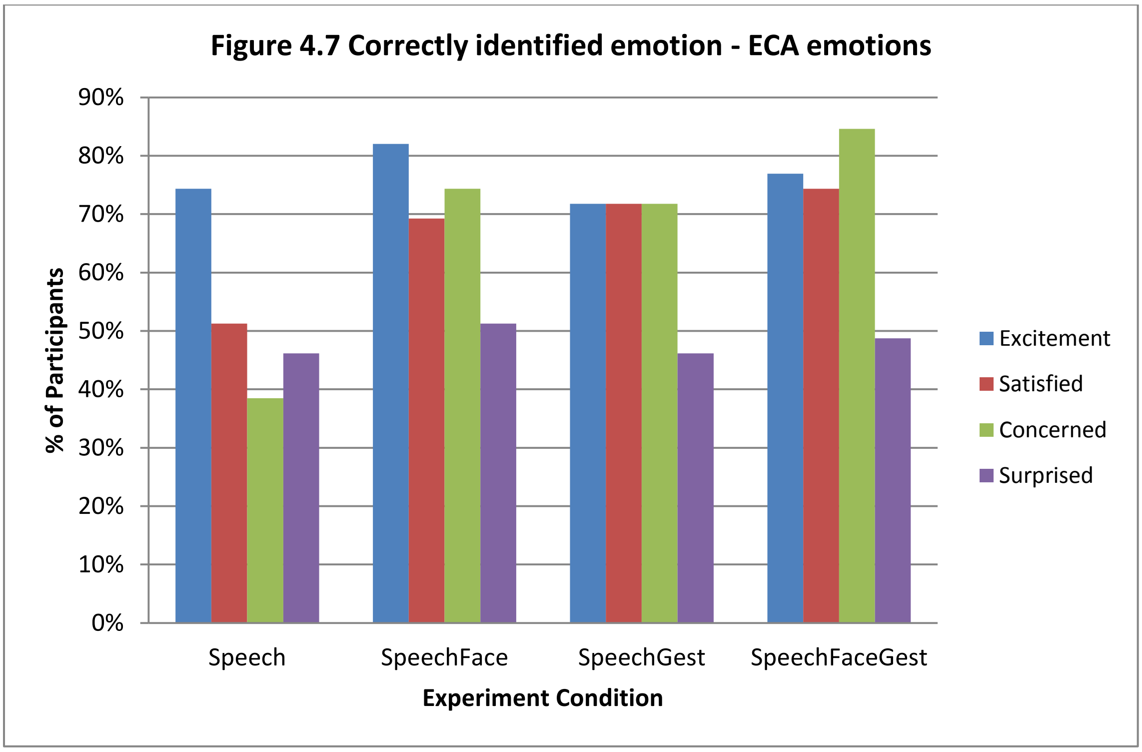


Figure 4.7 illustrates the differences between conditions and ECA emotions. In the speech facial expression and gesture condition, excitement, satisfied and concerned are correctly identified more often than surprised whilst in the speech condition excitement is correctly identified more often than satisfied, surprised and concerned.



Excitement is measured to have the highest average accuracy in all conditions particularly in the speech and facial expression condition whilst surprised appears to have the lowest average accuracy judgements across all conditions of the study. The high accuracy in identifying excitement which can be classed as a primary emotion is similar to findings by Tolksdorf et al. (2008).

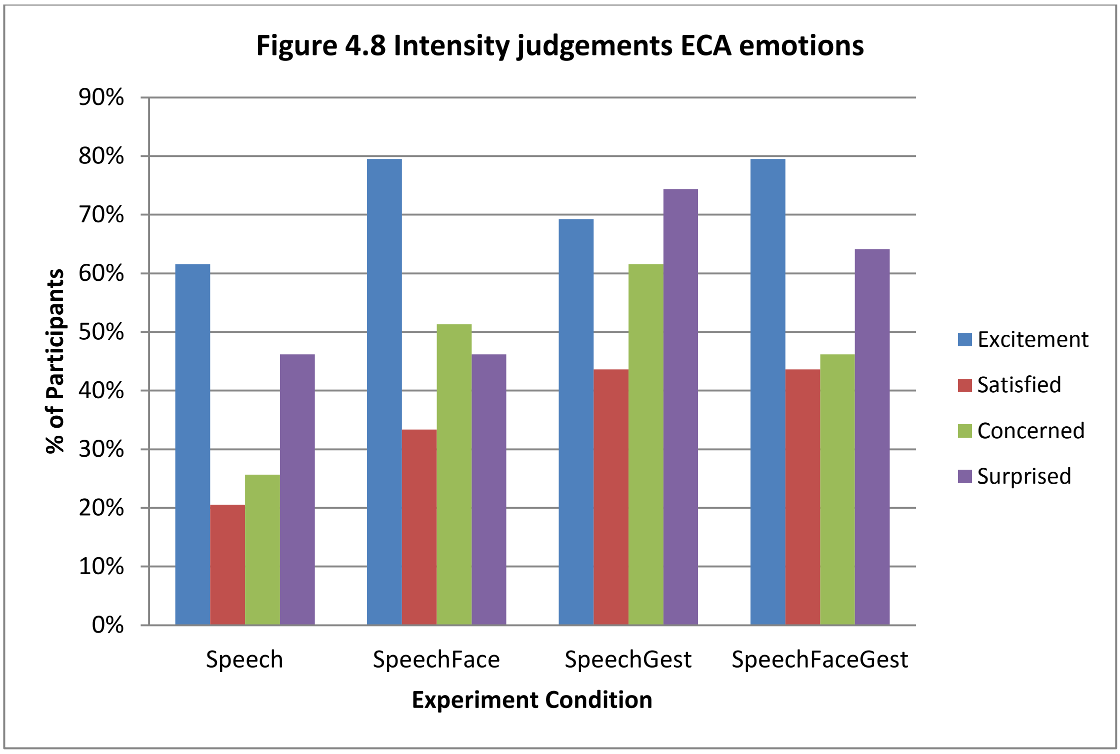
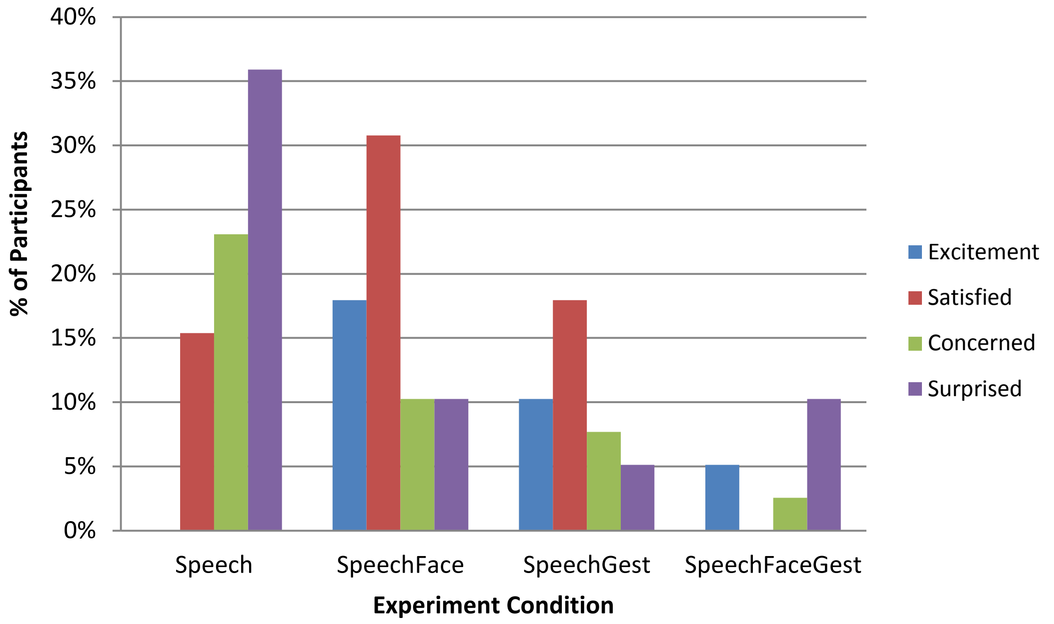


Figure 4.9 Repeat button usage - ECA emotions



Furthermore confidence judgements and intensity judgements show that excitement has higher average judgements across all conditions when compared with other emotions whilst surprise and satisfied have the lowest average confidence judgements. In addition, satisfied achieved the lowest intensity judgements in every condition as shown in Figure 4.8. The repeat usage button was used by learners to identify satisfied in three of the four conditions except the multimodal condition with three channels suggesting that this condition best expressed the emotion satisfied as shown in Figure 4.9.

Table 4.3 shows the Agent emotion in a confusion matrix in comparison to the emotion identified by participants in the speech condition. Concern is identified by 51% of participants as satisfaction indicating that these emotions are difficult to differentiate in the speech condition. Surprise is incorrectly identified in the speech facial expression gesture condition by 31% of participants as concern. In addition, satisfaction is identified by 21% of participants as concern (see Table 4.4 for details). These results suggest that further improvements on the accuracy of the ECA emotion expressed can be investigated in future studies that can investigate specific emotions and how their expression can be improved.

Table 4.3 Speech condition

		Agent emotion expressed			
		Excited	Satisfied	Concerned	Surprised
Emotion Identified by Participant	Excited	74%	5%	10%	21%
	Satisfied	13%	51%	51%	10%
	Concerned	5%	31%	38%	23%
	Surprised	8%	13%	0%	46%

Table 4.4 Speech, facial expression and gesture condition

		Agent emotion expressed			
		Excited	Satisfied	Concerned	Surprised
Emotion Identified by Participant	Excited	77%	3%	0%	10%
	Satisfied	15%	74%	10%	10%
	Concerned	3%	21%	85%	31%
	Surprised	5%	3%	5%	49%

4.2.6 Discussion

Our first hypothesis investigated the accuracy of the multimodal algorithm we had developed that used three channels. The results from the previous section confirmed that the ECA implementing multimodal agent behaviour generated expressions that users identified with higher accuracy in comparison with an ECA that generates agent behaviour using one channel. This study extends the results described by Tolksdorf et al. (2008) who suggested that more than one channel is required to correctly identify secondary emotions, such as concern, which are more complex.

The confidence judgements for this study were highest in the speech and facial expression condition. Furthermore, the intensity judgements were highest in the speech and gesture version. However, significantly higher differences for intensity judgements were identified in three out of the four emotions in the study namely excitement, satisfaction and surprise for the multimodal condition using three channels when compared with the speech condition. These results partially support our second hypothesis as speech, facial expression and gesture was expected to give higher confidence and intensity judgements across every emotion. Additionally, the speech, facial expression and gesture condition had the lowest mean for repeat button usage indicating confidence in the emotion identified albeit with the possibility of a learning effect as users completed the interaction.

The overall conclusion from this study is that participants accurately identified ECA emotions expressed in the speech, facial expression and gesture condition when compared to other conditions. In addition, they rated the same condition to have significantly higher intensity judgements for 3 out of 4 emotions when compared to the speech condition. However, participants rated emotions more confidently for satisfied in the speech, facial expression and gesture condition when compared to speech only, confirming the overall use of a multimodal strategy in ECAs to enhance the accuracy and intensity in expressed emotions.

There were some results such as poor accuracy when identifying surprise that must be considered. Although this is beyond the scope of this thesis, further studies can investigate a variety of combinations of how an ECA can express a specific emotion to make it easier for users to identify. This is important because participants achieve higher learning gains when they perceive that they are receiving accurate feedback in a learning context leading to better outcomes as discussed by Bickmore and Schulman (2007). Furthermore, future research can investigate the impact of individual channels as this is beyond the scope of the thesis activities.

4.3 Summary

This chapter described how multimodal affective feedback, implemented using a rule-based algorithm, can be used to express ECA emotions. Following on from this, we presented the results of an evaluation where participants had to accurately identify the ECA’s expression in four conditions and give confidence and intensity judgements for each emotion. The study looked at the impact of multimodal output on the accuracy of ECA expressions that were generated.

The overall results of the study show that a multimodal strategy that includes speech, facial expression and gesture generates ECA expressions that can be identified more accurately in comparison to an ECA using speech only. These results extend those obtained by Tolksdorf et al. (2008). The intensity judgements for the speech, facial expression and gesture condition are significantly higher for excitement, satisfaction and surprise when compared to speech alone suggesting that increasing the number of modalities can improve intensity. However confidence judgements appear to be significantly higher with only one of the four emotions, satisfied, when speech facial expression and gesture condition was compared to the speech condition.

The results of this study address the first research question: Which combination of channels used to portray ECA emotions are users able to identify most accurately? In this study, participants were able to correctly identify the ECA emotion making this a first step towards using this strategy to develop an empathic tutoring system that can impact positively on interactions. The next steps will be to develop and evaluate an empathic ECA tutor that uses multimodal feedback to generate empathic emotions. This is discussed in the following chapter.

These results contribute to the goals of the thesis by identifying a strategy that can be used by ECAs to improve the accuracy of communication with users. This extends previous evaluations that only looked at the impact of facial expression and provides evidence that generating multimodal ECA expressions that include facial expression, *speech and gesture* can enhance the accuracy of communication in human-computer interactions. As this study focusses on accurate communication between the user and the ECA, these results could apply to other systems that model human-to-human communication using ECAs**.**

Chapter 5 – Evaluating the ETS version 1

In our second set of studies we investigate whether a combined reactive and parallel feedback strategy gives better learning gains than a neutral strategy within a web-based tutorial environment. In our feedback strategy, empathy is used within an affective tutoring strategy that targets learner emotion whilst in a learning environment. Section 2.3.1 describes in detail a number of studies that successfully used this strategy in a similar manner to the current study to alleviate negative emotion within a computer-to-human tutoring context. This chapter describes the study that was conducted to establish the impact of this empathic feedback strategy, extended to reinforce positive emotions on learning outcomes within a formal educational setting using the ETS version 1.

This chapter addresses the second research question of this thesis: Can an empathic feedback strategy positively affect learner emotions whilst interacting with a tutoring system? In this study we evaluate an empathic tutoring strategy for an ECA that uses the multimodal algorithm described in the previous chapter to generate accurate empathic expressions whilst interacting with learners during a set of multiple choice quiz questions. In the evaluation learning gain was measured and subjective user judgements were elicited. Surprisingly, the results of this study do not establish any difference between the empathic feedback strategy when compared to a neutral feedback strategy. Nevertheless, we have learnt useful lessons that will be used in subsequent studies to enhance the impact of empathic tutorial feedback on learning outcomes in web-based tutoring systems.

This chapter is structured as follows. Section 5.1 describes the hypothesis being investigated whilst Section 5.2 discusses a short pilot study. Sections 5.3 – 5.5 describe the participants, methodology and materials of the study, respectively. Section 5.6 describes the results of this study and an assessment of the empathic feedback strategy is discussed in Section 5.7. Finally, in Section 5.8 we review the impact of the results of this evaluation study in the wider context of the thesis.

5.1 Hypothesis

This study is comparable to a previous study on modelling and evaluating parallel and reactive empathy in virtual agents conducted by McQuiggan and Lester (2006) and McQuiggan et al. (2008); which we discuss in Section 2.3.1. Other similarities with the current study include providing affective and task related support. There are also some notable differences. Whilst the previously mentioned authors are using an inductive method with training data to model empathy we are using a theoretical method based on Davis' theory (1994). In addition we are evaluating our empathic feedback strategy within a web-based tutorial environment with a single ECA similar to a study by Burleson (2006) whilst McQuiggan et al. (2008) have evaluated their CARE model in Crystal Island which is a narrative enquiry based interactive learning environment with multiple ECAs.

Following on from the discussion above, participants in this study should make progress in their learning and positively rate their learning experience confirming the success of parallel and reactive empathy in an empathic feedback strategy. The results of the studies described above inform the hypothesis for this evaluation:

1. An ECA implementing a parallel empathic strategy in response to positive self-reported user emotion, and reactive empathic strategy in response to negative self-reported user emotion, will improve learning in a web-based quiz system in comparison to a neutral strategy.

5.2 Pilot study

We carried out a small pilot study during April 2011 with two fourteen year old pupils during their IT lesson. The pupils were briefed about the study and given a demonstration on how to use the developed system. Each participant interacted with the developed system and completed a short questionnaire on their interaction. In this pilot, we found that the participants were not clear about the number of questions or sections within the interaction. As a result we included bread crumb information to help learners track where they were and how many questions or sections were remaining before completing the developed system, see Figure 5.1 below.

Figure 5.1 Section and number of questions added to quiz system

5.3 Participants

This study was conducted with two classes of 51 students in total aged 12-16 years old in a local comprehensive school. The students were randomly assigned by gender with 26 students in the neutral version and 25 students in the empathic version, with incomplete data sets due to absence (42 students, 19 – neutral, 23 – empathic), see Table 5.1.

Table 5.1 Study participants

Condition	Female	Male
Neutral	11	8
Empathic	12	11

As in the previous chapter learners participated in the study during their normal IT lessons over a number of weeks. The studies described here and in chapter 6 were approved by the ethics committee at the Open University and by the board of Governors at the school.

5.4 Methodology

Participants were told that the purpose of the study is to evaluate the use of animated characters in web-based quizzes to revise theoretical concepts in IT. During a typical lesson, participants were asked to complete a short pre-test in the domain of IT (see Appendix 3). A pre-test is a reliable method of measuring learning prior to interaction with the developed ETS. It was carefully designed and administered prior to interaction to reduce the chance of a learning effect prior to participant interaction with the ETS.

Approximately 4-6 weeks later, each participant interacted with one condition in a balanced between subjects study (see guidance sheet in Appendix 3). This research design was chosen in order to minimize any learning effects as learning gain was being measured. It allowed the comparison of groups of learners in the two experiment conditions. A drawback of this method is that the comparable groups within each experiment are similar but not identical which may impact on the results of the study.

Learners were given a new seating plan which alternated users between the empathic ETS and the neutral ETS, in addition, to being given clear verbal and written instructions on expectations on behaviour similar to test conditions (i.e. work in silence independently). Furthermore, the use of earphones or headphones allowed learners to engage with the ETS on their own screen.

Participants were asked to navigate to a link on the shared drive of their school network. They were instructed to open the web-based resource and the ECA appeared and delivered an introductory speech before prompting participants to enter their log in details. Participants completed the multiple choice quiz which was divided into four sections with five questions each.

Immediately after completing the quiz interaction, participants are asked to complete a short post-test in IT, which is a reliable method to measure learning gain after an interaction. Traditionally, participants regularly sit written examinations to measure summative learning after teaching and learning activities and as practice for final terminal public examinations. However, since the post-test is completed immediately after the interaction, it does not measure delayed transfer of information.

