

Department of Computer Science and Software Engineering
University of Canterbury



September 2007

Fitting Free-Form Question-Asking and Spatial Ability into ITS Development

A thesis submitted in partial fulfilment of the requirements
for the Degree of Master of Science in Computer Science
in the University of Canterbury by

Nancy Milik

Supervisor: Associate Professor Dr. Antonija Mitrović¹

Associate supervisors: Dr. Michael Grimley² and Associate Professor
Dr. Timothy Bell¹

Examiner: Vania Dimitrova³

¹Department of Computer Science and Software Engineering, University of Canterbury

²Department of Education, University of Canterbury

³School of Computing, University of Leeds

To my dearest parents & brothers

Acknowledgements

Thank you God for allowing me to survive the challenges of this research, especially writing a thesis!

I'd like to thank my awesome supervisor, Dr. Tanja Mitrović, for her constant and valuable support and guidance throughout this research project; it is a privilege to work under her supervision. Thank you to Dr. Mick Grimley for his advise on the *spatial ability* aspect of this project and feedback on drafts of this thesis. Thank you also to Dr. Tim Bell for feedback on earlier states of this project.

A huge thank you to my fellow ICTGers: Amali Weerasinghe, Amanda Nicholas, Devon Barrow, Jay Holland, Konstantin Zakharov, Moffat Mathews, Nilufar Baghaei, Pramudi Suraweera and Dr. Brent Martin. Thank you so very much for all the help, support, laughs, chats, jokes, technical tips and social time. Thanks for providing entertainment, good laughs and unforgettable memories, and for just being good friends.

I'd like to especially thank Amali, Jay and Kon for all their help and support throughout this research and in particular over the past couple of months. Thank you for the various tips and technical shortcuts, and for proof-reading bits and pieces of this document, especially thanks for *trying* to keep my sanity and for all the pep-talks. Also I've to thank Kon for introducing me to WinEdt and for the numerous latex tips and tricks that produced this document.

There are not enough words to express how blessed I am with my parents and brothers; where would I be without you or your prayers. In particular to this research, thank you dad and mum for all your encouragements that made surviving this challenge possible. My talented brother Fady, thanks for designing the posters presented in this research. My baby brother Bassam, thank you for putting up with me especially during the last period. Thank you for being the best family ever!

Also I don't know how to thank my cousin Ashraf and Isis for everything they've done, may the Lord reward you in His ways.

There is so much I'd like to say here but I realise this isn't an Oscar award speech. So to sum it up thank you to everyone who I have met over the duration of this research, each one has influenced this work in one way or another. Thanks to all those who brought a smile to my face, especially over the past period.

Finally, thanks is due to google for easing the information retrieval/research aspect of this research. This research was supported by an ICTG scholarship.

Contents

| | |
|---|------------|
| Acknowledgements | i |
| List of Figures | vii |
| List of Tables | ix |
| Abstract | i |
| Chapter 1 Introduction | 1 |
| 1.1 The Learning Battle | 1 |
| 1.2 The Learning Process | 3 |
| 1.2.1 Model of Learning | 3 |
| 1.2.2 Information Processing and Memory | 4 |
| 1.3 Research Content | 5 |
| 1.4 Thesis Structure | 7 |
| Chapter 2 Background | 9 |
| 2.1 Intelligent Tutoring Systems | 10 |
| 2.1.1 Student Modelling | 12 |
| 2.1.2 Constraint Based Modelling | 13 |
| 2.1.3 Constraint-Based Tutors | 14 |

| | | |
|------------------|---|-----------|
| 2.2 | Meta-Cognitive Skills | 15 |
| 2.3 | Individual Differences | 17 |
| Chapter 3 | ERM-Tutor | 21 |
| 3.1 | The Task | 22 |
| 3.2 | Overview of ERM-Tutor | 25 |
| 3.2.1 | Student Interface | 26 |
| 3.2.2 | Problem Representation | 30 |
| 3.2.3 | Problem Solver | 31 |
| 3.2.4 | Domain and Student Models | 32 |
| 3.3 | Enhancements to ERM-Tutor | 34 |
| 3.3.1 | The Constraint Set | 35 |
| 3.3.2 | The Interface | 36 |
| Chapter 4 | Question Asking | 39 |
| 4.1 | Question Generation | 40 |
| 4.1.1 | Nature of Questions | 41 |
| 4.1.2 | Types of Questions | 43 |
| 4.2 | Relevant Research in ITSs | 45 |
| 4.2.1 | ALPS | 45 |
| 4.2.2 | AutoTutor | 45 |
| 4.2.3 | Biology Circulatory System | 47 |
| 4.2.4 | VTA | 48 |
| 4.3 | Our Question-Asking Module | 48 |
| 4.3.1 | Questions Database | 50 |
| 4.3.2 | Information Retrieval Mechanism | 51 |
| 4.3.3 | Customisation of ERM-Tutor | 53 |