Finally, learners were asked to complete a short online survey on their interactions with the system. The *online* survey was chosen because it was an accurate and reliable method of obtaining user feedback on their subjective judgements immediately after interacting with the ETS. Further to this, collation and analysis was conducted quickly because information is easily transferrable to other mediums such as statistical packages. The online survey included a mixed design of closed questions to allow quantitative statistical analysis whilst open questions allowed qualitative analysis to obtain individual participant comments. This design allowed us to gather numerical information to easily summarise our findings and also allowed us to analyse specific examples of participants' observations when compared to learning gain to triangulate our results and draw accurate conclusions.

5.5 Materials

We have developed a set of materials to investigate our hypothesis. The ETS is based on a multiple-choice quiz environment that has been developed as an interactive teaching and learning tool. In addition we have developed a written pre and post-test to assess learning gain before and after user interaction with the ETS. Finally we have created an online survey to record user subjective judgements.

5.5.1 Pre-test and post-test

The written pre-test and post-test (see Appendix 3) has been developed to test learners on their domain knowledge of IT on the topic “Types of Computers” as a summative measure of learning in the domain of IT. The test includes low ability questions that require factual information. In addition, the test also includes middle ability questions requiring users to give examples of concepts and explain how concepts are used in industry. These short answer questions are based on the information provided in the ETS as multiple choice questions. The pre-test and post-test includes the same questions. However to minimise a learning effect, learners are given the pre-test at least 6 weeks prior to completing the rest of the study and they do not receive feedback on their results.

5.5.2 ETS version 1

We have developed the ETS as a short online web-based quiz environment that is described in Section 3.3 and the full implementation included in Appendix 3. The ETS acts as a teaching and learning exercise that uses formative assessment to give learners feedback on current progress.

In the empathic condition, when pupils gave a correct/incorrect answer for each multiple choice question the ETS gave an appropriate empathic response with a variety of appropriate non-verbal behaviour that includes gesture and facial expression. At the end of a section the ETS gave relevant empathic and metacognitive feedback related to self-reported emotion and score.

In the neutral condition, pupils received cognitive feedback that was accompanied by a neutral emotional expression and neutral affective feedback with no gesture. At the end of a section pupils are given neutral metacognitive feedback based on self-reported emotion and score.

5.5.3 Online Questionnaire

The online questionnaire was developed to obtain subjective user judgements from learners who interacted with the system. The survey included questions to obtain learners' personal information such as IT class, age and gender, however study condition and user name was not collected. In addition, closed questions based on a 5-point Likert scale (1 - Strongly Disagree – 5 Strongly Agree) were used to elicit detailed judgements on agent characteristics and the interaction itself. Finally a group of open-ended questions were used to obtain learner subjective comments on their interaction and "Julie" the ECA's characteristics, see Appendix 3 for details.

5.6 Results

5.6.1 Results of ETS version 1 on learning gain

Hypothesis 1: An ECA implementing a parallel empathic strategy in response to positive self-reported user emotion, and reactive empathic strategy in response to negative self-reported user emotion, will improve learning in a web-based quiz system in comparison to a neutral strategy.

Hypothesis 1 Result:-Unconfirmed

Our hypothesis investigated the success of an ECA implementing a parallel empathic strategy in response to positive self-reported user emotion, and reactive empathic strategy in response to negative self-reported user emotion in comparison to a neutral strategy. Both neutral and empathic conditions appear to significantly improve learning, as measured by the difference between pre-test and post-test scores, by an average 7.84 and 7.83 marks ($p<0.01$) respectively using the two-tailed t-Test. However, according to the one way ANOVA test on learning gain there is no significant difference between the two conditions. These findings do not support our overall hypothesis that an ETS using a parallel and reactive empathic strategy should improve learning with higher learning gains when compared to a neutral strategy.

5.6.2 Results of subjective user judgements

The questionnaire yielded the subjective judgements for the learners in our study. We used the ANOVA statistical test to establish significance. We did not collect specific user information such as name and condition during this study, which was a limitation. From Table 5.2 both groups of learners agreed that using the system was a good way to learn. However, learners who noticed Julie's expression agreed that Julie's facial expressions and gestures were appropriate, believable and a preferred way to learn when compared to learners who had not noticed Julie's facial expressions, see Table 5.3.

Table 5.2 Questionnaire results grouped by whether participants noticed Julie's facial expressions

Survey Question	Participants who noticed Julie's facial expression mean judgements	Participants who did not notice Julie's facial expression mean judgements
Q1 Good way to learn	3.14	3.14
*Q7 Facial expressions appropriate	3.29	2.43
*Q10 Julie's gestures are believable	3.33	2.62
*Q11 Julie's gestures are appropriate	3.24	2.57
Q17 Prefer to learn with a Character	3.24	2.52

*Statistically significant at p<0.05

Table 5.3 Questionnaire results grouped by whether participants had problems interacting with the system

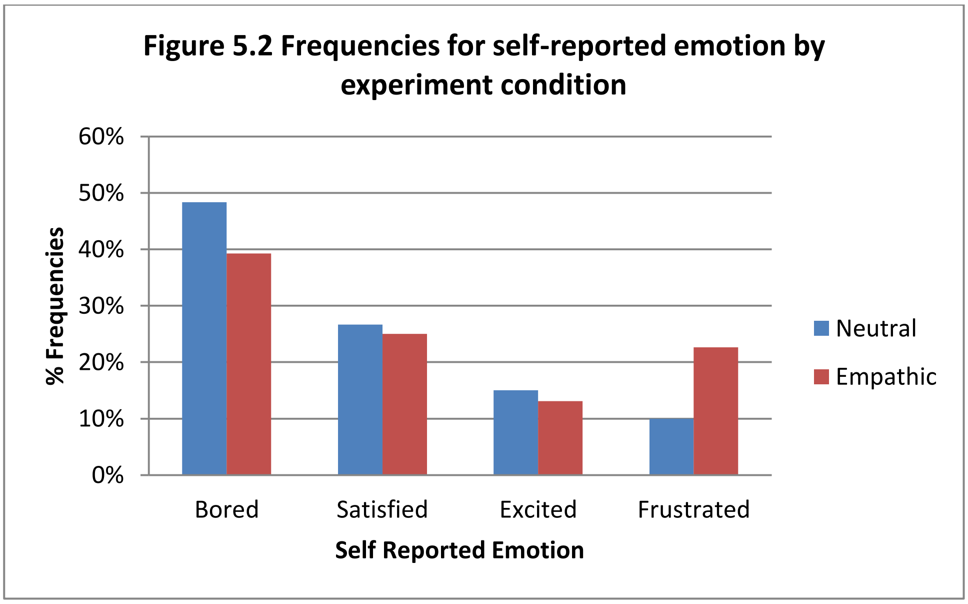
Survey Question	Participants who had problems mean judgements	Participants who did not have problems mean judgements
*Q1 Good way to learn	2.65	3.72
*Q4 Use system regularly	2.10	2.94
*Q18 Enjoy this type of activity	2.20	3.33

*Statistically significant at p<0.05

Some of the learners in this study were understandably upset because of the technical difficulties they encountered. Others did not report any technical difficulties although they considered the system slow. Technical difficulties occurred during interactions in both conditions in relation to the flash animation “freezing” or not submitting scores, and learners were moved to other machines where possible. These technical difficulties may be related to the configuration of these computers.

5.6.3 Results of self-reported emotion

During the study we automatically collected data on the score and the self-reported emotion for each section. We did not obtain any statistically significant data about the relationship between emotion and learning gain, which may be as a result of the incomplete data set collected (neutral = 15, empathic = 21), Figure 5.2 summarizes our findings:

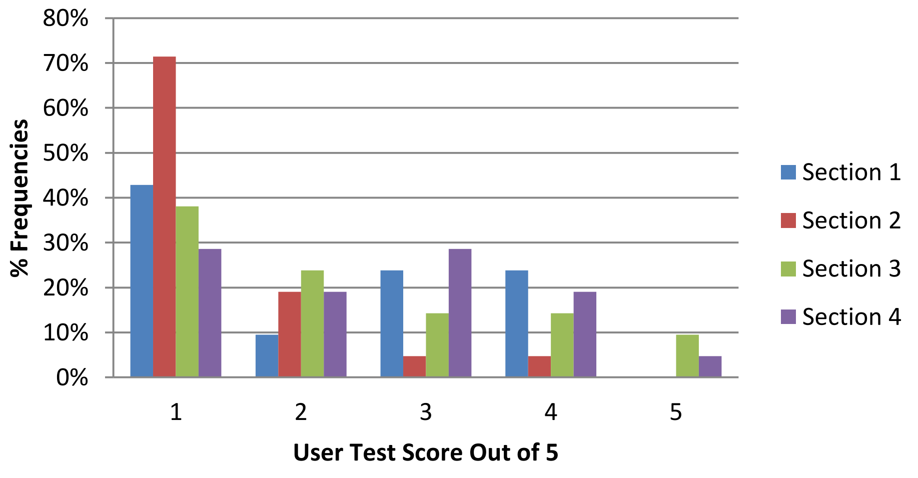


Learners appeared to self-report boredom most often especially in the neutral condition of the system. Surprisingly, learners were most frustrated in the empathic condition in comparison to the neutral condition. Satisfaction was the second highest emotion reported for both conditions and this was slightly higher in the neutral condition.

Learners could score from 0 up to 5 marks for each section. Most learners had low scores with the most frequent score for each section being 1. Section three and four appear to have higher scores suggesting more confidence from learners or more familiar material (see Figure 5.3).

The empathic strategy used parallel and reactive empathy to respond to self-reported learner emotion. High levels of negative emotion (i.e. frustration and boredom) were reported by participants in both conditions although this is higher in the empathic condition. This may be related to the increased duration of the feedback as well as technical difficulties on the day.

Figure 5.3 Frequencies for user test scores



This may have had an impact on the participants learning which was at times interrupted or delayed in addition to hampering data collection. Another frustration for all participants may relate to the number of interruptions suggesting that dividing the quiz interaction into sections after every 5th question, to obtain self-reported user emotion and for the ETS to respond appropriately, may reduce the “flow” or optimal learning state of the student which should be encouraged (Csikszentmihalyi 1990).

5.7 Discussion

These results show that participants in both conditions of the study significantly improved their learning when pre-test scores are compared to post-test scores. Although our overall hypothesis has not been confirmed, these results suggest an improvement in comparison to the pilot study described in Chapter 3. We implemented an affective & cognitive strategy vs. a neutral strategy with Julie. The difference between the two was not statistically significant when comparing pre-test and post-test scores for both conditions.

Although the current study has shown improved learning gains, the empathic ETS does not significantly improve learning gain when compared to the neutral ETS. However, in the study we did implement a specific empathic affective strategy in addition to addressing some technical issues such as using pre-recorded speech and improving timing.

Overall, learners from, both conditions, who had no problems with the system gave significantly higher subjective judgements, as they agreed that the system was enjoyable, a good way to learn and would be willing to use the system regularly.

Participant open-ended comments on Julie and the ETS ranged from positive learning experiences to negative learning experiences. Participants considered Julie “annoying, especially her voice” while most acknowledged that she helped them with the answer and this method of revising was better than reading a book or having “a boring teacher moaning”. Some learners appreciated the fact that Julie did not necessarily accept their self-reported emotion and sometimes reminded them to improve their effort or attainment. Some learners appeared to enjoy learning this way while others found Julie’s voice “annoying” and “repetitive”. This suggests that future systems should consider increasing the variation of statements from five in the current system.

Learners in the empathic condition in the pilot study, summarised in Chapter 3 and included in Appendix 1, and learners in the current study described in this study feel that the system is a useful way to revise in and out of lessons in comparison to the neutral condition in both studies.

5.8 Summary

This chapter described how empathy can be used within tutorial feedback in a web-based tutoring environment that uses ECAs. The study described in this chapter looked at the impact of parallel and reactive empathy as a feedback strategy within the ETS version 1. We presented the results of an evaluation where learning gain and subjective user judgements where measured for participants interacting with an ECA.

This second study addresses the second research question: Can an empathic feedback strategy positively affect learner emotions whilst interacting with a tutoring system? Participants in this study were able to show significant learning gains therefore showing that using an empathic feedback strategy that includes parallel and reactive empathy can impact positively on learning outcomes with ECAs. However, this empathic feedback strategy was not proved to improve on a neutral strategy, disproving the original hypothesis as no differences were found. Further to this, subjective measures from our survey may suggest that the multimodal feedback strategy generates appropriate and believable facial expressions, gestures and speech within a learning environment. Technical difficulties may have had an impact on the study and a future study may need to be conducted to confirm the results of this study.

Subsequent studies need to address the technical difficulties encountered. In addition, subjective user judgements suggest the increase of variation and reducing the duration of feedback to avoid the perception of repetition by the learner to ensure that the focus remains on the learning activities they are engaged in. The next step will be to enhance the tutoring system and evaluate the enhanced empathic ECA tutor with more learners. The resulting study is described in Chapter 6.

The activities in this chapter contribute to the goals of the thesis by demonstrating the impact of the empathic feedback strategy that uses parallel and reactive empathy on learning outcomes. The results of this evaluation extend previous evaluations that have not been conducted in formal educational settings such as a classroom situation and provide further evidence on the impact of parallel and reactive empathy on learning in quiz-based learning environments in particular.

Chapter 6 – Evaluating the ETS version 2

This chapter describes the third study where we evaluate the impact of the enhanced empathic ETS in the domain of Information Technology (IT). The evaluation consisted of a comprehensive study to establish the impact on learning and subjective user judgements taking into account different types of learners. Learners' empathic tendencies are established and used to investigate their impact on learning gains; Section 2.2.3 discusses the use of psychological assessments to establish empathic tendencies of participants. This chapter describes this final evaluation whose results do not suggest an overall effect but indicate possible groups of learners who may benefit from interaction with an empathic ETS in a classroom environment.

We revisit the third research question which looks at the impact of an empathic ETS on learning gains in computer-based learning. In this study we conduct a detailed evaluation based on the Empathic ETS that has been developed and enhanced based on the studies described in Chapters 4 and 5. The results of this study have been used to confirm previous results and suggest enhancements to the developed empathic tutoring system in addition to future studies to improve learning outcomes in affective tutoring systems.

The implementation for the empathic ETS that is being described in this chapter is described in detail in Chapter 3 (Sections 3.3 and 3.4).

The subsequent sections are structured as follows: Section 6.1 gives a detailed overview on the evaluation studies whilst Sections 6.2 – 6.4 describe the participants, methodology and materials used for the study. We revisit the hypotheses investigated by these evaluations and describe the results of the studies in Section 6.5. Discussion of the implications is presented in Section 6.6. We conclude with Section 6.7 where we discuss the implications of these results for future studies and in the wider context of the thesis objectives.

6.1 Hypotheses

The developed system responds to user self-reported emotion using an ECA implementing a multimodal feedback strategy in the domain of Information Technology. We hypothesized the following based on previous literature on empathy and professional judgement:

1. An empathic tutoring system will improve learning in comparison to a neutral tutoring system within a classroom setting in the domain of IT.
2. Learners with higher empathic tendencies interacting with an empathic tutoring system will achieve higher learning gains when compared to learners in the same condition with lower empathic tendencies.
3. Learners who interact with an empathic tutoring system will give the system higher subjective user judgements in comparison to learners interacting with the neutral tutoring system in the domain of IT.
4. Learners who are identified to have higher empathic tendencies interacting with an empathic tutoring system give the system higher subjective judgements in

comparison to learners in the same condition who have lower empathic tendencies.

6.2 Participants

The final evaluation was conducted with seven classes totalling 140 students (92 males, 48 females) aged between 12-16 years old in the same school under similar conditions to the study described in Chapter 5. Participants were taken from Years 8 - 11 studying IT and were asked to interact with the developed neutral and empathic tutoring systems in the domain of Information Technology (IT). Participant abilities were determined based on school administered Cognitive Abilities Tests (CAT) scores and teacher target grades, that split learners into 3 categories: low, middle and high ability. For example a high ability learner in Year 10 would have a CAT score of 124 and a teacher target grade of 'A'', whilst a low ability learner in Year 10 would have a CAT score of 85 and a teacher target grade of 'E', where the average CAT score is 100 (Forster and Metcalfe 2010) and an average teacher target grade is "C" at GCSE.

6.3 Methodology

The methodology of the study described in this chapter is similar to the methodology detailed in Section 5.4, but modified to include the Interpersonal Reactivity Index, a widely used psychological questionnaire, which we have used to establish participant empathic tendencies (see Appendix 4). Each participant interacted with one condition in a balanced between subjects study based on their empathic tendency score which was high, middle or low. These groups were derived from the interquartile ranges measured from the total group of participants.

During completion of the web-based quiz, after every five questions, each participant was asked to indicate their emotion. The ECA responded using an empathic feedback strategy based on the participant's self-reported emotion. Thereafter, learners completed a written post-test and an online questionnaire on their subjective judgements for the system (See Appendix 4).

6.4 Materials

We have developed a set of materials to investigate the hypotheses stated earlier, and the following sub sections give a detailed description.

6.4.1 Pre-test and post-test

The pre-test and post-test were developed to test learners on their knowledge of three topics for the current domain (see Table 6.1 below). The test was developed and administered in the exact same conditions as described in Section 5.5.1.

Table 6.1 Domain topics in IT

Topic
1. Hardware 2. Software 3. Health and Safety

6.4.2 Measuring learner empathic tendencies

Empathy was previously defined in Section 2.2 as a set of constructs having to do with "the responses of one individual's response to the experiences of another" (Davis 1994, p.12). The Interpersonal Reactivity Index (IRI) (Davis 1980) is widely accepted as a valid self-report multi-dimensional measure of empathy in various applications, as discussed in Section 2.2.3.The IRI is divided into four sub scales (of seven items each): Perspective- taking, Fantasy, Empathic Concern and Personal Distress, (see Appendix 4). Perspective-taking and Fantasy subscales are related to the cognitive component of empathy while Empathic Concern and Personal Distress subscales relate to the affective component. This multi-dimensional measure was used in the current study to measure participants' empathic tendencies. This enabled us to group participants using this measure and to investigate the correlation between empathic tendencies and learning outcomes. This information could form an important part of developing empathic tutoring systems for specific learners across varied applications in the future.

Expected results from previous studies (Hoffman 1977; Davis 1980) with adults suggest that females tend to score significantly higher than males. Scores on perspective-taking are expected to show the least difference between genders. The other three subscales could measure differences of up to 2.5 scale points. Finally, Perspective-taking scores should be positively associated with Empathic Concern and negatively with Personal Distress to validate the results achieved.

6.4.3 ETS version 2

Chapter 3 (Section 3.4) describes the enhancements that were made to the system based on the evaluation in Chapter 5 and the full implementation is included in Appendix 3. Similarly to the study detailed in Chapter 5 (Section 5.5.2), we compared two versions, a neutral and empathic ETS. The neutral condition of the study provides neutral feedback accompanied by a neutral expression without gesture to participants, whilst the empathic condition of the study provides empathic feedback with appropriate facial expression and gesture.

6.4.4 Questionnaire

An online questionnaire similar to the description in Section 5.5.3 was developed to obtain user judgements after their interaction with the empathic tutoring system. The survey included participants' basic information, judgements and comments on their interaction.

6.5 Results

6.5.1 Results of ETS version 2 on learning gain

The following section describes the hypothesis, results and analysis of the final study conducted with Information Technology pupils. This study evaluates the empathic tutoring system developed in the domain of Information Technology (IT).

Hypothesis 1: An empathic ETS will improve learning gain in comparison to a neutral ETS within a classroom setting in the domain IT.

Result: Unconfirmed

The results of this study indicate that participants in the neutral condition achieved a higher mean learning gain of 8.49 marks in comparison to the empathic condition where participants achieved a mean learning gain of 8.05 marks (see Table 6.2). Both scores are statistically significant suggesting that both conditions improve learner post-test scores when compared to pre-test scores as measured by the t-Test, however there are no significant differences between the two conditions when learning gain is compared. These results are similar to those reported in a previous study, discussed in Section 5.6, where learners averaged at least 7.5 marks improvement post-test scores when compared to pre-test scores. However, these results do not support our first hypothesis that learners in the empathic version of the system would achieve higher learning gains than learners in the neutral version of the system as learners achieved similar results.

Table 6.2 Paired samples t-Test - IT

Condition	Mean Learning Gain (Post-test – Pre-test)	Std. Deviation	N
Neutral	8.492*	6.692	65
Empathic	8.053*	6.615	75
*Statistically Significant			

Hypothesis 2 Learners with higher empathic tendencies interacting with an empathic tutoring system will achieve higher learning outcomes when compared to learners in the same condition with lower empathic tendencies.

Result: Unconfirmed

Our second hypothesis suggested that learners with high empathic tendencies (measured by Empathic Concern) would have the highest score in the empathic condition, whilst learners with low empathic tendencies would have the highest scores in the neutral condition. Table 6.3 shows that this is not the case. Learners in each group scored significant learning gain between their pre-test and post-test scores. However, low scorers on empathic tendency in the empathic condition scored the highest learning gain of 9.25 marks whilst high scorers on empathic tendency had the highest learning gain of 11.5 marks in the neutral condition, which appears to be the inverse of our original hypothesis. An independent t-Test measured a statistical difference between the learning gains for users with low empathic tendencies when compared to users with high empathic tendencies; $t(23) = -2.175$, $p = .040$). Furthermore, there is a significant correlation between higher empathic tendencies and higher learning gains in the neutral condition; $r(64) = .327$, $p = .008$. However, in the empathic condition, when learners with the highest empathic tendencies are ranked by empathic score, there is a significantly negative correlation of that to learning gains; $r(20) = -.487$, $p = .030$.

Table 6.3 Paired samples t-Test by empathic tendencies – (high, middle, low)

		Paired Differences		
Study Condition	Empathic Tendency	Mean Learning Gain**	Std. Deviation	N
Neutral	Low	5.467*	5.718	15
	Middle	8.590*	6.163	39
	High	11.500*	8.196	10
Empathic	Low	9.250*	7.841	12
	Middle	7.925*	6.810	40
	High	7.800*	5.996	20
*Statistically Significant ** Post-test - Pre-test				

Hypothesis 3:- Learners who interact with an empathic tutoring system will score higher subjective user judgements in comparison to learners interacting with the neutral tutoring system in both domains.

Result: Unconfirmed

Our third hypothesis looks at learners' subjective user judgements and our results do not suggest that there are higher judgements with participants in the empathic condition in comparison to participants in the neutral condition.

Hypothesis 4:-Learners who are identified to have higher empathic tendencies interacting with an empathic tutoring system will achieve higher subjective user judgements in comparison to learners in the same condition who have lower empathic tendencies.

Result: Unconfirmed

Table 6.4 Subjective user judgements from survey by empathic tendency in the neutral condition

	Mean Survey Scores		ANOVA Statistics			
Survey Question	**Low ET****	**High ET****	**df**	**F**	**p**	**ηp^2**
*Q1 Good Way to Learn	2.67	4.00	(2,61)	3.71	.030	.108
*Q2 Revise Other Subjects	2.53	4.00	(2,61)	3.93	.025	.114
*Q3 Good Outside Lessons	2.13	3.40	(2,61)	3.26	.045	.097

*Statistically significant at p<0.05

**Empathic Tendency measured by Empathic Concern Score from Interpersonal Reactivity Index

The fourth hypothesis looks at high empathic tendencies correlating to high learners' subjective judgements in the empathic condition. This hypothesis has not been proven by the results obtained from the current study. However, Table 6.4 above suggests that learners with high empathic tendencies in the neutral condition gave significantly higher mean judgements for three questions from the survey in the neutral condition when compared to learners with low empathic tendencies, as measured by the ANOVA statistical test and adjusted using the Bonferroni correction. This may suggest a preference for the neutral system by these particular users, however the effect size is small.

6.5.2 Results of student ability on learning gain

When learners are split by ability, Table 6.5 shows the paired samples t-Test results which show that most groups scored significantly higher scores in their post-test when compared to their pre-test score, with the exception of low ability learners in the empathic condition. Further to this, an independent samples t-Test indicated that high ability learners in the empathic condition scored significantly higher mean learning gains (M = 9.5, SD =7.19) when compared to low ability learners in the same condition (M = 1.9, SD = 3.49), $t(17) = 4.34$, $p < .001$, d =1.34) based on the independent samples t-Test. Levene's test measured unequal variances (F = 1.828, p = .183) therefore degrees of freedom were adjusted from 43 to 17. In addition, ability shows a significantly positive correlation to learning gain in the empathic condition, $r(76)$ = .299, p < .001. This suggests differences in learning gain between abilities in the empathic condition with a medium effect size.

The independent samples t-Test indicates that a higher mean learning gain is measured for low ability learners in the neutral condition (M = 6.9, SD = 5.87) in comparison with low ability learners in the empathic condition (M = 1.9, SD = 3.49), $t(13) = 2.14$, $p = .05$, $d = 1.04$, although the difference is not statistically significant. Levene's test measured unequal variances (F = 1.52, p =.238) therefore degrees of freedom were adjusted from 14 to 13. This does suggest that low ability learners in the neutral condition improved their learning when compared to low ability learners in the empathic condition who did not benefit from the interaction.

Table 6.5 Paired sample t-Test by ability

		Paired Differences		
Condition	Ability	Mean Learning Gain**	Std. Deviation	N
Neutral	Low Ability	6.889*	5.862	9
	Middle Ability	7.583*	6.358	24
	High Ability	9.625*	7.134	32
Empathic	Low Ability	1.857	3.485	7
	Middle Ability	7.710*	5.640	31
	High Ability	9.514*	7.186	37
*Statistically Significant ** Post-test - Pre-test				

6.5.3 Results of subjective user judgements

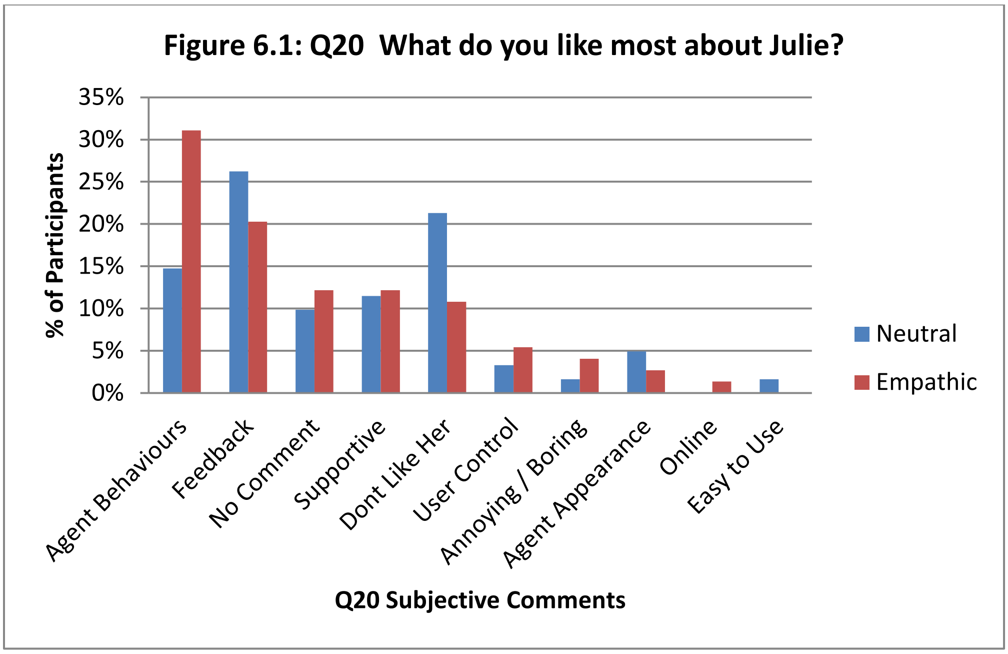
6.5.3.1 Subjective user judgements – survey responses

Participants were asked 21 questions that required multiple choice responses; this included 16 questions requiring a response on a five point Likert Scale and 4 questions requiring a "yes" or "no" response. Empathic tendency appears to have a significant effect on mean survey scores in the neutral condition of the study. The highest mean scores are consistently given by learners who have high empathic concern scores from the Interpersonal Reactivity Index. Learners with low empathic tendencies give low mean scores when compared to learners with high empathic tendencies who give higher mean scores for the survey questions included in Table 6.4.

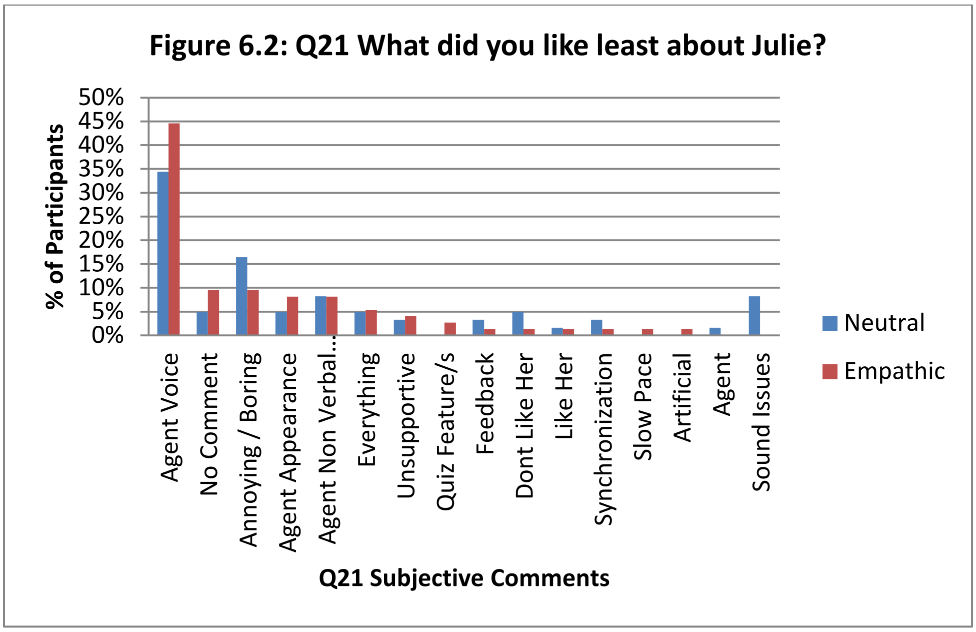
6.5.3.2 Subjective user judgements – open-ended questions

Participants were asked 6 open-ended questions on their interaction with Julie. A number of categories were identified from participants' comments using a thematic approach based on the frequency and importance of issues identified (see Figure 6.1 –

Fig 6.6). Descriptive data was used to summarise pupil comments and compare these to learner scores.



Pupils liked Julie's behaviours (15% - neutral, 31% - empathic) whilst pupils from both conditions (26% - neutral, 20% - empathic) liked the feedback they received; see Figure 6.1. Conversely, learners found Julie's voice the least attractive at 34% and 45% of participants in the neutral and empathic condition respectively. Pupils found her voice "creepy" or "weird". In addition pupils found her annoying whilst others mentioned her appearance and behaviours as the features they liked least of all; See Figure 6.2.



Learner subjective comments have beens split by high, middle or low empathic tendency as measured by Empathic Concern in Figure 6.3. The first two graphs in Figure 6.3 (A,B) compare subjective user judgements for low and high empathic tendency, showing that 33.3% of learners with low empathic tendencies did not like Julie and could not suggest any positive feature whilst 20% of learners with high empathic tendencies identified interactivity and working online as positive features. These subjective comments may shed light on average learning gains that were higher for learners with high empathic tendencies (11.5 marks) in the neutral condition when compared to learners with low empathic tendencies (5.46 marks).

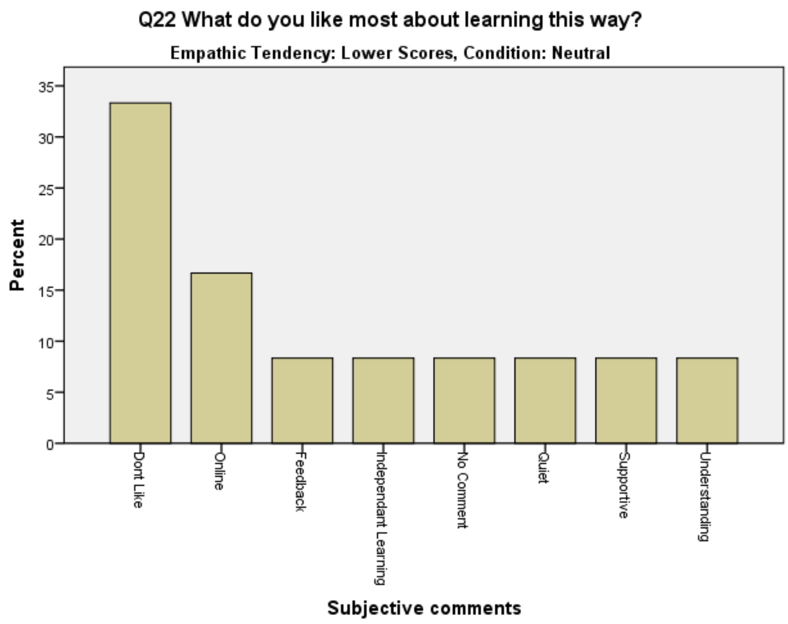
Conversely, in Figure 6.4 graph A and B shows that learners in the empathic condition with low empathic tendencies identified feedback (25%) and engagement (25%) as positive features whilst learners with high empathic tendencies identifed interactivity (30%).

In addition, Figure 6.5 shows that sound was another issue identified by learners in the neutral condition; however average learning gains for this group were surprisingly high at 11.5 marks when compared to other groups as has been previously mentioned.

Most pupils did not identify any problems whilst interacting with the system. However, Figure 6.6 looks at learners' subjective comments and sound is identified by learners in the neutral condition as problematic. Other problems identified were difficult questions, issues with submission and issues with Julie not working properly. When pupils were asked for any further comments, most did not give any comments whilst those that did stated that they did not like the interaction.

Figure 6.3 Subjective user judgements by empathic tendency: neutral condition

A)



B)

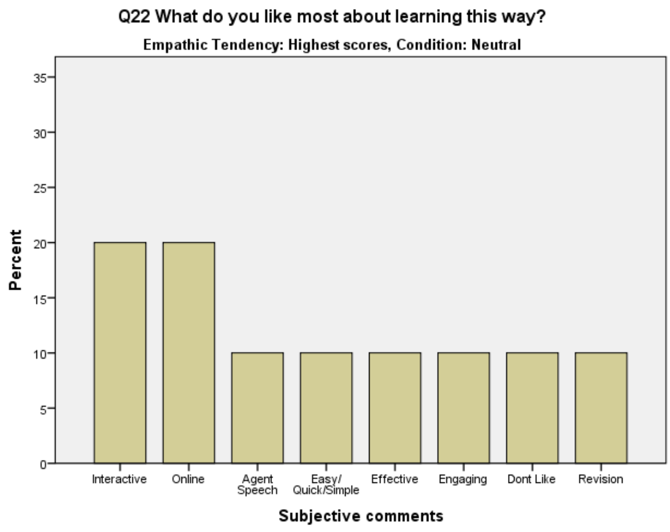
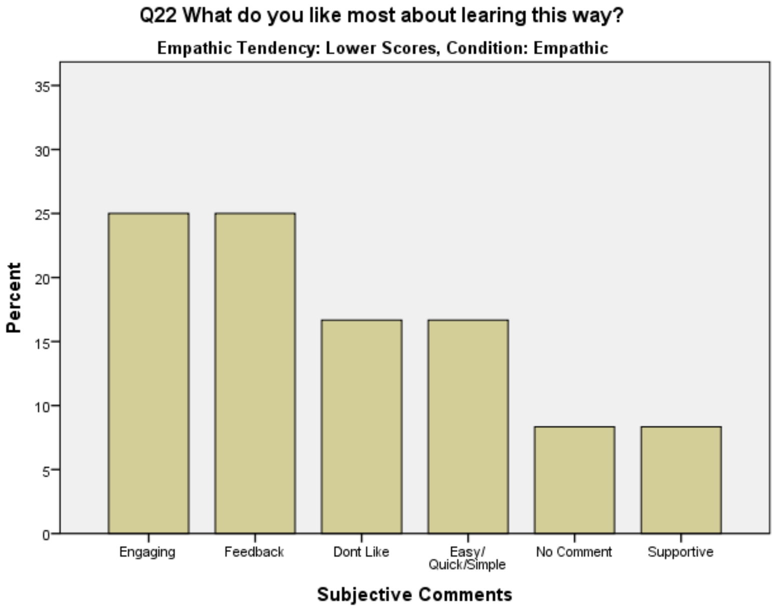


Figure 6.4 Subjective user judgements by empathic tendency: empathic condition

A)



B)