| | | |
|------------------|--|------------|
| Chapter 5 | Spatial Ability | 59 |
| 5.1 | Cognitive Theory of Multimedia Learning | 60 |
| 5.2 | Spatial Ability | 65 |
| 5.2.1 | Measuring Spatial Ability | 67 |
| 5.3 | Our Spatial Ability Module | 70 |
| 5.3.1 | Feedback Messages | 71 |
| 5.3.2 | Customisation of ERM-Tutor | 73 |
| | | |
| Chapter 6 | Evaluation | 77 |
| 6.1 | Preliminary Evaluation (2005) | 78 |
| 6.1.1 | Procedure (2005) | 79 |
| 6.1.2 | Results (2005) | 80 |
| 6.1.3 | Summary (2005) | 82 |
| 6.2 | Evaluation Study (2006) | 82 |
| 6.2.1 | Procedure (2006) | 84 |
| 6.2.2 | Results (2006) | 85 |
| 6.2.3 | Summary (2006) | 90 |
| 6.3 | Evaluation Study (2007) | 90 |
| 6.3.1 | Variations in contrast to the 2006 Study | 91 |
| 6.3.2 | Procedure (2007) | 93 |
| 6.3.3 | Results (2007) | 94 |
| 6.4 | Discussion | 115 |
| | | |
| Chapter 7 | Conclusions | 123 |
| 7.1 | Research Contribution | 124 |
| 7.1.1 | Question-Asking Module | 125 |
| 7.1.2 | Spatial Ability Module | 127 |

| | |
|--|------------|
| 7.2 Future Work | 128 |
| References | 133 |
| Appendix A Information and Consent Form | 145 |
| Appendix B Domain Knowledge Tests | 147 |
| B.1 Test Version A | 147 |
| B.1.1 solution | 150 |
| B.2 Test Version B | 150 |
| Appendix C Questionnaire | 153 |
| Appendix D Spatial Ability Tests | 157 |
| D.1 Paper Fold Test | 157 |
| D.2 Card Rotation Test | 160 |
| Appendix E ITS 2006 Short Paper | 163 |
| E.1 Presented Poster | 167 |
| Appendix F NZCSRSC 2007 Paper | 169 |
| Appendix G AIED 2007 Short Paper | 179 |
| G.1 Presented Poster | 183 |
| Appendix H College Poster Competition | 185 |

List of Figures

| | | |
|-----|---|----|
| 1.1 | Model of Learning | 3 |
| 1.2 | Architecture of our memory system | 4 |
| 2.1 | Architecture of a typical ITS | 11 |
| 3.1 | Architecture of ERM-Tutor | 27 |
| 3.2 | Snapshot of the original version of ERM-Tutor | 28 |
| 3.3 | Graphical problem representation in ERM-Tutor | 30 |
| 3.4 | A snapshot of the modified version of ERM-Tutor | 37 |
| 3.5 | Steps-line showing Step 4 as the current step | 38 |
| 4.1 | Snapshot of question-asking in ALPS | 46 |
| 4.2 | Basic idea of the question asking module | 49 |
| 4.3 | Architecture of ERM-Tutor | 54 |
| 4.4 | ERM-Tutor interface with Questions module | 56 |
| 4.5 | Close up on the Questions module | 57 |
| 5.1 | The dual learning channel | 62 |
| 5.2 | An annotated illustration | 63 |
| 5.3 | Paper folding test instruction | 68 |
| 5.4 | Paper folding test example | 68 |

| | | |
|-----|---|-----|
| 5.5 | Card rotations test instruction | 69 |
| 5.6 | Card rotations test example | 69 |
| 5.7 | Example of a constraint in multimedia representation | 73 |
| 5.8 | Architecture of ERM-Tutor | 74 |
| 5.9 | Snapshot of ERM-Tutor with a multimedia feedback message | 76 |
| 6.1 | The newly added question to tests A and B | 93 |
| 6.2 | Groupings of participants | 97 |
| 6.3 | Pre- and post-test scores for the four groups | 99 |
| 6.4 | Number of attempted problems by the four groups | 104 |
| 6.5 | Learning curve for all participants | 106 |
| 6.6 | Learning curves for the four groups | 107 |
| 6.7 | Learning curves for the different groupings of participants | 108 |
| 6.8 | Number of constraints learnt for the four groups | 109 |
| 6.9 | Subjective ratings from the four groups | 114 |

List of Tables

| | | |
|------|---|-----|
| 3.1 | Summary of ER diagram notation | 24 |
| 5.1 | Principles for the design of multimedia messages | 64 |
| 6.1 | Mean system interaction details (standard deviation) (2005) | 80 |
| 6.2 | Percentage of question types asked (2005) | 81 |
| 6.3 | Evaluation study experimental design (2006) | 83 |
| 6.4 | Spatial ability test scores (2006) | 86 |
| 6.5 | Participants assignment to groups (2006) | 86 |
| 6.6 | Pre-test and post-test scores (2006) | 87 |
| 6.7 | Mean (<i>sd</i>) pre-test and post-test scores for groups (2006) | 87 |
| 6.8 | Spatial ability test scores (2007) | 95 |
| 6.9 | Participants assignment to groups (2007) | 96 |
| 6.10 | Pre-test and post-test scores (2007) | 98 |
| 6.11 | Mean (<i>sd</i>) pre-test and post-test scores for groups (2007) | 98 |
| 6.12 | Mean (<i>sd</i>) pre-test and post-test scores for all classifications (2007) . . . | 100 |
| 6.13 | ANCOVA results (2007) | 101 |
| 6.14 | Summary of means (<i>sd</i>) of system interaction results (2007) | 103 |
| 6.15 | Percentage of question types asked (2007) | 110 |
| 6.16 | Summary of means (<i>sd</i>) of subjective results for ERM-Tutor (2007) . . | 113 |

| | |
|--|-----|
| 6.17 Examples of students subjective comments (2007) | 116 |
| 6.17 Examples of students subjective comments (2007) continued | 117 |

Abstract

Intelligent Tutoring Systems (ITSs) are problem-solving environments that provide individualised instruction and are able to adapt to the abilities and needs of each individual student in order to maximise effective learning. They provide feedback on students' actions, but a problem arises when students do not always understand the feedback they receive. Therefore, it would be beneficial for students to be able to ask for additional clarifications at any time, and to receive feedback customised to their individual differences. This research focuses on providing an additional help channel in ITSs where students are able to ask free-form questions, as well as accounting for the students' psychometric measure of spatial ability.