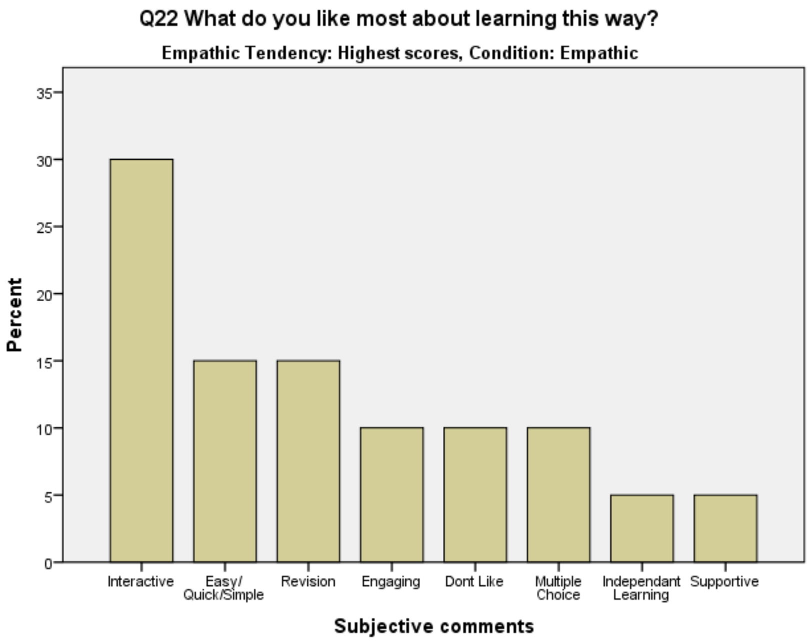
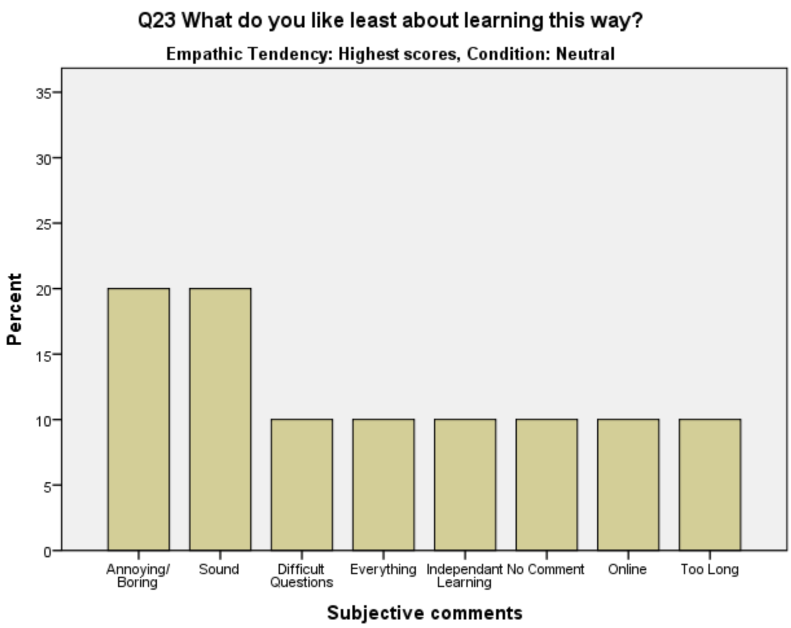


Figure 6.5 Subjective user judgements by high empathic tendency: experiment condition

A)



B)

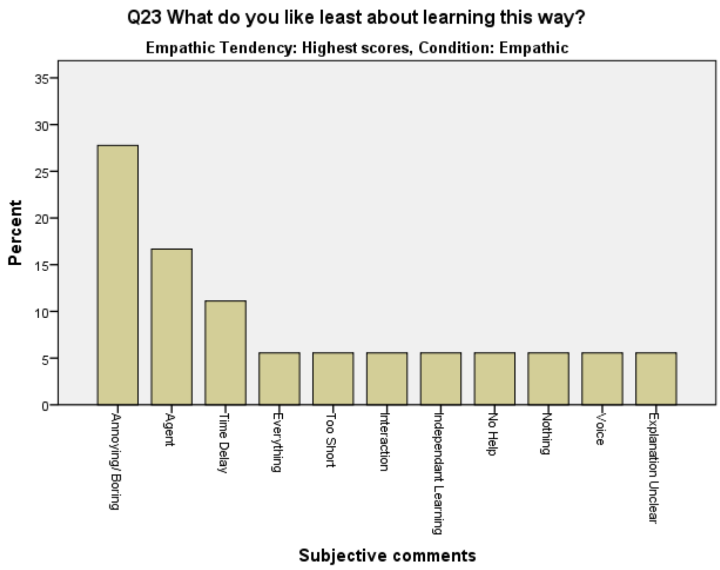
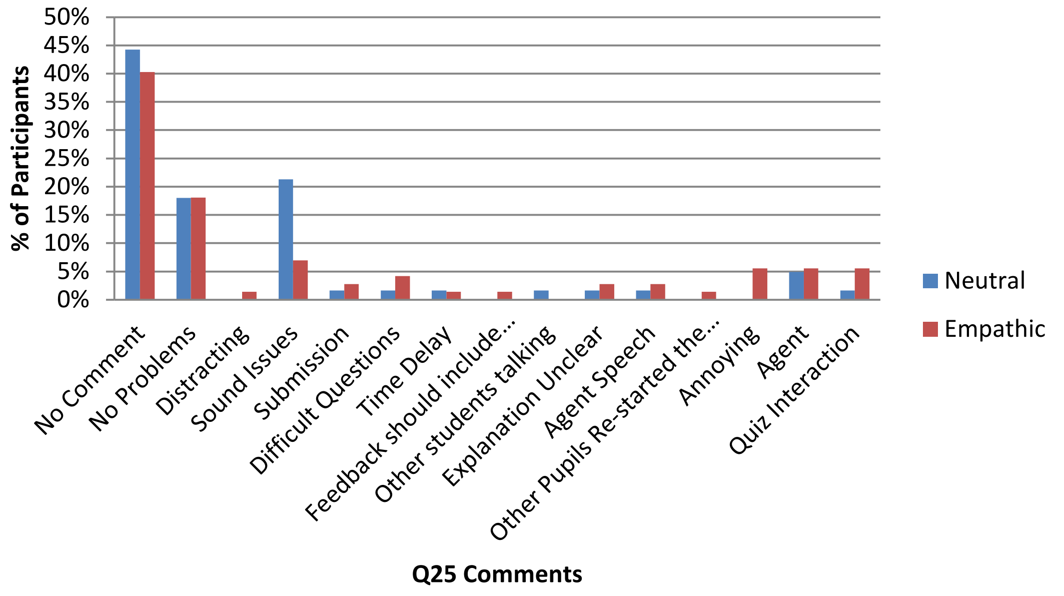
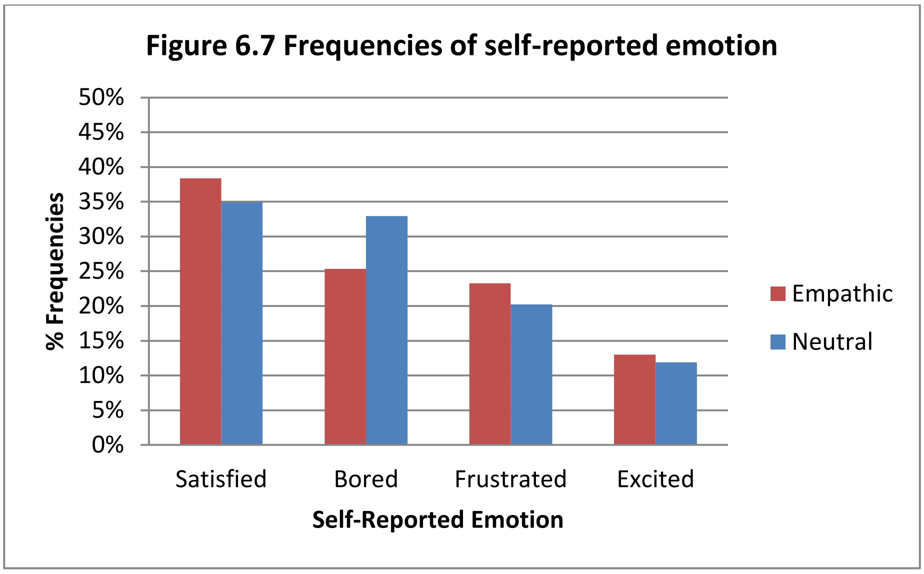


Figure 6.6 Q25 Problems with the system



6.5.4 Results of self-reported user emotion

Automated data on self-reported emotion was collected and Figure 6.7 shows the overall frequency of emotions reported by learners in this study. Satisfaction is reported most often in both conditions (35% - neutral, 38% - empathic) whilst excitement is reported least often (12% - neutral, 13% - empathic). Boredom was reported more often in the neutral condition with 33% as compared to the empathic condition at 25%.

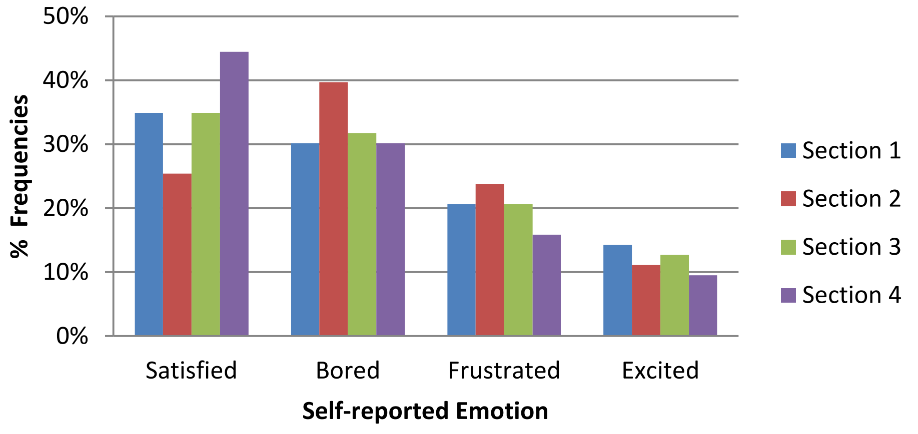


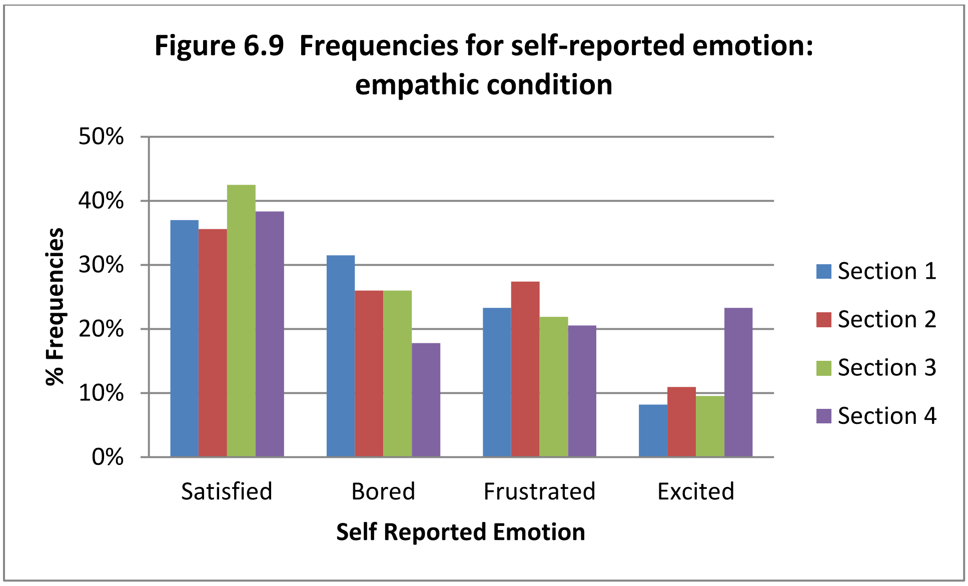
The ANOVA statistical test showed no differences in learning gains between the conditions of the experiment. However post hoc tests using Turkey's HSD revealed that learners who reported satisfaction in each section of the quiz achieved significantly higher learning gains than learners who reported boredom in both conditions (Section1 p = .033, Section2 p < .001, Section3 p = .006, Section4 p = .004). Additionally, learners who reported satisfaction achieved significantly higher learning gains than those reporting frustration in Section 2 for both conditions (p = .012). However, the overall effect sizes are small as shown in Table 6.6, which also shows the significant between-subjects effects for self-reported emotion on learning gain for each section of the quiz interaction. Figures: 6.8 and 6.9 show the frequencies of self-reported emotions for each condition.

Table 6.6 Between-subjects effects on learning gain: emotion

Section Number	**df**	**F**	***p***	η p2
Section 1	3 128	3.14	.028	.068
Section 2	3 128	7.49	.000	.149
Section 3	3 128	3.73	.013	.080
Section 4	3 128	4.81	.003	.101

Figure 6.8 Frequencies for self-reported emotion: neutral condition





6.6 Discussion

Our first hypothesis looked at the empathic ETS v2 improving learning in the domain of IT, however we found no evidence of this. Learners in the neutral condition scored a higher average learning gain of 8.49 marks in comparison to the empathic condition where learners scored an average of 8.05 marks (but the difference was not statistically significant).

Further to this, our second hypothesis looked at learners with high empathic tendencies scoring higher learning gains than those with low empathic tendencies in the empathic condition. Our results were not statistically significant when learning gains in the empathic condition were compared to learning gains in the neutral condition based on empathic tendency. However, learners measuring high empathic tendencies in the neutral condition scored higher learning gains when compared to learners with low empathic tendencies in the same condition. In addition, empathic tendency correlated positively with learning gain in the neutral condition. A further result showed that when learners in the empathic condition with the highest empathic tendencies are ranked by empathic tendency score, there is a negative correlation of that with learning gain.

Learners gave higher subjective user judgements in the neutral condition when compared to the empathic condition, which did not support our third hypothesis. However, learning gain appeared to correlate with higher subjective user judgements in both conditions, suggesting that a positive interaction with the ETS encourages learning. Learners were not given feedback on their pre-test or post-test scores prior to completing the online questionnaire, however after each section of the ETS, learners received a score which may have an impact on the previously mentioned results.

In our final hypothesis, we looked at whether learners with high empathic tendency would give the highest subjective judgements in the empathic condition. The results suggest that learners with high empathic tendency gave the highest subjective judgements in the neutral condition rather than the empathic condition as expected.

These results confirm a previous smaller study, discussed in Chapter 5, where learners showed similar learning gains when a neutral ETS was compared to an empathic ETS. In addition, these results extend those conducted by Andallaza and Rodrigo (2013) who found little difference between two versions of the ECA Grimace, within Aplusix, an affective tutoring system in Algebra, differing in the affect sensitivity and frequency of intervention.

Empathic tendencies appear to correlate with learner judgements especially in the neutral version of the system suggesting that this version of the system appeals to learners with high empathic tendencies. This suggests that further studies could look at using empathic tendency to group learners to appropriately responsive tutoring systems.

When students were asked what they didn't like about learning with the ETS both in the neutral and empathic condition the single largest response (22% neutral, 16.9% empathic) was that they found the ETS annoying and boring. In addition these learners achieved an average learning gain (neutral 9.69 marks, empathic 9.58 marks) that was higher than the overall average learning gain of 8 marks. This may suggest that this negative emotion might not have had an effect on their learning.

Furthermore, although Julie's voice was identified as a feature that learners liked the least, the average learning gains for these learners is higher than the total average learning gain of 8 marks, (neutral 11.19 marks, empathic 9.94 marks). This may suggest that agent (ECA) voice and being annoyed or bored may have a lower effect on learning gain when compared to other features.

The learners who felt the system was useful for revision achieved an average learning gain of 15.4 marks in the neutral condition in comparison to 9.00 marks in the empathic condition, whilst learners who did not feel there was anything they liked about the system scored a lower average of 4.75 marks in the neutral condition and 1.50 marks in the empathic condition.

The results on learner emotion extend previous results as satisfaction which can be described as a similar emotion to flow (engagement) (Csikszentmihalyi 1990) was the most frequently reported emotion in both conditions. This was confirmed by comparisons of self-reported emotion to learning gain which showed higher learning gains for satisfaction although significant differences were only established between satisfaction and boredom across the four sections of the quiz. Further to this our study extends a recent survey by San Pedro et al. (2013) where boredom appears more frequently than frustration, when they developed and then applied automated detectors of student affect to previous tutorial logs from the ASSISTment system, an ITS in Mathematics.

Although our results are not statistically significant, they suggest that future research can investigate how best to detect and respond to boredom as opposed to concentrating on the detection and response to user confusion and frustration as previous research has done (Klein et al., 1999; Burleson 2006; McQuiggan et al., 2007) since it may occur more often than frustration. In addition, physiological sensors, follow-up interviews and observation of participants during the interaction using video or using think-aloud protocols to record participants' views can clarify student emotion to enable the ETS to respond appropriately.

6.7 Summary

In this chapter, we have described the final evaluation of the developed empathic tutoring system which was conducted in the domain of Information Technology. We discussed the results of the study which included participant learning gain and subjective user judgements.

The studies described in this chapter address the second, third and fourth research questions: Does an empathic feedback strategy positively affect learning gains and subjective user judgements in an ETS; what is the impact of learner empathic tendency on learning gain in an empathic ETS; how do the results of these evaluations compare with other studies conducted in different settings?

The hypotheses discussed investigated whether an empathic ETS would achieve higher learning gains and subjective user judgements when compared to a neutral ETS. We did establish that learning gain was significantly improved in both conditions however, there did not appear to be any significant differences between the two conditions, as they gave similar mean scores for learning gain, confirming the results from Chapter 5. There were no significant differences in subjective user judgements between the two conditions.

However, although not statistically significant overall, interactions between empathic tendency and learning gain suggest that there may be a link between these factors that needs to be investigated further. Learners with the highest empathic tendencies achieved the highest learning gains with the neutral ETS in addition to a positive correlation between empathic tendency and learning gain in the same condition. In the empathic condition, learners with the highest empathic tendencies showed a negative correlation with learning gain. In addition, learners with high empathic tendencies gave significantly higher subjective user judgements for the neutral ETS than those for the empathic ETS which disproved our original hypothesis. Further investigations indicated a significant difference between learners with different abilities: in the empathic condition high ability learners achieved higher learning gains when compared with low ability learners. However, there was no such difference between high and low ability learners in the neutral condition.

Qualitative analysis of subjective user judgements from open-ended questions in the survey for learners with high empathic tendencies and high learning gains in the empathic condition suggest they found the system “engaging” and identified “feedback” as features they liked most when learning with the system. However, these learners identified the ECA “voice” and described the system as “annoying/boring” when asked which features they liked least about learning with the system.

These results extend previous studies by including an evaluation on the impact of empathic tendency on learning gain and judgements with learners using an empathic

ETS. In addition these results provide useful information on the impact of self-reported emotion on learning gain with an empathic ETS in the classroom.

Future studies could use learner empathic tendency to appropriately respond to learner emotion by using empathy or a neutral strategy more robustly within a formal classroom environment.

Further to this, subsequent studies could look at varying the empathic ETS further in terms of detecting emotion using a combination of methods for example self-report with physiological sensors or context related information, to ensure that emotion is accurately detected.

Chapter 7 – Conclusion

The aim of this research was to explore the impact of an empathic ETS on learning gains in a formal educational context. The main objectives were (1) to develop an accurate ECA emotion expression strategy; (2) to develop an effective empathic feedback strategy combining affective and cognitive feedback that imitates one-to-one human tutoring within a classroom setting; (3) to evaluate the impact of user empathic tendency on learning gains and subjective user judgements during interactions with an ETS in Information Technology.

In this chapter, we discuss the contributions of the thesis; Section 7.1 revisits the research questions from Chapter 1 (Section 1.2) and describes the results obtained. In Section 7.2 we discuss the implications of developing an ETS and the impact on similar tutoring systems. Finally, in Section 7.3 we discuss possible future work to explore issues raised in this thesis on developing and implementing an ETS.

7.1 Research summary

In order to understand how emotion and learning are linked and whether this link can be used in computer-to-human tutoring contexts, we developed an empathic ETS that uses an ECA to provide affective and cognitive feedback to users during a quiz-based interaction within the domain of Information Technology. This differed from previous research which *investigated explanation-centred interactions in domains such as Physics and Mathematics*. The empathic ETS was developed to add to the limited research on the impact on learning of affective tutoring systems in classroom based situations *with teenagers*. To address the issue of evaluating the accuracy of ECA emotions *prior to implementation* within an ETS, we developed a multimodal algorithm to generate accurate ECA emotions. The accuracy of the algorithm was confirmed in an evaluation with users demonstrating the usefulness of this method when developing ECA output in Chapter 4. Following on from this we applied the multimodal algorithm when implementing an empathic feedback strategy using parallel and reactive empathy within an ETS. Previous studies had only evaluated the impact of parallel and reactive empathy within *virtual learning environments with multiple ECAs implemented as actors in a social context in comparison with our ECA implemented in an affective tutoring system within a learning context*. Two successive versions of the ETS were implemented and evaluated with users in classrooms to establish impact on learning gain in Chapters 5 and 6. The quantitative results from these studies did not establish a difference between the empathic ETS and a neutral ETS, however, user empathic tendency may impact on learning gains achieved with particular groups of learners. *Previous studies had not investigated the impact of learner empathic tendency on learning gains with an ETS*.

This thesis makes the following contributions based on the research activities we have conducted. We have shown that

- A multimodal strategy that uses speech, facial expression and gesture generates ECA expressions that users can accurately identify when compared to a strategy that uses speech alone.

- There is no evidence of differences in learning gains achieved by learners who interact with an empathic ETS when compared with a neutral ETS.
- Learners with high empathic tendencies achieved higher learning gains when compared with learners with low empathic tendencies in the neutral ETS.
- Empathic tendency correlates positively with learning gain in the neutral ETS.
- Empathic tendency correlates negatively with learning gain in the empathic ETS when learners with the highest empathic tendencies are ranked by empathic tendency score.
- High ability learners achieved higher learning gains when compared to low ability learners in the empathic condition. Also, ability correlates positively to learning gain in the same condition.
- Subjective user judgements suggest that positive learner judgements correlate with higher learning gains in both conditions.
- Learners self-reporting satisfaction, achieve higher average learning gains when compared to learners self-reporting boredom in both conditions. Further to this, boredom is the most frequently self-reported negative emotion in this type of human-computer interaction setting.

In Section 1.2 we stated the three research aims from which we identified the research questions that we investigated. The following subsections give a summary of the results that addressed each question and discuss our contributions in more detail.

7.1.1 Expressing ECA emotion

Q1: Which combination of channels used to portray ECA emotions are users able to identify most accurately?

The first research question set out to investigate the impact of carefully modelling emotion in an ECA to establish whether this improves emotion recognition by users. In Chapter 4, we identified empathy related emotions relevant to an educational setting that an ECA would express: excitement, satisfaction, concern and surprise. Subsequently we developed a multimodal algorithm to generate these emotions by combining one or more communication channels from the following: speech, facial expression and gesture. Then, we conducted a study where users were asked to identify and rate the confidence and intensity of each emotion that Julie, the ECA used throughout the thesis activities, expressed. In this study, users identified emotions expressed in four different conditions: speech; speech and facial expression; speech and gesture; speech, facial expression and gesture.

In the study, users identified emotions expressed through speech, facial expression and gesture significantly more accurately than emotions expressed through speech only. These results suggest that a multimodal strategy that uses speech, facial expression and gesture can generate accurate emotions for ECA use by using three communication channels to express an emotion more accurately. Therefore, this strategy could be used in human-computer interaction specifically an educational setting such as tutoring.

7.1.2 Developing an empathic ETS

Q2: Does an empathic ETS feedback strategy positively affect students' learning gains and student subjective user judgements?

For our second research question, we looked at whether an empathic ETS feedback strategy can positively affect learning gains and subjective user judgements. In Chapter 3, we developed an empathic ETS that implemented a theoretically based empathic feedback strategy based on parallel and reactive empathy with users. This affective feedback strategy was implemented using the multimodal algorithm discussed in Chapter 4.

We conducted user evaluations, by comparing a version of the empathic ETS (version 1 and version 2) with a neutral ETS in Chapters 5 and 6 in terms of users' learning gain and subjective user judgements in the domain of Information Technology.

The results in Chapters 5 and 6 indicate that users interacting with the empathic ETS when compared to a neutral ETS have similar learning gains in the domain of Information Technology. Overall, subjective user judgements indicated that more learners liked ECA behaviours when interacting with the empathic ETS in comparison to the neutral system whereas both groups of learners did not like the agent voice. These results are similar to previous research on dialogue systems that have identified the agent voice as less preferred but having little impact on learning gain (Forbes-Riley et al., 2006).

Since there is no evidence that empathy can improve the overall effectiveness of an ETS, this may suggest that although people treat media as people, as suggested by the media equation (Reeves and Nass 1996) and project social conventions based on gender on to ECAs implemented as learning companions (Kim et al., 2007), this may not extend to expecting media to imitate human-to-human empathy. However, previous research has found some benefits in adding emotion to an ECA. For example, Lane et al. (2013) implemented an ECA in the role of coach in an informal learning environment with results showing improved self-efficacy. In addition, subjective user judgements from this research may suggest that further work could investigate the impact of individual agent features and learner empathic tendencies on learning as discussed in Section 7.3.

7.1.3 User empathic tendency

Q3: What is the impact of student empathic tendencies on learning gain and students' subjective judgements whilst interacting with an empathic ETS?

The third research question looked at the impact of learner empathic tendencies on learning gains and subjective user judgements in an empathic ETS. In Chapter 6, user empathic tendencies were measured prior to the interaction with the empathic ETS v2 using the Interpersonal Reactivity Index and users were grouped into high, low and middle, based on their score for Empathic Concern. The results of the study for Information Technology suggest that users with high empathic tendencies had higher learning gains with the neutral ETS v2 when compared with users with low empathic tendencies in the same condition. Additionally empathic tendency correlated positively with learning gain in the neutral ETS v2. Conversely, empathic tendency had a negative correlation with learning gain in learners with the highest empathic tendencies in the empathic ETS v2. Subjective user judgements suggested that learners who achieved higher learning gains gave more positive judgements when compared to learners who achieved lower learning gains within the same condition.

Although the quantitative results achieved are surprising as we expected users with high empathic tendencies to achieve the highest learning gains with the empathic ETS (v2) whilst users with low empathic tendencies to achieve the highest learning gains in the neutral ETS, the qualitative results may suggest a link between positive judgements for the ETS and improved learning gains. This indicates that empathic tendency may impact on learning gain however further research could establish these results as discussed in Section 7.3. Another possibility may be that learners with high empathic tendencies are not as responsive to artificially generated empathy, when compared to human empathy because they may not relate to the empathic ETS behaviours, and therefore learn better when interacting with a neutral ETS.

7.1.4 Evaluations with an ETS

Q4: How do results from evaluations with users in a classroom environment compare with results from studies conducted in other settings or different users with affective tutoring systems?

The evaluations with the developed ETS are comparable to a recent study (Andallaza and Rodrigo, 2013) where two versions of an affective tutoring system that implements an ECA to interact with teenagers were compared. There were no significant differences measured in learning gain between the two conditions, similar to the results achieved in this thesis. However, our evaluations suggest that there may be an interaction between user empathic tendency and the use of an empathic strategy within feedback. In addition, the emotions that have been identified during the interaction with the ETS are similar to previous studies. However, boredom was self-reported more frequently than frustration. Therefore, this result could be confirmed in a different domain or formal setting to establish generalization.

7.2 Implications of developing an empathic ETS

The implications of the results described above and the design decisions made during the thesis activities are discussed in this section. The ETS presented in this thesis cannot be directly compared with other systems as their implementation can vary considerably and this remains a significant consideration within this research area.

1. Can schools successfully implement an ETS?

The classroom environment can be a challenging area to successfully implement complex technologies for an ETS. The school's ethos and support for the required technologies is an important part of getting this right as school wide targets and objectives would need to encompass the benefits and drawbacks of implementing an ETS to augment classroom teaching. Teaching staff would require committed support from the school in terms of training and technical support to understand how these technologies can be used across a variety of subjects. Further to this, parental support is required through the granting of permission for the use of specific technologies such as web cams for use with children as there are ethical considerations that parents may be concerned about.

The studies described in this thesis demonstrate the ethical, technological and design issues that should be considered and the importance of pilot studies and a school champion to maintain momentum and evaluate the impact of an ETS on student learning gain to establish success. Thereafter financial changes such as software and/or equipment, physical changes to classroom layout or structural changes to lessons (e.g. duration or staffing) may be required to improve ETS success. However, in the near future, schools are required to adapt to the changes in assessment within important national examinations (GCSEs and 'AS' levels) which place a greater emphasis on performance in final examinations in comparison to coursework or controlled assessment during lessons. Therefore schools need to investigate a variety of teaching and learning strategies, such as an ETS, that can engage learners in regular revision to improve learning gain inside and outside the classroom.

2. Does emotion matter in computer-based learning?

Previous research has suggested a link between certain emotions and learning in computer-based tutorial environments (Kort et al., 2001). For example, in the domain of computer literacy using natural language (Craig et al., 2004). The current study has investigated the link between emotion and learning within a quiz-based learning environment in Information Technology. Similar to the previous study, learners experienced significantly higher average learning gains when self-reporting satisfaction when compared to average learning gains when self-reporting boredom in both conditions. Consequently, this suggests that if positive emotion can be encouraged or maintained within an educational setting, by an ETS within computer-to-human tutoring, this can impact positively on learning gain. However, as previously mentioned, overall learning gains in both the empathic and neutral conditions were similar, although the impact of user empathic tendency on empathic interventions could be investigated further to confirm the results of this study. This suggests that further research needs to confirm these results in a variety of domains and learning settings.

3. Can an ETS be implemented to imitate human-to-human tutoring?

To some extent, the developed system extends previous work based on ITS' that give cognitive feedback on learning by using a similar strategy with an empathic ETS in a quiz-based learning environment with teenagers. In particular, the successful work done over a substantial period by Graesser et al. (1999; 2005) using AutoTutor, to imitate one-to-one human tutoring strategies of untrained tutors has been used as a theoretical and empirical foundation for the tutorial strategies used within our ETS. However, the differences in implementation and context by adding the affect dimension may impact on the overall effectiveness of the ETS we have developed.

The two successive versions of the ETS that were developed indicate the iterative nature of such implementations and the mixed results achieved in the studies indicate that further studies would be required to firmly establish the impact of affective tutoring systems in classrooms. This is supported by the continued work to add affect detection and expression with established ITS such as AutoTutor (D'Mello et al., 2009) and Aplusix (Andallaza and Rodrigo 2013).

4. Developing an ETS based on Davis' theory of empathy.

Davis' (1994) theory of empathy is the theoretical basis of the feedback strategy developed for the ETS. His theory encapsulates both cognitive and affective empathy and clearly defines parallel and reactive empathy which we have used to inform how the ETS responds to positive and negative emotion respectively. In addition, we use parallel and reactive empathy to provide affective feedback for correct and incorrect responses for each multiple choice quiz question. One limitation could be that a theoretically based empathy strategy may be inferior to one developed through training data from human tutors and students in the specific setting. Therefore, future studies could make the comparison to establish effectiveness as discussed in Section 7.3.

5. ETS with manually implemented ECA empathic behaviours.

Although the system designs were based on theoretical or empirical evidence, the implementation through manually inserting behaviours rather than using automated methods for the ETS limit scalability. For example, the qualitative results achieved in our studies in Chapters 5 and 6 suggest that the ETS voice can impact on user subjective judgements even when varied pre-recorded human speech is used.

Therefore, using sophisticated text-to-speech engines or speech and emotional variations in pre-recorded speech should be considered carefully to provide greater flexibility in an automated system whilst maintaining “human emotion” within this communication channel. This may be directly linked to emotion in the voice gaining a higher importance in an ETS when compared to systems that give cognitive feedback alone where a neutral voice is used.

6. Implications of ECA characteristics:

The developed ETS used a female, half-boded ECA as the tutor. The use of this ECA was directly related to the subset of tutors available for research purposes. However, the choice of a half bodied agent who appears to be seated at a desk was intuitively related to maximising the ECA size and therefore the effect of facial expression, gesture and other non-verbal communication on users. Further research may look at how ECA characteristics such as gender, size and/or body impact on learning within an ETS as discussed in Section 7.3.

7. IRI Index used to measure learner empathic tendencies:

In Chapter 6, the users’ empathic tendencies have been measured by the Interpersonal Reactivity Index (Davis 1980), which is a psychological tool that was developed to measure empathic tendency in adults. Consequently, a limitation to our measure of empathic tendency is that users in the thesis study were aged between 12 – 16 years of age. However, this tool is the most reliable measure of empathic tendency within the field as confirmed by Hoffman (2000). Furthermore, the mean user scores were appropriately validated by comparing them to Davis' (1980) expected mean scores. These expected scores contrasted between gender and the relation between the four subsections of the tool. As expected, girls gave higher mean scores for empathic tendency when compared to boys.

8. Developing a quiz based multiple choice ETS

The ETS described in this thesis is developed around a quiz-based system that was used as a teaching and learning tool. The development of the ETS is not comparable to any existing system as most systems use explanation centred learning with short or extended answers required. The advantage of this system is its adaptability to different domains. However, explanation centred learning encourages deeper learning such as analysis and synthesis which improves delayed transfer. Therefore future studies could investigate the impact of the developed quiz ETS on delayed learning gains and transfer using a delayed post-test to establish effects.

9. Implications of deploying technology across the classroom and conducting whole class evaluations.

Deploying ETS technology across an institution with an established IT infrastructure can be a challenge. This can be especially difficult to adapt technologies in schools with an established network and operating system aimed at a teaching and learning environment. ETS technology demands specific underlying hardware and software specifications to run effectively.

Therefore, designing evaluations should take these factors into consideration. Further to this, during the evaluations, provision can be made for technical support and task guidance limiting interference with the learning interaction between the ETS and the user. In Chapter 5 (Section 5.6), we gained empirical evidence identifying key issues with technical support that were carefully looked at and addressed before the final study described in Chapter 6 took place with Information Technology users. This resulted in a reduction in learners who had problems with their interaction when giving subjective user judgements and gives a firmer validity to the quantitative results achieved.

Therefore, a second criticism of the study is the extent that classroom based interactions can be based on independent learner interactions with an ETS without supportive evidence such as interviews, video recordings or observations to validate results.

7.3 Future work

The modelling, implementation and evaluation of the ETS in the current thesis has in part addressed the four research questions set out in Chapter 1. However, there are aspects of the current research that require further investigation. In this section, we present suggestions for further work and extensions to the current ETS model.

1. Extending the ETS feedback strategy through accurate emotion expression.

The current ETS identified a suitable multimodal algorithm to develop and implement accurate emotion expression. However, the evaluation described in Chapter 4 compared different emotions within the four conditions of the study. Having established that multimodal feedback gives the most accurate ECA emotion, a further study could establish for each emotion, which combination of multimodal emotion expression is more accurately and confidently identified by users to maximise emotion recognition by users for more complex emotions such as motivated.

One further extension to generating accurate emotion is using an ECA that allows finer control of non-verbal behaviour such as gesture length. The ability to make these adjustments could be used to tailor ECA output towards the target user through evaluations used as training data to establish a baseline for each emotion that is then used in the subsequently developed tutoring system.

2. Extending the ETS to use physiological sensors or context data.

State-of-the-art systems use physiological sensors to augment self-report, which can negatively impact the interaction or alter subsequent emotion, when detecting user affect within computer-based learning environments. Although this research is in its infancy in relation to which emotion each sensor can accurately measure (Arroyo et al., 2009a; Woolf et al., 2009), in addition to the intrusive nature of some sensors, they could be used to provide further input for an ETS to respond more appropriately to learner emotion in future. Alternatively, context related data from current or previous interactions can also be used to enhance affect detection (Andallaza and Rodrigo 2013), although quiz-based learning environments may include limited data of this nature.

In summary, the current system can be extended to incorporate a variety of sensors such as webcams, wrist bands and pressure seats, or context related information such as time taken to complete a question, in order to establish the most effective method of affect detection to improve learning gains within a tutorial environment in a classroom context.

3. Extending the ETS to establish system generalization.

The current study was not able to evaluate the impact of the developed ETS across a variety of domains and learning contexts. Therefore evaluating the ETS in another subject area across a number of schools can establish whether the system does generalise across domains.

In addition, the developed ETS investigated one type of learning style; quiz-based learning. The ETS could be extended in order to compare the results in this study with other learning environments such as explanation centred learning to establish the effects on learning. One question to consider is whether some learning styles are inappropriate for formal learning environments such as classrooms when compared to less formal learning environments such as self-study at home.

References

Albiero, P., G. Matricardi, D. Speltri and D. Toso (2009). "The assessment of empathy in adolescence: A contribution to the Italian validation of the "Basic Empathy Scale"." Journal of Adolescence **32**(2): 393-408.