We describe ERM-Tutor, the test-bed ITS chosen for implementing our research framework. ERM-Tutor is a constraint-based tutoring system for teaching logical database design. Students practise this procedural task in ERM-Tutor by solving each step and receiving feedback on their solutions.

We also present our approach to addressing the meta-cognitive skill of question-asking in ERM-Tutor. We added a question-asking module that enables students to ask free-form questions and receive the most appropriate answers stored in the system. In addition, we investigated the potential of tailoring the feedback messages towards the learners' psychometric measure of spatial ability. We modified ERM-Tutor to provide not only textual feedback messages, but also multimedia messages, containing a combination of text and pictures.

We performed a series of evaluation studies in order to evaluate the effectiveness of our proposed solutions. All our studies were conducted with tertiary students enrolled in

an introductory database course. The students had attended lectures on logical database design and were asked to use ERM-Tutor to develop and practise their mapping skills.

The results show an overall improvement in performance and learning gain for all students using ERM-Tutor. Interactions with the question-asking module show that most questions asked by students were task-focused, directly requesting help on specific errors. The results confirm the need for addressing students' questions inside an ITS environment. Furthermore, there were no conclusive results to support a difference in effectiveness of the textual versus multimedia feedback presentation modes with respect to the students' spatial ability. However, we observed a number of trends indicating that matching the instruction presentation mode towards the students spatial ability influences their perception of the system and motivation to use it, more than their learning gain. Our results show promising indications for further explorations.

We present our approaches, full analyses of the collected data from our evaluation studies, as well as our research contributions to the ITSs field. We also portray a number of future directions that will contribute towards maximising the effectiveness of learning in ITSs.

CHAPTER 1

Introduction

Learning, or constructing new knowledge, is an inevitable operation that faces everyone in their daily life. We have long been aware that individuals differ in their perception, processing and storage of information. In particular, we differ in how we select as well as extract the meaningful insights from the ‘big picture’, how we comprehend and make sense of any given feedback, and how we mentally represent the ‘finished’ piece of knowledge. Nonetheless there has been little research in the way of catering for our individual differences in learning. Through this research, we present our contribution to enhancing the learners’ experience by providing an additional channel for asking questions as well as catering for their spatial ability.

1.1 The Learning Battle

Let us start by defining the term ‘learning’. Some scholars have defined learning as solving, or fitting in the pieces of, an unassembled jigsaw puzzle. In this view, information is the scrambled pieces of the jigsaw puzzle and knowledge is when it is constructed into its coherent shape or scene, that is, when it makes sense together.

We constantly face the challenge of acquiring new skills and knowledge. Rapid and widespread developments in technology have made information available and easily accessible more than ever before. Such ease of access alone however, does not necessarily result in a better learning gain. This has led to an apparent trend in utilising technology to facilitate learning. Although e-learning tools, such as WebCT [WebCT and Lane, 2002], are becoming more popular in educational institutions, they still do not effectively support learning. While they make it easier for teachers to present instructional material and carry out some administrative tasks, they do not provide students with individualised feedback based on their performance, which is crucial for successful learning. An effective solution that provides adaptive pedagogical assistance for each student is Intelligent Tutoring Systems (ITSs).

ITSs are interactive computerised tutors that provide an environment where students carry out problem-solving activities, and receive feedback on their actions. As the student interacts with the system, it tracks their behaviour and produces, as well as maintains, a model of the student's knowledge. This model is used in adapting the environment towards the needs, knowledge, learning abilities and preferences of the student. This includes decisions about the timing and content of teaching actions and feedback to be presented to each individual student. Such adaptations have been shown to result in significant improvement over simplistic e-learning tools, especially in fields that require practical proficiency [Koedinger et al., 1997; Mitrović et al., 2001]. Nevertheless, this is still a growing discipline that is utilising findings in educational and psychological theories, new developments in the field of Artificial Intelligence and advancements in software and hardware technologies.

1.2 The Learning Process

If learning is about knowledge, its construction and retention, then it is necessary to take a closer look at the existing model of learning as well as the learner's memory system.

1.2.1 Model of Learning

A commonly used model of learning in today's society [Kort and Reilly, 2002] is illustrated in Figure 1.1. It begins with a set of data given to the student. This data is a collection of answers to some unasked questions. The data *becomes* information only when the appropriate question is asked, that is, in order for the data to be transformed into a piece of information, the learner must find the *Question-Answer* pairs that tie the data together. These Question-Answer pairs are organised into a structure, with new questions arising as questions are answered and stored. A deeper thought process, which includes reasoning about the new information and integrating it with previous experiences, is needed for assembling information into Knowledge.



Figure 1.1: Model of Learning

From this model it can be concluded that the process of building and correcting knowledge structures is driven by the questions we ask. Therefore, asking good questions is considered as a crucial meta-cognitive skill that plays a central role in learning. Meta-cognition is often referred to as “thinking about thinking”, which can be used to help students *learn how to learn*. In other words, meta-cognition refers to higher order

thinking that involves active control over the thinking processes involved in learning. Research has shown that those with greater meta-cognitive abilities tend to be more successful thinkers.