Aleven, V. A. W. M. M. and K. R. Koedinger (2002). "An effective metacognitive strategy: learning by doing and explaining with a computer-based Cognitive Tutor." Cognitive Science **26**(2): 147-179.

Andallaza, T. and M. M. Rodrigo (2013). Development of an Affect-Sensitive Agent for Aplusix. Artificial Intelligence in Education. H. C. Lane, K. Yacef, J. Mostow and P. Pavlik, Springer Berlin Heidelberg. **7926:** 575-578.

Arroyo, I., D. Cooper, W. Burleson, B. Woolf, K. Muldner and R. Christopherson (2009a). Emotion Sensors Go To School. AIED 2009. V. Dimitrova, R. Mizoguchi, B. du Boulay and A. C. Graesser. Brighton, UK, IOS Press. **200:** 17-24.

Arroyo, I., K. Muldner, W. Burleson, B. Woolf and D. Cooper (2009b). Designing Affective Support to Foster Learning, Motivation and Attribution. AIED 2009, Brighton, UK, IOS.

Arroyo, I., B. Woolf, J. M. Royer and M. Tai (2009c). Affective Gendered Learning Companions. AIED 2009. V. Dimitrova, R. Mizoguchi, B. du Boulay and A. C. Graesser. Brighton, UK, IOS Press. **200:** 41-48.

Atkinson, R. C. (1974). "Teaching Children to Read Using a Computer." American Psychologist **29**: 169-178.

Atkinson, R. K., R. E. Mayer and M. M. Merrill (2005). "Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice." Contemporary Educational Psychology **30**(1): 117-139.

Batson, C. D., C. L. Turk, L. L. Shaw and T. R. Klein (1995). "Information function of empathic emotion: Learning that we value the other's welfare." Journal of Personality and Social Psychology **68**(2): 300-313.

Baylor, A., J. Ryu and E. Shen (2003). The Effects of Pedagogical Agent Voice and Animation on Learning, Motivation and Perceived Persona. World Conference on Educational Multimedia, Hypermedia and Telecommunications 2003. Honolulu, Hawaii, USA, AACE. **1:** 452-458.

Beale, R. and C. Creed (2009). "Affective interaction: How emotional agents affect users." International Journal of Human-Computer Studies **67**(9): 755-776.

Beven, J. P., A. O'Brien-Malone and G. Hall (2004). "Using the Interpersonal Reactivity Index to assess empathy in Violent offenders." International journal of Forensic Psychology **1**: 33-41.

Bickmore, T. (2003). Relational Agents: Effecting change through Human-Computer Relationships Media Arts & Sciences, Massachusetts Institute of Technology. **Doctoral Dissertation:** 284.

Bickmore, T. and D. Schulman (2007). Practical approaches to comforting users with relational agents. CHI '07 extended abstracts on Human factors in computing systems. San Jose, CA, USA, ACM**:** 2291-2296.

Bickmore, T. W., R. Fernando, L. Ring and D. Schulman (2010). "Empathic Touch by Relational Agents." Affective Computing, IEEE Transactions on **1**(1): 60-71.

Bloom, B. S. (1956). Taxonomy of Educational Objectives: Handbook of Cognitive Domain. New York, McKay.

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**: 4-16.

Brown, P. and S. C. Levinson (1987). Politeness: Some universals in language use. New York, Cambridge University Press.

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. . Cambridge, MA, Massachusetts Institute of Technology. **Doctoral Dissertation**.

Cassell, J. (2000). "Embodied Conversational Interface Agents." Communications of the ACM **43**(4): 70-78.

Cassell, J., T. Bickmore, H. Vihjalmsson and H. Yan (2000). More Than Just a Pretty Face: Affordances of Embodiment. Intelligent User Interfaces. New Orleans**:** 52-59.

Cassell, J., C. Pelachaud, N. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost and M. Stone (1994). Animated Conversation: Rule-Based Generation of Facial Expression, Gesture and Spoken Intonation for Multiple Conversational Agents. . Siggraph 94, ACM SIGGRAPH, Addison Wesley**:** 413-420.

Castellano, G., A. Paiva, A. Kappas, R. Aylett, H. Hastie, W. Barendregt, F. Nabais and S. Bull (2013). Towards Empathic Virtual and Robotic Tutors. Artificial Intelligence in Education. H. C. Lane, K. Yacef, J. Mostow and P. Pavlik, Springer Berlin Heidelberg. **7926:** 733-736.

Cliffordson, C. (2001). "Parents' judgments and students' self-judgments of empathy: The structure of empathy and agreement of judgments based on the Interpersonal Reactivity Index (IRI)." European Journal of Psychological Assessment **17**(1): 36-47.

Conati, C. (2002). "Probabilistic assessment of user's emotions in educational games." Applied Artificial Intelligence **16**(7-8): 555-575.

Conati, C., A. Gertner and K. VanLehn (2002). "Using Bayesian Networks to Manage Uncertainty in Student Modelling." User Modelling and User-Adapted Interaction **12**(4): 371-417.

Craig, S., A. Graesser, J. Sullins and B. Gholson (2004). "Affect and learning: An exploratory look into the role of affect in learning with AutoTutor." Journal of Educational Media **29**(3): 241-250.

Csikszentmihalyi, M. (1990). Flow: The Psychology of Optimal Experience, Harper and Row.

D'Mello, S., B. Lehman, J. Sullins, R. Daigle, R. Combs, K. Vogt, L. Perkins and A. Graesser (2010). A time for emoting: When affect-sensitivity is and isn't effective at promoting deep learning. Intelligent Tutoring Systems, Springer**:** 245-254.

D'Mello, S. K., S. Craig, J. Sullins and A. C. Graesser (2006). "Predicting Affective States expressed through an Emote-Aloud Procedure from AutoTutor’s Mixed-Initiative Dialogue." International Journal of Artificial Intelligence in Education **16**: 3-28.

D'Mello, S. K., S. D. Craig, K. Fike and A. C. Graesser (2009). "Responding to Learners’ Cognitive-Affective States with Supportive and Shakeup Dialogues." Human Computer Interaction, HCI 2009: 595-604.

Davis, M. (1994). Empathy: A Social Psychological Approach. Madison, WI, Brown and Benchmark.

Davis, M. H. (1980). "A Multidimensional Approach to Individual Differences in Empathy " JSAS Catalog of Selected Documents in Psychology **10**: 85.

De Carolis, B. (2005). MyTutor: a personal tutoring agent. Intelligent Virtual Agents, Springer**:** 478-488.

Demetriou, A., A. Efklides and M. Platsidou (1993). "The architecture and dynamics of developing mind: Experimental structuralism as a frame for unifying cognitive developmental theories." Monographs of the society for Research in Child Development **58**(234).

Dennis, M., J. Masthoff, H. Pain and C. Mellish (2011). Does Self-Efficacy Matter When Generating Feedback? Artificial Intelligence in Education. G. Biswas, S. Bull, J. Kay and A. Mitrovic, Springer Berlin Heidelberg. **6738:** 444-446.

Doddannara, L. S., S. M. Gowda, R. S. Baker, S. M. Gowda and A. M. J. B. De Carvalho (2011). Exploring the relationships between design, students' affective states, and disengaged behaviors within an ITS. Artificial Intelligence in Education. G. Biswas, S. Bull, J. Kay and A. Mitrovic. **6738**.

Dweck, C. S. (1999). Self-theories: Their role in motivation, personality and development. Philadelphia, The Psychology Press.

Eisenberg, N. and R. Fabes (1990). "Empathy: Conceptualization, measurement, and relation to prosocial behaviour." Motivation and Emotion **14**(2): 131-149.

Eisenberg, N., R. A. Fabes, B. Murphy, M. Karbon, P. Maszk, M. Smith, C. O'Boyle and K. Suh (1994). "The Relations of Emotionality and Regulation to Dispositional and Situational Empathy Related Responding." Journal of Personality and Social Psychology **66**(4): 776-797.

Eisenberg, N. and R. Lennon (1983). "Sex differences in empathy and related capacities." Psychological Bulletin **94**(1): 100-131.

Ekman, P. and W. V. Friesen (1977). Facial Action Coding System, Consulting Psychologists Press, Stanford University, Palo Alto.

Forbes-Riley, K., D. Litman, S. Silliman and J. R. Tetreault (2006). "Comparing Synthesized versus Pre-Recorded Tutor Speech in an Intelligent Tutoring Spoken Dialogue System." American Association for Artificial Intelligence.: 509-514.

Forster, N. and I. Metcalfe (2010). Using MidYIS Individual Pupil Records (IPRs) to inform Teaching and Learning, Centre for Evaluation and Monitoring (CEM), Durham University.

Goleman, D. (1995). Emotional Intelligence. New York, Bantam Books.

Graesser, A., S. Lu, G. Jackson, H. Mitchell, M. Ventura, A. Olney and M. Louwerse (2004). "AutoTutor: A tutor with dialogue in natural language." Behavior Research Methods, Instruments, & Computers **36**(2): 180-192.

Graesser, A. C., G. T. Jackson, E. C. Mathews, H. H. Mitchell, A. Olney, M. Ventura, P. Chipman, D. Franceschetti, X. Hu and M. M. Louwerse (2003). Why/AutoTutor: A test of learning gains from a physics tutor with natural language dialog. Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society**:** 1-6.

Graesser, A. C., D. S. McNamara and K. Van Lehn (2005). "Scaffolding deep comprehension strategies through Point&Query, AutoTutor, and iStart." Educational psychologist **40**: 225-235.

Graesser, A. C., N. K. Person and J. P. Magliano (1995). "Collaborative dialogue patterns in naturalistic one-to-one tutoring." Applied cognitive psychology **9**(6): 495-522.

Graesser, A. C., K. Wiemer-Hastings, P. Wiemer-Hastings and R. Kreuz (1999). "AutoTutor: A simulation of a human tutor." Journal of cognitive systems research: 35-51.

Gratch, J. and S. Marsella (2005). "Lessons from emotion Psychology For The Design Of Lifelike Characters." Applied Artificial Intelligence **19**: 215-233.

Hershkovitz, A., R. S. J. D. Baker, G. R. Moore, L. M. Rossi and M. Velsen (2013). The Interplay between Affect and Engagement in Classrooms Using AIED Software. Artificial Intelligence in Education. H. C. Lane, K. Yacef, J. Mostow and P. Pavlik, Springer Berlin Heidelberg. **7926:** 587-590.

Heylen, D., A. Nijholt and R. op den Akker (2005). "Affect in Tutoring Dialogues." Applied Artificial Intelligence **19**: 287-311.

Heylen, D., M. Vissers, R. o. d. Akker and A. Nijholt (2004). Affective Feedback in a Tutoring System for Procedural Tasks. Affective Dialogue Systems 2004. E. Andre and e. al, Springer**:** 244-253.

Hill, R., J. Belanich, C. H. Lane, G. M. Core, M. Dixon, E. Forbell, J. Kim and J. Hart (2006). Pedagogically Structured Game-Based Training: Development Of The ELECT BiLAT Simulation. Army Science Conference. University of Southern California Marina Del Rey CA Institute for Creative Technologies. **25**.

Hoffman, M. L. (1977). "Sex differences in empathy and related behaviors." Psychological Bulletin **84**(4): 712-722.

Hoffman, M. L. (1987). The contribution of empathy to justice and moral judgment. Empathy and its development. N. Eisenberg and J. Strayer. New York, NY US, Cambridge University Press**:** 47-80.

Hoffman, M. L. (2000). Empathy and Moral Development: Implications for Caring and Justice. Cambridge, UK, Cambridge University Press.

Hussain, M. S., O. AlZoubi, R. A. Calvo and S. K. D'Mello (2011). Affect Detection from Multichannel Physiology during Learning Sessions with AutoTutor. Artificial Intelligence in Education. G. Biswas, S. Bull, J. Kay and A. Mitrovic, Springer Berlin Heidelberg. **6738:** 131-138.

Jolliffe, D. and D. P. Farrington (2006). "Development and validation of the Basic Empathy Scale." Journal of Adolescence **29**: 589-611.

Kapoor, A., W. Burleson and R. W. Picard (2007). "Automatic prediction of frustration." International journal human computer studies **65**: 724-736.

Ke, F. (2008). "Alternative goal structures for computer game-based learning. ." International Journal of Computer-Supported Collaborative Learning **3**: 429-445.

Keir, G. (2012). Oxford Dictionaries. Oxford Dictionaries Online. J. Pearsall, Oxford University Press.

Keleş, A., R. Ocak, A. Keleş and A. Gűlcű (2008). "ZOSMAT: Web-based Intelligent Tutoring System for Teaching–learning Process." Expert Systems with Applications (ESWA 2437): 1-11.

Keller, J. M. (1987). IMMS: Instructional Materials Motivation Survey, Florida State University.

Kendon, A. (2002). "Some uses of the head shake." Gesture **2**(2): 147-182.

Kim, Y., A. L. Baylor and E. Shen (2007). "Pedagogical agents as learning companions: the impact of agent emotion and gender." Journal of Computer Assisted Learning **23**: 220-234.

Klein, J., Y. Moon and R. W. Picard (1999). This computer responds to user frustration. CHI '99 extended abstracts on Human factors in computing systems. Pittsburgh, Pennsylvania, ACM**:** 242-243.

Klein, J., Y. Moon and R. W. Picard (1999). "This computer responds to user frustration: Theory, design, and results." Interacting with Computers **14**(2): 119-140.

Kohler, W. (1929). Gestalt psychology. Oxford England, Liveright.

Kort, B., R. Reilly and R. W. Picard (2001). An affective model of interplay between emotions and learning: reengineering educational pedagogy-building a learning companion. Advanced Learning Technologies, 2001. Proceedings. IEEE International Conference on**:** 43-46.

Lane, C. H. (2010). Characters that Help You Learn: Individualized Practice with Virtual Human Role Players. Adaptation and Personalization in E-B/Learning using Pedagogic Conversational Agents (APLEC 2010). D. Perez-Marin, I. Pascual-Nieto and S. Bull. Hawaii, USA**:** 7-12.

Lane, C. H. and K. Van Lehn (2005). "Teaching the Tacit Knowledge of Programming to Novices with Natural Language Tutoring." Computer Science Education **15**(3): 183-201.

Lane, H. C., C. Cahill, S. Foutz, D. Auerbach, D. Noren, C. Lussenhop and W. Swartout (2013). The Effects of a Pedagogical Agent for Informal Science Education on Learner Behaviors and Self-efficacy. Artificial Intelligence in Education. H. C. Lane, K. Yacef, J. Mostow and P. Pavlik, Springer Berlin Heidelberg. **7926:** 309-318.

Lane, H. C., D. Noren, D. Auerbach, M. Birch and W. Swartout (2011). Intelligent Tutoring Goes to the Museum in the Big City: A Pedagogical Agent for Informal Science Education. Artificial Intelligence in Education, Springer Berlin Heidelberg. **6738:** 155-162.

Lang, P. J., M. M. Bradley and B. N. Cuthbert (1999). International affective picture system (IAPS): Technical manual and affective ratings, Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

Lee, J. and S. Marsella (2006). Nonverbal Behaviour Generator for Embodied Conversational Agents. Intelligent Virtual Agents**:** 243-255.

Lepper, M. (1993). Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. . Computers as Cognitive Tools,. S. Lajoie and S. Derry, Hillsdale: Lawrence Erlbaum Associates.**:** 75–105.

Lepper, M. R. and R. W. Chabay (1998). Socializing the Intelligent Tutor: Bringing Empathy to Computer Tutors. Learning Issues for Intelligent Tutoring Systems. H. Mandl and A. Lesgold, Springer-Verlag**:** 242-257.

Matsuda, N., W. W. Cohen, K. R. Koedinger, G. Stylianides, V. Keiser and R. Raizada (2010). Tuning Cognitive Tutors into a Platform for Learning-by-Teaching with

SimStudent Technology. Adaptation and Personalization in E-B/Learning using Pedagogic Conversational Agents (APLEC 2010), Hawaii, USA.

McClave, E. Z. (2000). "Linguistic functions of head movements in the context of speech." Journal of Pragmatics **32**(7): 855-878.

McDaniel, M., H. Roediger and K. McDermott (2007). "Generalizing test-enhanced learning from the laboratory to the classroom." Psychonomic Bulletin & Review **14**(2): 200-206.

McQuiggan, S. W., S. Lee and J. C. Lester (2007). Early prediction of student frustration. Affective Computing and Intelligent Interaction. A. C. R. Paiva, R. Prada and R. W. Picard, Springer Berlin Heidelberg. **4738:** 698-709.

McQuiggan, S. W. and C. J. Lester (2006). "Learning Empathy: A Data-Driven Framework for Modelling Empathetic Companion Agents " AAMAS'06: 961-968.

McQuiggan, S. W., J. L. Robinson, R. Philips and C. J. Lester (2008). "Modelling Parallel and Reactive Empathy in Virtual Agents: An Inductive Approach " AAMAS'08 **167-174**.

Moller, S., J. Krebber and P. Smeele (2006). "Evaluating the speech output component of a smart-home system." Speech Communication **48**(1): 1-27.

Moreno, R. and R. E. Mayer (2005). "Role of guidance, reflection and interactivity in an agent-based multimedia game." Journal of Educational Psychology **97**(1): 117-128.

Mueller, C. M. and C. S. Dweck (1998). "Praise for intelligence can undermine children’s performance." Journal of Personality and Social Psychology **75**(1): 33-52.

Nii, P. (1986). "The blackboard model of problem solving." AI Mag. **7**(2): 38-53.

Paiva, A., J. Dias, D. Sobral, R. Aylett, S. Woods, L. Hall and C. Zoll (2005). "Learning by Feeling: Evoking Empathy with Synthetic Characters
" Applied Artificial Intelligence **19**: 235-266.

Payr, S. (2003). "The virtual university's faculty: an overview of educational agents." Applied Artificial Intelligence **17**: 1-19.

Pekrun, R., T. Goetz, W. Titz and R. P. Perry (2002). "Academic Emotions in Students' Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research." Educational Psychologist **37**(2): 91-205.

Pelachaud, C. (2005). Multimodal expressive ECAs. in Proceedings of the 13th annual ACM international conference on Multimedia MM'05. Singapore, ACM: 684-689.

Picard, R. W., S. Papert, W. Bender, B. Blumberg, C. Breazeal, D. Cavallo, T. Machover, M. Resnick, D. Roy and C. Strohecker (2004). "Affective Learning - a manifesto." BT Technology Journal **22**(4): 253-269.

Poggi, I. and C. Pelachaud (2000). Emotional Meaning and Expression in Animated Faces. Affective Interactions, LNAI. A. Paiva, Springer-Verlag. **1814:** 14.