1.2.2 Information Processing and Memory

The learning process involves three main phases, as illustrated in Figure 1.2; sensing, processing and storing information. Firstly, we use our sensors to select the relevant information from our environment. Secondly, we organise the selected information into its logical structure, construct it into new knowledge and integrate it with our previous knowledge, all of which takes place in our working, or short-term, memory. Finally, knowledge is then transformed into permanent storage, inside our long-term memory, via the process of encoding.

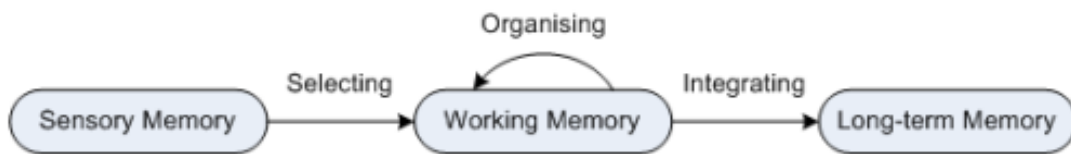


Figure 1.2: Architecture of our memory system

Successful learning requires a great deal of *attention*. Dedicated attention is needed in selecting and processing the appropriate information. Moreover, in contrast to our long-term memory, the working memory is limited in its capacity. Therefore, to maximise the learning gain, it is necessary not to overload the learner with information. It would be ideal for an effective learning environment to simplify the selecting phase and aid in the processing phase.

A possible way of making the selection of information simpler for learners is to account for their spatial ability. *Spatial ability* is the ability to visualise and manipulate

objects in space. We differ in the level of which we can spot, comprehend, and process information in its various forms. For instance, some learners may cope better with written text than with diagrams and charts. Since we react differently to the various forms of information, learning through our *most effective* form will minimise the overhead processing needed.

1.3 Research Content

One-to-one tutoring is a powerful method of promoting knowledge construction. There is substantial empirical evidence that human tutoring is extremely effective when compared to typical classroom environments [Bloom, 1984; Cohen et al., 1982; Corbett, 2001]. However, the high student-to-teacher ratio in many educational environments makes it unrealistic to provide the ideal one-to-one tutoring for every student. ITSs are an effective solution to this problem.

ITSs are believed to be a powerful technology in the educational process with the goal of maximising the learning effects. ITSs are effective learning tools due to the level of adaptive pedagogical assistance they provide. They make decisions about the timing and content of the teaching instructions and feedback to each student based on their individual state. Although ITSs provide feedback on students' actions, the students do not always understand the feedback they receive. Therefore, it would be beneficial for students to be able to ask for additional clarifications at any time, and to receive feedback customised to their individual differences.

In particular, research has proven the need for further investigations of modelling meta-cognitive skills that enhance knowledge acquisition and facilitate deep learning. It is also shown that sophisticated learning environments should stimulate learner questions and facilitate the process of receiving answers to the questions that learners ask,

thereby facilitating their ability to learn. Furthermore, it is believed that a *one-size-fits-all* scheme when dealing with learners is not the most effective nor efficient way of teaching.

In this research, we develop an environment that allows students to ask free-form questions, as well as provides feedback corresponding to the learner's spatial ability, which are aimed to improve the effectiveness of the learners by enhancing deeper/meaningful learning. There are four main objectives to our research:

- Develop an environment designed to engage learners in question-asking during problem solving.
- Evaluate the effectiveness of such environment on learning.
- Extend the feedback module to incorporate a multimedia representation for learners with high spatial ability.
- Evaluate the effectiveness of accommodating for different spatial ability levels.

Building on successful work in constraint-based tutors, the project involves incorporating a question-asking module into an existing ITS, mainly because building an ITS is beyond the main focus of this research. ERM-Tutor (Entity Relationship Mapping Tutor) [Marshall, 2004; Milik et al., 2006], an ITS developed at the Intelligent Computer Tutoring Group (ICTG)¹, at the University of Canterbury², has been chosen for its well-structured, close-ended task. The tutor allows students to design the relational schema for a given ER (Entity Relational) diagram using the standard seven-step mapping algorithm [Elmasri and Navathe, 2000].

This research looks into the implication of providing a question-asking module that allows students to ask free-form questions. It also investigates the potential of tailoring

¹<http://www.cosc.canterbury.ac.nz/tanja.mitrovic/ictg.html>

²<http://www.canterbury.ac.nz>

the feedback messages to students' ability to engage in spatial cognition. These goals are being tested through two hypotheses. The first hypothesis is that answering students' open-ended questions will clarify their understanding and result in higher performance. The second hypothesis is that presenting the system's responses tailored to the students' individual differences will lead to more effective learning and higher learning gain.

The evaluation of these hypotheses involved conducting a number of studies at the University of Canterbury, in a second-year introductory database course offered by the department of Computer Science and Software Engineering, involving students learning entity relationship mapping.

The main contribution of this research is therefore, to evaluate the added value of question-answering in combination with adaptation to the learners' spatial ability in an ITS learning environment.

1.4 Thesis Structure

The remainder of this thesis presents the context, approach and contributions of our research. Chapter 2 presents an overview of the relevant background information related to this research. In particular it describes ITSs and their typical architecture, as well as the constraint-based modelling approach to implementing them. This approach is used in implementing ERM-Tutor, the test-bed system for our research. We describe ERM-Tutor in Chapter 3, giving an overview of its problem-solving task, architecture and a number of key modifications made to enhance the learning experience while interacting with it.

Chapter 4 presents the question-asking aspect of our project. It describes the relevant background information to question-asking, focusing on the nature and importance of generating questions in a learning setting. It also discusses the question-asking module,

which was added into ERM-Tutor in order to respond to the students free-form questions while interacting with the system.

Chapter 5 details our approach to accounting for the students' spatial ability. It discusses the Cognitive Theory of Multimedia Learning, which is used as the foundation for this research in customising the instructional messages to maximise the students' effective learning.

Chapter 6 describes the series of evaluation studies conducted to test the hypotheses of this research. It presents a description of the studies, detailed analyses of the collected data and a discussion of the results. Finally, Chapter 7 concludes this thesis by presenting the conclusions of this research, as well as outlining future directions that will extend this research.

CHAPTER 2

Background

One-to-one human tutoring provides the most effective learning environment [Bloom, 1984]. However, high student-teacher ratio in traditional educational environments proves that one-to-one tutoring for every student is unrealistic. ITSs are an effective solution to this problem. ITSs are knowledge-based systems that provide an individualised learning environment, adapting to the knowledge, learning abilities and individual differences of each individual learner.

Research has found evidence that the effectiveness of ITSs is due to the level of the adaptive pedagogical assistance they provide. This ability to adapt is largely based on student models constructed by the ITS as a representation of the student's current knowledge state. In recent years, there has been an increasing interest in modelling and accounting for the student's meta-cognitive skills and abilities to enhance the learning experience. In addition, there is a shift in the field towards implementing educational and psychological theories, such as the effect of motivation on learning and how ITS can increase motivation [de Vicente and Pain, 2002].

This chapter introduces ITSs and highlights some relevant concepts to this research in the field. Section 2.1 presents a high level description of ITSs and their typical ar-

chitecture. It also describes the constraint-based modelling approach to implementing them, which is used in ERM-Tutor, the test-bed system for this research. Section 2.2 presents an overview of the meta-cognitive skill of question-asking, while Section 2.3 presents an overview of individual differences that have an effect on learning. Additional background material relevant to question-asking and spatial ability are discussed in the Chapters 4 and 5 respectively.

2.1 Intelligent Tutoring Systems

A question that is often asked is what makes ITSs intelligent? Pioneers in the field have attributed the intelligence of an ITS to its behaviour in terms of its adaptive course sequencing, problem-solving support, student diagnosis, adaptive feedback generation, adaptive problem generation, fading of scaffolding, and adaptive dialogue management.

ITSs are being developed in accordance to many educational theories, ranging from the ones that support Socratic teaching, to the ones that accommodate for collaborative learning. The underlying architecture however, remains fairly consistent across these different types. A typical ITS consists of four main components [Alpert et al., 1999]: a domain module, a student modeller, a pedagogical module and an interface, as shown in Figure 2.1. The domain module contains explicit representation of the material to be taught. This material is displayed via the interface which is the communication medium between the student and the system. Upon the student's interactions with the system the student modeller develops a representation of the student's state, keeping track of the student's progress over time. Then, based on the student model, the pedagogical module makes decisions about the timing and content of the instructions and feedback.

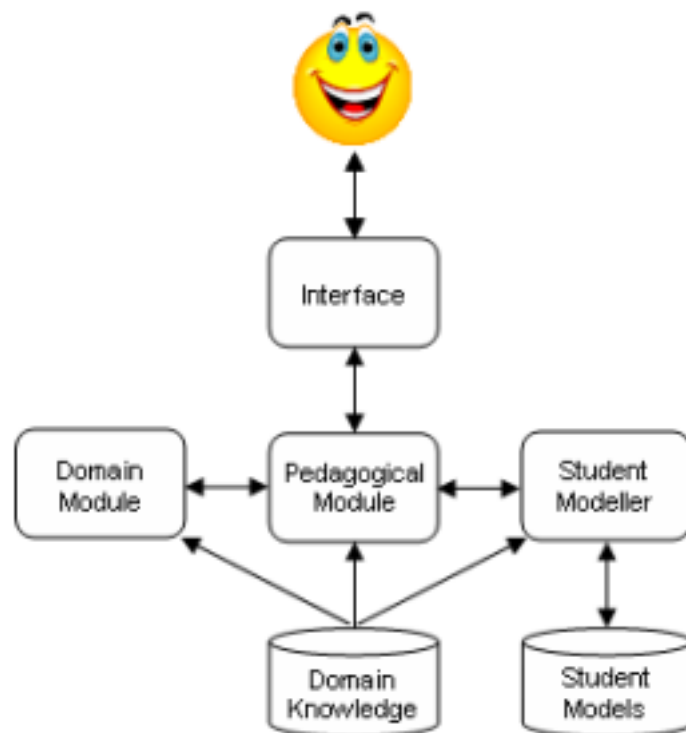


Figure 2.1: Architecture of a typical ITS

2.1.1 Student Modelling

Student modelling is the process of gathering relevant information to infer the current cognitive state of the student, thus forming a student model. The student model serves as an analogue to the representation of the student's knowledge and skill level that a human tutor develops during their interactions. Human tutors however, have an inherent advantage over computer tutors. In addition to assessing their students' skills, they observe and infer their meta-cognitive behaviour, personalities, learning styles, and affect states. They simply have a much wider variety of cues to work with when perceiving their students that the current computer tutors do not have. Although it has been shown that a student model can be useful even without being very accurate [Self, 1990], it is clear that the more accurate it is, the more effective it becomes.