Prendinger, H., H. Dohi, H. Wang, S. Mayer and M. Ishizuka (2004). Empathic Embodied Interfaces: Addressing Users’ Affective State. Affective Dialogue Systems**:** 53-64.

Prendinger, H., J. Mori and M. Ishizuka (2005). "Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game." International journal human computer studies **62**: 231-245.

Putnam, R. T. (1985). Live Tutoring of Addition: protocols. Unpublished data.

Rau, M. A. (2013). Conceptual learning with multiple graphical representations: Intelligent Tutoring Systems support for sense-making and fluency-building processes. Human-Computer Interaction Institute. Pittsburgh, Carnegie Mellon University. **Doctoral Dissertation**.

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

Rickel, J., S. Marsella, J. Gratch, R. Hill, D. Traum and W. Swartout (2002). "Toward a new generation of virtual humans for interactive experiences." IEEE Intelligent Systems: 33-38.

Roscoe, R. D. and M. T. H. Chi (2007). "Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions." Review of Educational Research **77**(4): 534-574.

San Pedro, M. O. Z., R. J. D. Baker, S. M. Gowda and N. T. Heffernan (2013). Towards an Understanding of Affect and Knowledge from Student Interaction with an Intelligent Tutoring System. Artificial Intelligence in Education. H. C. Lane, K. Yacef, J. Mostow and P. Pavlik, Springer Berlin Heidelberg. **7926:** 41-50.

Sarrafzadeh, A., S. Alexander, F. Dadgostar, C. Fan and A. Bigdeli (2007). "How do you know that I don't understand?'' A look at the future of Intelligent Tutoring Systems." Computers in Human Behaviour: 1022.

Schunk, D. H. (1991). "Self-efficacy and academic motivation." Educational psychologist **26**: 207-231.

Stankov, S., M. Rosić, B. Žitko and A. Grubiŝić (2007). "TEx-Sys model for building Intelligent Tutoring Systems." Computers & Education: 1-20.

Stotland, E. (1969). Exploratory Investigations of Empathy. Advances in Experimental Social Psychology. L. Berkowitz. New York, Academic Press. **4:** 271-314.

Tan, J. and G. Biswas (2006). The Role of Feedback in Preparation for Future Learning: A Case Study in Learning by Teaching Environments. ITS 2006. M. Ikeda, K. Ashley and T.-W. Chan, Springer-Verlag. **LNCS 4053:** 370-381.

Tolksdorf, J., C. Becker-Asano and S. Kopp (2008). Do You Know How I Feel? Evaluating Emotional Display of Primary and Secondary Emotions. Intelligent Virtual Agents 2008. H. Prendinger, J. Lester and M. Ishizuka, Springer Berlin / Heidelberg. **5208:** 548-549.

van der Sluis, I. and C. Mellish (2009). Towards Empirical Evaluation of Affective Tactical NLG. In Empirical methods in natural language generation, European Workshop on Natural Language Generation (ENLG 2009). . E. J. Krahmer and M. Theune. Athens, Greece, Association for Computational Linguistics**:** 242-263.

Verplanken, B., G. Hofstee and H. Janssen (1998). "Accessibility of Affective Versus Cognitive Components of Attitudes." Eur. Journal of Social Psychology **28**: 23-35.

Wang, N., L. W. Johnson, R. E. Mayer, P. Rizzo, E. Shaw and C. Heather (2008). "The politeness effect: Pedagogical agents and learning outcomes." International journal human computer studies **66**: 98-112.

Weiner, B. (1985). "An Attributional Theory of Achievement Motivation and Emotion." Psychological Review **92**: 548-573.

Wiemer-Hastings, K., A. C. Graesser and D. Harter (1998). The Foundations and Architecture of Autotutor. ITS '98. G. e. a. E. B. P., Springer-Verlag. **1452**.

Woolf, B., T. Dragon, I. Arroyo, D. Cooper, W. Burleson and K. Muldner (2009). Recognizing and Responding to Student Affect. Human-Computer Interaction. Ambient, Ubiquitous and Intelligent Interaction**:** 713-722.

Woolf, B. P., E. Aimeur, R. Nkambou, S. Lajoie, S. D'Mello, R. Taylor, K. Davidson and A. Graesser (2008). Self Versus Teacher Judgments of Learner Emotions During a Tutoring Session with AutoTutor. Intelligent Tutoring Systems 2008, LNCS, Springer Berlin Heidelberg. **5091:** 9-18.

Zinn, C., J. Moore and M. Core (2005). Intelligent Information Presentation for Tutoring Systems. Multimodal Intelligent Information Presentation, Springer Netherlands. **27:** 227-252.

Appendix 1: Extract from Technical Report

Pilot study results:

Experiment design

The experiment investigates the impact of an Intelligent Tutoring System's strategy (affective & cognitive vs. cognitive only) on learners' progress as determined by three measures: the learners' motivation during a task, their subjective judgement of the ITS, and their learning through post-tests.

2.1 Participants

The study population is taken from secondary school pupils at a comprehensive school in the south east of England. Pupils in Year 9 aged 13-14 years studying an IT course equivalent to ½ GCSE A*-G are used in the study. One class of approximately 25 pupils at a local school in Year 9 has been requested to participate in the study. Matched random sampling is used to match at least 4 pupils by ability (low, high and middle) and gender to each condition.

Experimental condition

The study will have 2 conditions as shown in table A1.1 below with an even distribution of participants by ability and gender.

Table A1.1 Experimental condition

	High Ability	Middle Ability	Low Ability
Affective	4	4	4
Non- Affective	4	4	4

Equipment and procedure

The study was based in a local comprehensive school in the United Kingdom. One class of year 9 pupils aged 13-14 years were asked to participate in the study as part of their lessons in IT. A group of 5 students in year 10 aged 14-15 years were asked to participate in evaluating an example system as part of their lesson in Business studies. The materials used include the following:

- Computer Room
- Install tutoring system software in all computers.
- Install software for automated data capture in all computers.
- Questionnaires
- Pre-test
- Post-test

Procedure

Participants are told that the purpose of the pilot study is to provide quizzes that use animated characters to help their revision of theoretical concepts in Information Technology. Participants are given an information sheet which describes how to log in and how to interact with Julie the embodied agent.

The participants are asked to complete a short test within class time. They interact with the system for five fifteen minute sessions in lesson time during one month. The class is asked to make subjective comments on the tutoring system and complete a post test to measure their progress.

Results and analysis

To assess the suitability of the developed tutoring system we carried out a study during March and April of 2009. A class of 24 students aged 13-14 years in Year 9 were asked to participate in the study. The aim of the study was to evaluate the DeanesITQuiz System which implemented an affective and cognitive strategy to teach Information Technology (IT) in comparison to a cognitive strategy. Students were randomly allocated to either condition with both genders and abilities represented.

Students were briefed about the study and given a demonstration on how to use the DeanesITQuiz System. The study began with a pre-test to measure subject knowledge followed by three separate interactions with the system during consecutive half hour periods of each timetabled lesson, see Table A.1.2. The study concluded with a post-test to measure learning and a short questionnaire to obtain user subjective comments. The results of the study are discussed in the following sections.

Table A1.2 Experiment timetable

Date	Activity	Lesson
Pre-test	13/03/2009	09:45 – 10:30
Section 1	20/03/2009	12:30 -13:00
Section 2	20/03/2009	14:15-14:45
Section 3	27/03/2009	10:00-10:30
Post-test	03/04/2009	12:00-12:45
Questionnaire	03/04/2009	12:45-13:00

Results on Learning

The preliminary results for the study suggest an overlap between the two conditions. We conducted a two-tailed t-Test giving a value p of 0.36 which does not show a significant statistical difference.

The difference in learning was calculated by subtracting pre-test scores from post-test scores and our study suggests that the mean for the affective & cognitive condition 1.25 is at least twice the size of the cognitive condition which is 0.5. Table A1.3 gives summary measures for the study.

The standard deviation in the affective & cognitive condition was 6.08 marks in comparison to 3.78 marks in the affective condition. There were an equal number of participants with the lowest difference in learning of -14 reported in the affective and cognitive condition while both conditions had at least one participant with the maximum learning difference of eight.

Table A1.3 Results summary

Measure	Affective & Cognitive	Cognitive
Mean	1.25	0.5
Standard Deviation	6.08	3.78
Standard Error Mean	1.75	1.09
Number of Participants	12	12
Minimum Learning Diff	-14	-5
Maximum Learning Diff	8	8

Results from subjective user judgements:

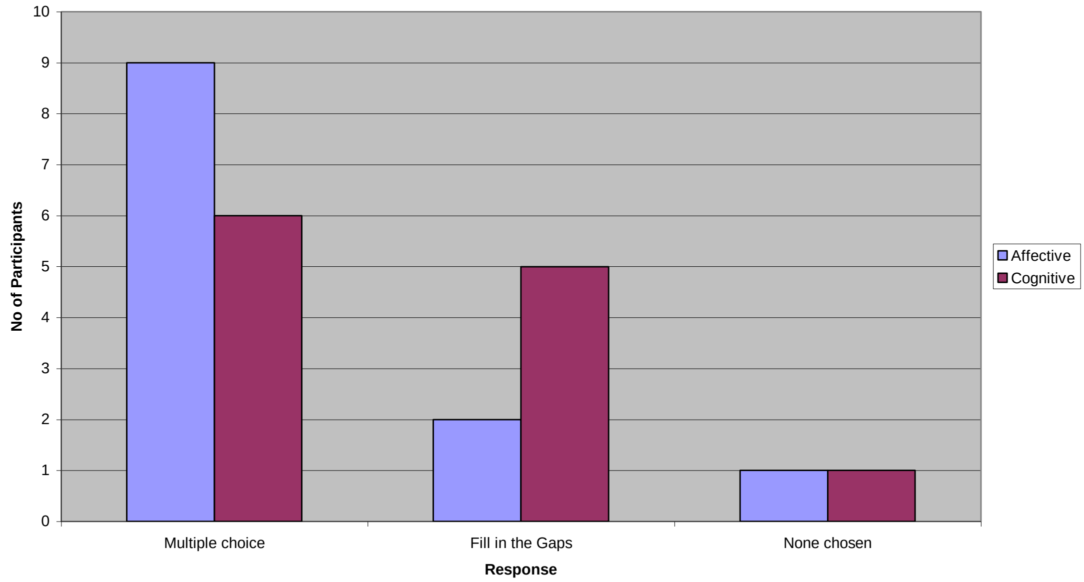
Participants completed an online survey of 14 questions based on their interaction. Responses were recorded using a Likert scale from 1(Strongly disagree) to 5 (Strongly agree). Eight participants (67%) in the affective and cognitive condition rate the usefulness of the tutoring system at 3 or higher in comparison to six participants (50%) from the cognitive condition.

Three students (25%) in the cognitive condition rated their enjoyment during use at 3 (average) or higher in comparison to nine students (75%) in the affective and cognitive condition.

The most popular type of quiz in the tutoring system is multiple choice which received 75% of the responses from the users in the affective and cognitive condition. In the cognitive condition users preferred both multiple choice and fill in the gaps at 50% and 42% respectively.

Figure A1.1 below illustrates the responses for this question while Table A1 in the appendix illustrates the rating for each type of quiz within the tutoring system.

Figure A1.1 Most enjoyable type of quiz



Users rated Julie the animated agent's facial expressions, gestures and overall appearance. Table A1.4 illustrates higher judgements by users in the affective and cognitive condition in comparison to the cognitive condition across Julie's physical features.

Table A1.4 – Q6 Reported user judgements on Julie's features

	Facial Expressions		Gestures		Overall Appearance	
Response - Likert Scale	**Affective**	**Cognitive**	**Affective**	**Cognitive**	**Affective**	**Cognitive**
Disagree Strongly	5 (42%)	6 (50%)	4 (33%)	5 (42%)	3 (25%)	6 (50%)
Disagree	1 (8%)	3 (25%)	3 (25%)	4 (33%)	2 (17%)	3 (25%)
Average	5 (42%)	2 (17%)	3 (25%)	3 (25%)	5 (42%)	2 (17%)
Agree	0 (0%)	1 (8%)	2 (17%)	0 (0%)	0 (0%)	1 (8%)
Agree Strongly	1 (8%)	0 (0%)	0 (0%)	0 (0%)	2 (17%)	0 (0%)

Eight out of twelve users (67%) in the affective and cognitive group agreed that it was useful to indicate their affective state at the end of a quiz.

Most users in the cognitive condition did not wish to continue to revise using the tutoring system in IT (34%) or out of lessons (33%), in comparison to the affective and cognitive version at 58% and 50% respectively. Some positive and negative comments from users indicate views on their interaction with the system:

"Although in the future I will think about using this very clever idea of have a fiction character to help me along the way"

"I like it is good for revision"

"I enjoyed the filling in the gaps. I think this would help us if we were to do this once a week, also if we could pick what one to do."

"her voice was quite off putting and I would be on question ten and she would still be saying "that's great well done .... " (for example). so she needs to keep up a bit more overall I found it useful. :) x x x"

"I found that sometimes that Julie would be 10 questions behind the one I was on"

"I liked the experience because it was a different fun learning experience"

"I hate her voice and she is to slow"

"I think that it was useful but next time on the fill in the gap questions I think that there shouldn't have been the button which gave you a clue by giving you the next letter because you wouldn't get that option in a test and I also think that you should have been able to choose a character instead of just been give Julie."

"Well I thought that Julie was helpful however I think she may need a different voice and stop talking all the time as it got very annoying to me"

Appendix 2: System implementation

The following appendix includes pseudo code for key parts of the ETS implementation. In addition this section includes detail on predefined ECA expressions and emotions.

Figure A2.1 Pseudo code: system initialization

Julie appears on screen, she has a neutral expression and waves at the user;
Julie gives a randomly selected welcome speech and prompts the user to log in
When the user has logged in Julie gestures at the quiz interface and verbally asks the user to read the instructions before they continue

Figure A2.2 Pseudo code: system response to correct answer

If user response is “correct” then:
> *Julie is happy; she nods and randomly selects a “positive verbal response”.*
> *Julie has a neutral expression as she waits for the next user response.*

Figure A2.3 Pseudo code: system response to incorrect answer

If the user response is "wrong" then:
> *Julie expresses concern; she leans forward, and randomly selects a negative verbal response,*
> *Julie states the correct answer and emphasises this explanation with a random gesture,*
> *Julie has a neutral facial expression as she waits for the next user response.*

Figure A2.4: Pseudo code: system responses to self-reported emotion and current score for boredom

If User reports "Bored"
If current score >=4 then:
Julie is concerned but surprised, Julie responds verbally by praising their high score and attributing the boredom to the material.
If current user score =3 then:
Julie is concerned, Julie responds verbally by encouraging the good score and acknowledging that the material is not interesting. She emphasises her point using gesture.
If current score < 3 then:
Julie is concerned, Julie responds verbally by acknowledging the user's emotion and providing metacognitive feedback on increasing effort.
She suggests an immediate action of repeating the current section. She emphasises her point using gesture.

Figure A2.5 Example user log in details (student details.txt)

```
topic=User Log users=user1:password;name:
user2:password;name2:
user3:password;name3:
```

Figure A2.6 Question set format: section 1: Information Technology, ETS version 1

(users.txt)

```
topic=Hardware and software&questions=Which of the following retains its data even after the computer is switched off?:CPA;Cache;ROM;RAM;REM:3:
What does RAM stand for?:random accessible memory;random access memory;random allowable memory;read access memory;read allowable memory:2:
What is a computer's main internal backing store?:hard disk;USB;CD-ROM;ROM;RAM:1:
How best is the size of a hard disk measured?:bytes;terabytes;kilobytes;megabytes;gigabytes:5:
What type of media is a hard disk?:optical;magnetic;laser;desk-jet;solid state:2:
```

Table A2.1 ECA expressions

Uncertain
Ironic
Blink
Surprised
Sad
Angry
Happy
Suspicious
Tragic
Satiric
Relaxed
Confused
Concerned
Charming
Smile
Kiss
Basic

Table A2.2 ECA animations

Restpose
Acknowledge
GestureDown
GestureLeft
GestureUp
GestureUser
Hearing
LookDownLeft
LookDownRight
LookDown
LookLeft
LookRight
LookUpLeft
LookUpRight
LookUp
Offended
Sad
StartListening
Surprised
Threatened

Uncertain
Wave
GetAttention
LookDown2
Announce
Confused
Congratulate
Decline
Explain
GestureRight
Pleased
Process
Search
GestureUpLong
LookRightLong
ExplainSuite
ChangePosition
Speak_01
Speak_02
Speak_03
Speak_04
Speak_05
Smack_01
Count_01
Smack_02
Count_02
Count_03
Count_04
Count_05

Appendix 3: Experiment material: Chapter 5

A3.1: Pre-post test

Mock test

Please answer the following questions. Include as much detail as possible and answer in full sentences.

Question Number:	Question:	Marks:
1	Rank the following types of computers in order of processing power. 1 2 3 4 Netbook Mainframe Supercomputer PC	4
2	What does PC stand for?	2
3	Give 1 example use of a PC.	1
4	What is a mainframe computer?	1
5	State two organizations that use mainframes and explain how	4

	they are used in these organizations.	
6	Describe how a supercomputer differs from a mainframe computer?	2

Question Number:	Question:	Marks:
7	State two organizations that use supercomputers and explain how they are used in these organizations.	4
8	Explain the 2 advantages of a laptop over a PC?	2
9	Netbooks use ______________________. Flash memory Hard disk	1
10	What does PDA stand for?	3
11	What was the original purpose for PDAs?	1
12	Describe 2 features of a tablet computer.	2

13	Give 1 example of a tablet computer and explain how it is used.	2
14	Why do some devices have an embedded computer?	1
15	List 5 devices that have an embedded computer and explain how they work.	10
16	What is another term for an embedded computer?	2

Total out of 42

The End

A3.2 Pupil guidance sheet

ICT revision with Julie

Guidance sheet:

You will be involved in a short study during the next 2 weeks. The study seeks to find out the impact of animated characters on learning. You will be interacting with the system that includes a set of quizzes and working through these during 3 of your lessons this term.

How to open the quiz system:

Go to

Rmshared/IT/Short Course/DeanesITQuizver3

Open the file **Index.htm**

Type in your allocated user name and password.

What happens next:

1. Follow the instructions given by Julie the animated character, see picture above.
2. Start answering the quiz questions.

3. After every 5 questions Julie will ask you how you feel about your score and progress.
4. At the end of quizzes, click on Finish at the top and wait for instructions on how to submit your marks.