Several studies have focused on investigating different aspects of the student models, including what attributes to model and how to provide means for visualisation and inspection. It is evident that modelling mouse clicks and key presses is not sufficient enough to customise the learning process for the student. Many efforts therefore, have been invested in mimicking several of the features that human tutors effortlessly use when tutoring their students.

Various methods have been utilised to construct more personalised student models. These include inducing meta-cognitive skills from student interactions with the system, for example the use of help [Aleven et al., 2004] and self-explanation [Chi et al., 1994], and using additional tools ranging from personality questionnaires to new technologies such as input speech recognition, facial expression monitoring, and eye tracking.

2.1.2 Constraint Based Modelling

Constraint based modelling (CBM) [Ohlsson, 1994] is one of many modelling techniques that are used to model the student's state as well as drive the pedagogical process. Based on the theory of learning from performance errors [Ohlsson, 1996], the domain is represented in terms of state constraints. Each constraint is an ordered pair $\langle C_r, C_s \rangle$, where C_r is the relevance condition and C_s is the satisfaction condition. The student solution is matched against C_r . The constraint whose relevance condition matches the student solution is considered satisfied if the student solution satisfies the C_s . Violated constraints signify errors that violate fundamental concepts of the domain. In other words, constraints are of the form:

“If \langle relevance condition \rangle is true, then \langle satisfaction condition \rangle had better also be true, otherwise something has gone wrong.”

For example, a constraint for driving in New Zealand roads, where the speed limit is 50km/h, could be written as:

“If \langle I am driving in New Zealand, and my speed limit is x km/h \rangle , then it ought to be the case that \langle x is less than or equal to 50km/h \rangle , otherwise something has gone wrong.”

The CBM approach reduces computational effort of student modelling to just pattern matching. It does not require a runnable expert module which are difficult to build for many domains, or extensive bug libraries which enumerate students' misconceptions about the domain. Moreover, since CBM evaluates the problem state rather than the path taken to arrive at the state, it stands robust in the face of creative solutions from students as well as inconsistent problem solving strategies.

A number of constraint-based tutors have been developed within the Intelligent Computer Tutoring Group (ICTG) at the University of Canterbury [Mitrović et al., 2004]. We present an overview of constraint-based tutors in the following subsection.

2.1.3 Constraint-Based Tutors

Constraint-based tutors represent the domain knowledge as a set of constraints. These constraints are used for evaluating the student's solution for syntax and semantic errors. This knowledge base enables the ITS to identify correct student solutions by first matching all relevance patterns in the student solution against the problem state, then testing the satisfaction conditions. If the satisfaction condition is met by the student solution, the solution is correct, otherwise it is incorrect and the appropriate pedagogical action is taken. The student knowledge is also represented as a set of constraints. The short-term student model consists of a list of satisfied and a list of violated constraints for the current solution attempt. The long-term model includes the history of usage for each constraint. This information is used by the pedagogical module to adapt the ITS towards the students needs, for example to select the next appropriate problem for the student to work on, and to generate feedback.

Constraint-based tutors have been shown to enhance learning in a variety of domains [Mitrović et al., 2001]. In EER-Tutor [Suraweera and Mitrović, 2004; Mitrović et al., 2004] students learn how to develop conceptual database schemas, and in SQL-Tutor [Mitrović, 1998; Mitrović and Ohlsson, 1999; Mitrović et al., 2002] students learn how to pose queries on relational databases. These two tutors teach open-ended design tasks, while NORMIT [Mitrović, 2002; Mitrović et al., 2004; Mitrović, 2005] teaches the procedural task of data normalisation. ERM-Tutor [Marshall, 2004; Milik et al., 2006] is another database tutor that teaches logical database design. This tutor com-

pletes the constraint-based database suite [Mitrović et al., 2004]. We chose ERM-Tutor as the test-bed for our research based on its relatively small domain-knowledge size and its well-structured, close-ended task. We present a detailed description of ERM-Tutor in Chapter 3.

Constraint-based tutors have also been developed in the area of language learning. In particular, CAPIT [Mayo and Mitrović, 2001] teaches the rules of punctuation and capitalisation in English and LBITS [Martin, 2001] teaches vocabulary. Also in the area of foreign language acquisition there is a system that teaches verb endings in a German language tutor [Martin and Nicholas, 2007]. Moreover, in the area of Object-Oriented software design, COLLECT-UML [Baghaei et al., 2005] is a tutor that encourages collaboration between students while learning UML design.

2.2 Meta-Cognitive Skills

Meta-cognition is defined in Webster's New Millennium Dictionary of English [Webster, 2004] as the awareness and understanding of one's thinking and cognitive processes. This awareness and understanding differentiates deep from shallow learning. Deep learning, as defined by Bloom [1956], requires the ability to use higher-order cognitive skills such as analysis (compare, contrast) and synthesis (integrate components into a new whole, draw relationships between concepts). Those learners who master such skills are intrinsically motivated and incorporate new ideas they are learning with existing knowledge and personal experiences, hence what they learn goes beyond getting the right answer or reproducing knowledge, and often lasts well after the course of study has ended [Bransford et al., 2000; Chi et al., 1994; Graesser and Person, 1994; Snow, 2002].

This area has been studied quite extensively in social contexts such as classrooms [Karabenick, 1998]. Theories, such as inquiry learning, have focused on encouraging learners to ask questions, formulate hypotheses, plan tests of hypotheses, collect and analyse data, explain results, and communicate findings to peers. However, it remains an interesting area for further exploration in interactive learning environments such as ITSs.

Researchers in the field of ITSs have explored various strategies to enhance meta-cognitive skills, such as self-explanation [Chi et al., 1994], help-seeking [Roll et al., 2007] and question-asking [Corbett et al., 2005]. The latter is a new concept that we are interested in exploring further.

Recent research in the field of ITSs has focused on assisting students with additional help facilities, investigating the students' behaviour towards such facilities and evaluating their impact on learning. Research has shown that learners' do not take advantage of many such opportunities because they either abuse the help function, or do not use the help function when it is appropriate [Aleven et al., 2004]. This has shifted the research to not only investigate how to get the students to learn the desired meta-cognitive skills but also how to get students to use them [Roll et al., 2007].

ITSs have the potential of responding to students' questions as well as developing the students' question generation skills. Allowing students to ask questions in a controlled learning environment, such as an ITS, will ensure that the answers returned are relevant to the problem-solving task. This is because for instance, although popular search engines, such as *google*¹, are effective in retrieving a good match to the submitted query, they do not always return the answer relevant for a particular domain. For example, when "*what is an entity*" is submitted into the google search engine, the returned results include general descriptions of an entity as a concept as well as de-

¹<http://www.google.com>

scriptions from a number of fields including law, open system architecture, J2EE SDK, database management systems (DBMS) and computer games. In contrast, a controlled environment provided by an ITS has the added advantage of assisting the student to stay focused on the problem solving task. We present our approach to incorporating question-asking in an ITS environment in Chapter 4.

2.3 Individual Differences

It is clear that a one-size-fits-all in teaching does not meet the needs of all learners. *Individual difference psychology* examines how people are similar as well as different in their thinking, feeling and behaviour [Tyler, 1965]. Moreover, establishments by modern cognitive science and educational psychology have discussed many categories of student diversity that have an impact on learning. Each person has an individual profile of characteristics, abilities and challenges that influence their perception and processing of information and learning. Examples of such diversities typically considered in research include gender, ethnicity, age, personality, motivation, intelligence, IQ, abilities, prior experiences, creativity, cognitive styles, learning styles, interests, self-efficacy, and the capacity to process information.

The literature is rich with theories and models that have been developed for classifying and addressing such diversities to maximise the individual's learning experience. In particular, *psychometrics* is the field of study concerned with the theory and technique of educational and psychological measurements. Psychometric measurements include the measurement of knowledge, abilities, attitudes, and personality traits. The field is primarily concerned with two research tasks: first, the construction of instruments and procedures for measurement; second, the development and refinement of theoretical approaches to measurement.

The first psychometric instruments were designed to measure the concept of intelligence. The commonly known historical approach is the *Stanford-Binet IQ Test* [Thorndike et al., 1986; Fancher, 1985], involving measurements of attention, memory and verbal skills. An alternative to measuring intelligence has been measuring cognitive capacities. Cognitive capacities within individuals include a general component, or general intelligence factor [Jensen, 1999], as well as cognitive capacity specific to a given domain. Psychometrics is applied widely in educational assessment to measure cognitive abilities in domains such as reading, writing, and mathematics. Examples of such assessments include the *Classical Test Theory* [Crocker and Algina, 1986], *Item Response Theory* [Lord, 1980] and *Rasch Measurement Models* [Rasch, 1980].

Another major focus in psychometrics have been on personality testing. There is a wide range of theoretical approaches to conceptualising and measuring personality. Some of the most known instruments include the *Minnesota Multiphasic Personality Inventory* [Butcher, 1992], the *Five-factor Model* (or “*Big 5*”) [McCrae and Costa Jr, 1987] and the *Myers-Briggs Type Indicator* [Pittenger, 1993]. Individual attitudes and subjective impressions have also been studied extensively in psychometrics. A common approach to their measurement is the use of the *Likert* scales.

There is an apparent trend towards the development of ITSs with pedagogical material being adaptively presented to users according to their domain-independent cognitive abilities. ITSs are essentially multimedia environments, utilising multiple forms of information content in presenting instructional messages to the users. Some of the skills that are required to engage fully in multimedia learning seem to closely resemble the definition of spatial ability. Therefore, spatial ability is of interest to us.

Spatial ability refers to the ability of mentally manipulating two-dimensional and three-dimensional objects/figures. It is typically measured with simple cognitive tests and is suggested to be predictive of user performance with some types of user interfaces

and domains [Steinke et al., 2004]. We present our approach to accounting for the students' spatial ability in Chapter 4.

CHAPTER 3

ERM-Tutor

ITSs have proven to be successful learning tools, producing significant learning gains in a variety of domains. However, there is still room for expansion and improvement. For instance, there is still a gap between the theories ITSs currently address and research in related fields such as education and psychology; that is, there are a number of theories and strategies presented in the literature that have not been utilised in the development of ITSs. Our research addresses two such theories; students need to ask questions as well as receive instruction and feedback in accordance to their spatial ability, described in detail in Chapters 4 and 5, respectively.

In order to evaluate the effectiveness of our research an implementation of the proposed environments was necessary. We decided to use an existing ITS, mainly because building an ITS is beyond the main focus of this research. ERM-Tutor (Entity Relationship Mapping Tutor) [Marshall, 2004; Milik et al., 2006], an ITS developed at the Intelligent Computer Tutoring Group (ICTG)¹, at the University of Canterbury, was chosen as our test-bed ITS for its well-structured, close-ended task.