Remember:

To save your marks, click Finish, and then click on the Submit button to save your marks. Wait for the confirmation page before you close the browser window.

A3.3 Online survey:

After you have finished with learning and revising IT Theory with the help of Julie, please fill out the following questionnaire. Most items in the questionnaire ask you to what extent you agree with a specific statement. You can choose one from five options. The options go from "1 Strongly Disagree" to "5 Strongly Agree". Most questions are about "the system", that is the system (including Julie) which you have used to learn and revise IT Theory. Thank you in advance for completing this questionnaire.

* Required

What is your approximate age? *

- ☐ 11 years
- ☐ 12 years
- ☐ 13 years
- ☐ 14 years
- ☐ 15 years
- ☐ 16 years

What is your gender? *

- ☐ Female
- ☐ Male

1 This system is a good way to learn or revise IT theory. *

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	○	○	○	○	○

2 This system would be a good way to revise some of my other subjects *

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	○	○	○	○	○

3 This would be a good system to use outside of lessons. *

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	○	○	○	○	○

4 I would use this system regularly. *

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	○	○	○	○	○

5 It is useful to reflect on my feelings after the end of a set of questions. *

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	○	○	○	○	○

6 Did you notice Julie's facial expressions? *

- ☐ Yes
- ☐ No

7 Julie's facial expressions are appropriate to this activity. *

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	○	○	○	○	○

8 Julie's facial expressions are believable. *

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	○	○	○	○	○

9 Did you notice Julie's gestures? *

- ☐ Yes
- ☐ No

10 Julie's gestures are believable. *

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	○	○	○	○	○

11 Julie's gestures are appropriate to this activity. *

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	○	○	○	○	○

12 Did you notice what Julie said? *

- ☐ Yes
- ☐ No

13 What Julie said was helpful to my learning *

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	○	○	○	○	○

14 Julie was interested in my progress *

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	○	○	○	○	○

15 Julie cares about me as an individual *

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	○	○	○	○	○

16 I liked Julie the animated character. *

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	○	○	○	○	○

17 I prefer to learn with an animated character rather than without one. *

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	○	○	○	○	○

18 I enjoy this type of activity. *

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	○	○	○	○	○

19 How much time would you prefer to spend on this type of activity in one session? *

- ☐ 0-10 minutes
- ☐ 11-20 minutes
- ☐ 21-30 minutes
- ☐ 31-40 minutes
- ☐ >40 minutes

20 What do you like most about Julie *

21 What do you like least about Julie *

22 What do you like most about learning this way. *

23 What do you like least about learning this way. *

24 Did you have any problems during the session. *

- ☐ Yes
- ☐ No

25 Please explain any problem/s that you did have.

26 Any other comments

Appendix 4: Experiment material & survey results: Chapter 6

This section includes the material used in the experiment and subjective results from the online survey.

A4.1: Interpersonal Reactivity Index questionnaire

Participants complete the Interpersonal Reactivity Index below as an online form.

IRI questionnaire

Please fill out the following questionnaire. Most items in the questionnaire ask you to what extent you agree with a specific statement. You can choose one from five options. The options go from "0 Strongly Disagree" to "4 Strongly Agree". Most questions are about your feelings in various situations. This information will be used in a study to investigate learning in IT.

---

* Required

Please enter your name and surname*

Please choose your IT class from the list below* 10S1

What is your approximate age?*

- 13
- 14
- 15
- 16
- 12

What is your gender?*

- Male
- Female

1. I daydream and fantasize, with some regularity, about things that might happen to me.*

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:					

2. I often have tender, concerned feelings for people less fortunate than me. *

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:					

3. I sometimes find it difficult to see things from the “other guy’s” point of view. *

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

4. Sometimes I don’t feel very sorry for other people when they are having problems. *

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

5. I really get involved with the feelings of the characters in a novel.*

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

6. In emergency situations, I feel apprehensive and ill at ease.*

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

7. I am usually objective when I watch a movie or play, and I don’t often get completely caught up in it. *

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

8. I try to look at everybody’s side of a disagreement before I make a decision.*

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

9. When I see someone being taken advantage of, I feel kind of protective towards them. *

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

10. I sometimes feel helpless when I am in the middle of a very emotional situation.*

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

11. I sometimes try to understand my friends better by imagining how things look from their perspective. *

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

12. Becoming extremely involved in a good book or movie is somewhat rare for me. *

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

13. When I see someone get hurt, I tend to remain calm.*

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

14. Other people's misfortunes do not usually disturb me a great deal. *

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

15. If I'm sure I'm right about something, I don't waste much time listening to other people's arguments.*

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

16. After seeing a play or movie, I have felt as though I were one of the characters. *

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

17. Being in a tense emotional situation scares me.*

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

18. When I see someone being treated unfairly, I sometimes don't feel very much pity for them.*

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

19. I am usually pretty effective in dealing with emergencies. *

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

20. I am often quite touched by things that I see happen.*

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

21. I believe that there are two sides to every question and try to look at them both.*

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

22. I would describe myself as a pretty soft-hearted person. *

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

23. When I watch a good movie, I can very easily put myself in the place of a leading character. *

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

24. I tend to lose control during emergencies.*

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

25. When I'm upset at someone, I usually try to "put myself in his shoes" for a while. *

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

26. When I am reading an interesting story or novel, I imagine how I would feel if the events in the story were happening to me.*

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

27. When I see someone who badly needs help in an emergency, I go to pieces. *

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:	○	○	○	○	○

28. Before criticizing somebody, I try to imagine how I would feel if I were in their place.*

	0 Strongly Disagree	1 Disagree	2 Neither Agree nor Disagree	3 Agree	4 Strongly Agree
Choose one of the following:					

A4.2A: Pre-post-test: IT

IT questions

Please answer the following questions. Include as much detail as possible and answer in full sentences.

Question Number:	Question:	Marks:
1	Rank the following in order of size? 1 2 3 4 5 Terabytes Kilobytes Bytes Gigabytes Megabytes	5
2	What does RAM stand for?	2
3	Describe the use of ROM	1
4	Link the following pairs of hardware on the left with media on the right. Hard disk Deskjet CD Magnetic Printer Optical	3
5	Which is the best media for backing up a school server at night? Explain your answer	2
6	Which is the best media for transferring a file from home to school? Explain your answer	2

Question Number:	Question:	Marks:
7	How is a mouse used? Give an example	2
8	How is a scanner used? Give an example	2
9	What input device would you use to create a report? Why	2
10	Which document is Magnetic Ink Character Recognition found on?	1
11	Rank the following in order from smallest to largest memory size. 1 2 3 CD-ROM, DVD, Floppy Disk	3
12	What is a graphic? Give one example	2
13	What are pixels?	1
14	What does CAD stand for?	1
15	What problem occurs when storing graphics as a bitmap?	1

16	Give 4 features commonly found in painting/graphics packages? (Explain each using at least 1 full sentence)	8
17	How would an architect produce large plans?	2

Total out of 40

The End

A4.3: Participant guidance sheet for IT:

IT revision with Julie

Guidance sheet:

You will be involved in a short study during the next couple of lessons. The study seeks to find out the impact of animated characters on learning. You will be interacting with the system that includes a set of quizzes and working through these in a lesson.

How to open the quiz system:

Julie should be already loaded on your screen. Check that you have the right version of Julie loaded up.

Index_E.html or Index_N.html

Type in your allocated user name and password.

What happens next:

1. Follow the instructions given by Julie the animated character, see picture above.
2. Start answering the quiz questions.
3. After every 5 questions Julie will ask you how you feel about your score and progress.
4. At the end of quizzes, click on Finish at the top and wait for instructions on how to submit your marks.

Remember: To save your marks, click Finish, and then click on the Submit button to save your marks. Wait for the confirmation page before you close the browser window.

A4.4: Online survey:

Changes were made to the survey used in the experiment from Chapter 5 and pupils were required to give their names, surname and class to allow comparisons with learning gain.

Survey

After you have finished with learning and revising IT Theory with the help of Julie, please fill out the following questionnaire. Most items in the questionnaire ask you to what extent you agree with a specific statement. You can choose one from five options. The options go from "1 Strongly Disagree" to "5 Strongly Agree". Most questions are about "the system", that is the system (including Julie) which you have used to learn and revise IT Theory. Thank you in advance for completing this questionnaire.

---

* Required

What is your Name and Surname* [ ]

What is your Class?* [ ]

What is your approximate age?*

- ☐ 11 years
- ☐ 12 years
- ☐ 13 years
- ☐ 14 years
- ☐ 15 years
- ☐ 16 years

What is your gender?*

- ☐ Female
- ☐ Male

1 This system is a good way to learn or revise IT theory.*

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	○	○	○	○	○

2 This system would be a good way to revise some of my other subjects*

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	○	○	○	○	○

3 This would be a good system to use outside of lessons.*

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	○	○	○	○	○

4 I would use this system regularly.*

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	○	○	○	○	○

5 It is useful to reflect on my feelings after the end of a set of questions.*

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	○	○	○	○	○

6 Did you notice Julie's facial expressions?*

- ☐ Yes
- ☐ No

7 Julie's facial expressions are appropriate to this activity.*

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	○	○	○	○	○

8 Julie's facial expressions are believable.*

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	○	○	○	○	○

9 Did you notice Julie's gestures?*

- ☐ Yes
- ☐ No

10 Julie's gestures are believable.*

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	◌	◌	◌	◌	◌

11 Julie's gestures are appropriate to this activity.*

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	◌	◌	◌	◌	◌

12 Did you notice what Julie said?*

- ☐ Yes
- ☐ No

13 What Julie said was helpful to my learning*

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	◌	◌	◌	◌	◌

14 Julie was interested in my progress*

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	◌	◌	◌	◌	◌

15 Julie cares about me as an individual*

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	◌	◌	◌	◌	◌

16 I liked Julie the animated character.*

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	◌	◌	◌	◌	◌

17 I prefer to learn with an animated character rather than without one.*

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	○	○	○	○	○

18 I enjoy this type of activity.*

	1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree
Choose one of the following:	○	○	○	○	○

19 How much time would you prefer to spend on this type of activity in one session?*

- ☐ 0-10 minutes
- ☐ 11-20 minutes
- ☐ 21-30 minutes
- ☐ 31-40 minutes
- ☐ > 40 minutes

20 What do you like most about Julie*

21 What do you like least about Julie*

22 What do you like most about learning this way.*

23 What do you like least about learning this way.*

24 Did you have any problems during the session.*

- ☐ Yes
- ☐ No

25 Please explain any problem/s that you did have.

26 Any other comments

A4.5: Survey data: collated subjective comments

Table A4.5.1: Q20 What do you like most about Julie? :Subjective comments vs. learning gain

	Neutral Condition			Empathic Condition		
Subjective Comments:	**No. Participants**	**Mean Learning Gain**	**Std. Deviation**	**No. Participants**	**Mean Learning Gain**	**Std. Deviation**
No Comment	6	6.33	6.532	9	5.22	3.866
Don't Like Her	**13**	**6.46**	**5.666**	**8**	**4.00**	**7.270**
Supportive	7	9.14	9.227	9	8.11	4.285
Feedback	**16**	**11.13**	**7.839**	**15**	**11.53**	**4.389**
Annoying / Boring	1	1.00		3	6.67	5.508
Agent Appearance	3	5.33	3.215	2	10.00	0.000
Agent Behaviours	9	11.11	6.333	23	9.48	7.537
Communication	3	8.67	2.309	0		
User Control	2	7.00	5.657	4	6.75	8.382
Easy to Use	1	12.00		0		
Online	0			1	10.00	
Total	**61**	**8.74**	**6.797**	**74**	**8.38**	**6.276**

Table A4.5.2: Q21 What do you like least about Julie? :Subjective comments vs. learning gain

	Neutral Condition			Empathic Condition		
Subjective Comments:	**No. Participants**	**Mean Learning Gain**	**Std. Deviation**	**No. Participants**	**Mean Learning Gain**	**Std. Deviation**
No Comment	3	4.67	3.055	7	7.86	6.040
Don’t Like Her	3	7.67	9.292	1	5.00	
Annoying / Boring	**10**	**4.80**	**5.594**	**7**	**6.29**	**7.088**
Feedback	2	12.50	.707	1	1.00	
Agent	1	8.00		0		
Agent Appearance	3	6.33	2.517	6	7.17	6.047
Unsupportive	2	10.50	.707	3	3.33	.577
Sound Issues	5	6.40	6.542	0		
Agent Voice	**21**	**11.19**	**8.122**	**33**	**9.94**	**5.517**
Everything	3	4.67	4.163	4	1.00	6.633
Like Her	1	15.00		1	5.00	
Agent Non Verbal Behaviour	5	9.60	6.107	6	10.33	7.285
Synchronization	2	15.50	3.536	1	24.00	
Artificial	0			1	12.00	
Quiz Feature/s	0			1	8.00	
Slow Pace	0			2	9.50	4.950
Total	**61**	**8.74**	**6.797**	**74**	**8.38**	**6.276**

Table A4.5.3: Q22 What do you like most about learning this way? : Subjective user judgements vs. learning gain

	Neutral Condition			Empathic Condition		
Subjective Comments:	**No. Participants**	**Mean Learning Gain**	**Std. Deviation**	**No. Participants**	**Mean Learning Gain**	**Std. Deviation**
No Comment	3	3.33	1.155	10	6.50	7.792
Nothing	3	7.33	3.055	0		
Supportive	1	6.00		3	13.33	5.774
Feedback	4	11.25	5.852	8	10.00	6.392
Easy/ Quick/Simple	**8**	**9.25**	**10.833**	**11**	**11.36**	**6.454**
Effective	2	3.00	0.000			
Interactive	7	12.00	5.972	8	8.25	6.274
Engaging	7	9.57	5.028	11	7.18	5.913
Revision	**5**	**15.40**	**8.264**	**7**	**9.00**	**4.933**
Online	5	10.20	5.630	1	3.00	
Multiple Choice	1	2.00		3	3.67	1.528
Don’t Like	**8**	**4.75**	**5.007**	9	6.11	6.133
Understanding	1	6.00		0		
One-to-One	1	15.00		0		
Independent Learning	3	5.67	5.508	1	8.00	
Challenging	0			1	18.00	
Agent Speech	1	12.00		1	7.00	
Quiet	1	1.00				
Total	**61**	**8.74**	**6.797**	**74**	**8.38**	**6.276**

Table A4.5.4: Q23 What do you like least about learning this way? : Subjective user judgements vs. learning gain

	Neutral Condition			Empathic Condition		
Subjective Comments:	**No. Participants**	**Mean Learning Gain**	**Std. Deviation**	**No. Participants**	**Mean Learning Gain**	**Std. Deviation**
No Comment	5	6.40	7.537	10	6.50	4.696
Everything	4	4.75	5.123	6	1.50	5.206
Unsupportive	1	9.00		1	2.00	
Annoying/ Boring	**13**	**9.69**	**7.825**	**12**	**9.58**	**4.719**
Easy	**0**			2	19.00	1.414
Too Short	2	1.00	2.828	2	12.50	2.121
Online	3	10.33	2.082			
Sound	3	14.33	11.846	1	12.00	
Interaction	0			2	16.50	2.121
No Help	0			1	17.00	
Multiple Choice	2	10.00	2.828	2	13.50	2.121
Agent	2	8.50	6.364	**7**	**8.14**	**4.100**
Distraction	0			1	24.00	
Independent Learning	1	3.00		1	1.00	
Time Delay	5	11.60	6.229	5	8.80	9.418
Nothing	8	9.13	7.180	5	5.80	4.324
Difficult Questions	3	5.33	6.110	4	9.75	7.136
Eye Strain	0			1	2.00	
Voice	0			5	7.00	7.616
Too Long	3	12.00	7.211	0		
Voice	3	5.00	3.000	0		
Explanation Unclear	1	20.00		1	13.00	
Repetitive	1	10.00		0		
Old Fashioned	0			1	10.00	
Can't Interact with Classmates	0			1	9.00	
Total	**60**	**8.83**	**6.813**	**71**	**8.54**	**6.360**

Table A4.5.5: Q25 Did you have any problems? : Subjective user judgements vs. learning gain

	Neutral			Empathic		
Subjective Comments:	**No Participants**	**Mean Learning Gain**	**Std. Deviation**	**No Participants**	**Mean Learning Gain**	**Std. Deviation**
No Comment	27	8.48	6.635	29	6.14	5.449
No Problems	11	8.82	5.231	**13**	**7.00**	**4.000**
Distracting	0			1	24.00	
Sound Issues	**13**	**12.08**	**8.450**	**5**	**9.60**	**4.278**
Submission	1	0.00		2	20.50	3.536
Difficult Questions	1	4.00		3	12.33	7.506
Time Delay	1	16.00		1	4.00	
Feedback should include targets		0		1	8.00	
Other students talking	1	2.00		0		
Explanation Unclear	1	5.00		2	10.00	4.243
Agent Speech	1	2.00		2	3.50	2.121
Other Pupils Re-started the Quiz2	0			1	15.00	
Annoying	0			4	8.25	9.323
Agent	**3**	**6.00**	**4.359**	4	15.00	4.163
Quiz Interaction	1	3.00		4	12.00	4.320
Total	**61**	**8.74**	**6.797**	**72**	**8.53**	**6.187**

Table A4.5.6: Q26 Do you have any other comments? : Subjective user judgements vs. learning gain

Subjective Comments:	No Participants	Mean Learning Gain	Std. Deviation	No Participants	Mean Learning Gain	Std. Deviation
No comment	**44**	**7.25**	**5.657**	49	7.16	5.818
Julie Pointless	0			1	15.00	
Improve Sound	0			1	11.00	
Don't Like Interaction	1	10.00	5.657	**10**	**8.60**	**8.072**
Good Way to Learn	1	12.00		2	12.00	0.000
Julie is Robotic	0			1	15.00	
More Challenging	0			1	16.00	
Change Agent Voice	2	16.00		1	20.00	
Add Mute Button	1	9.00		1	12.00	
Skip Feedback	0			2	13.00	9.899
Emotion Section Skipped	1	20.00		0		
Try Another Program	1	13.00	**2.646**	0		
Sound Annoying	1	28.00		1	8.00	
Annoying	**3**	**5.00**	1.414			
Make Easier	1	4.00		1	18.00	
Feedback	2	10.00		1	3.00	
Useful to Revise at Home	1	19.00		0		
Time Delay	1	29.00	6.797	0		
More Options on Emotions Section	0			1	7.00	
Male Character	0			1	8.00	6.276
Add emotion to Voice	1	3.00		0		
Total	**61**	**8.74**		**74**	**8.38**	