¹<http://www.cosc.canterbury.ac.nz/tanja.mitrovic/ictg.html>

In this chapter, we present ERM-Tutor, a constraint-based tutor that teaches logical database design (i.e. mapping conceptual to logical database schemas). ERM-Tutor is a problem-solving environment, in which students practice this procedural task. The tutor allows students to design the relational schema for a given ER (Entity-Relationship) diagram using the standard seven-step mapping algorithm [Elmasri and Navathe, 2000]. ERM-Tutor complements classroom teaching; it is assumed that students have already learnt the mapping algorithm in lectures and are familiar with the database theory. Students are led sequentially through the seven steps of the mapping algorithm, through which the system analyses their solutions, and provides the students with tailored feedback messages based on their knowledge.

ERM-Tutor was first developed in 2004 as part of an honours project [Marshall, 2004]. We have made a number of notable changes to ERM-Tutor in order prepare it for use in our evaluation studies. Some of these changes are presented in this chapter.

In the next section (Section 3.1), we present a general description of the problem solving task in ERM-Tutor. Section 3.2 describes the general architecture and functionality of the system, while Section 3.3 outlines a number of the key enhancements made to ERM-Tutor.

3.1 The Task

Developing a database consistent with the user requirements is a complex task [Elmasri and Navathe, 2000]. Initially, the user requirements need to be analysed, and then a high-level representation of the database is produced; this process is called *conceptual data modelling*. Conceptual data modelling is probably the most labour intensive and time consuming phase of the development process. A database schema is a conceptual representation of the data structures that are required by a database. The data structures

include the data objects, the associations between data objects, and the rules which govern operations on the objects. The goal of the database schema is to make sure that the all data objects required by the database are completely and accurately represented. Because the database schema uses easily understood notations and natural language, it can be reviewed and verified as correct by the end-users.

A database schema is also detailed enough to be used by the database developers as a *blueprint* for building the physical database. The information contained in the database schema will be used to define the relations, primary and foreign keys. A poorly designed database will require more time in the long-term. Without careful planning for example, a database may be created that omits data required to create critical reports, produces results that are incorrect or inconsistent, and is unable to accommodate changes in the user's requirements.

The most commonly used data model for conceptual database design is the Entity-Relationship (ER) model. This model, originally proposed by Chen [1976], describes the conceptual structure of a database in the form of entities and relationships; a summary of the ER diagram notation is shown in Table 3.1. Because there are no database management systems (DBMS) based on conceptual data models, this high-level schema needs to be translated to a schema in a data model supported by the chosen DBMS; this process is known as the *logical database design*. The relational data model is usually taught in introductory database courses, as well as the algorithm for mapping ER schemas into relational ones.

ERM-Tutor teaches the ER-to-Relational mapping algorithm as defined in [Elmasri and Navathe, 2000]. The algorithm consists of seven well-defined steps. Each step in the algorithm maps one concept from the ER diagram by either creating a new relation, or altering previously created relations by adding foreign keys and attributes. The seven steps of the algorithm are as follows:

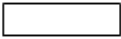
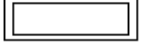













| Symbol | Meaning |
|---|--------------------------|
|  | Regular Entity |
|  | Weak Entity |
|  | Relationship |
|  | Identifying relationship |
|  | Simple Attribute |
|  | Multivalued Attribute |
|  | Composite Attribute |
|  | Derived Attribute |
|  | Key Attribute |
|  | Partial Key Attribute |
|  | 1:1 Cardinality Ratio |
|  | 1:N Cardinality Ratio |
|  | M:N Cardinality Ratio |
|  | Total Participation |
|  | N-ary Relationship |

Table 3.1: Summary of ER diagram notation (Adapted from [Elmasri and Navathe, 2000])

- Step 1: Map all regular entities and their simple attributes.
- Step 2: Map all weak entities and their simple attributes.
- Step 3: Map all 1:1 relationship types.
- Step 4: Map all 1:N relationship types.
- Step 5: Map all M:N relationship types.
- Step 6: Map all multivalued attributes.
- Step 7: Map all N-ary relationship types.

Although the algorithm is well-defined and short, students typically find it hard to learn and apply consistently. Therefore the existence of an intelligent and adaptive learning environment would be beneficial for students. The architecture of ERM-Tutor is discussed in the following section.

3.2 Overview of ERM-Tutor

ERM-Tutor is a web-based ITS, with a centralised architecture, where all tutoring functions are performed on the server side and all the data structures are also kept on the server side. Figure 3.1 illustrates the architecture of ERM-Tutor and the interaction between its components. ERM-Tutor is implemented in Lisp and runs within AllegroServe, an extensible web server provided with Allegro Common Lisp (ACL)². The main components include the pedagogical module, problem solver, student modeller, session manager and user interface. The tutor also contains a set of predefined problems along with their ideal solutions which are specified by a human expert, a set of

²Franz Inc. Allegro Common Lisp (<http://www.franz.com/>)

constraints representing the domain knowledge and student models representing each student's state.

The student interacts with the system through its user interface (shown in Figure 3.2). First, the student is required to log into the system. Upon a successful login, a session is established and the student's model is retrieved, or a new model is created for those who login for the first time. The student can choose any problem to work on. The problems are ER diagrams to be mapped into relational schemas.

The problem-solving process is broken into seven tasks, corresponding to the seven steps of the mapping algorithm. The order of the steps is fixed, and the student is required to go through the steps of the algorithm in their specified order. Each step is presented to the student on a separate page. The student has to correctly complete the current step in order to move on to the next step. Once the student has completed a step, the interface for composing the solution for the next step is presented, and the student has access to all previously mapped relations. The student can at any time ask ERM-Tutor to analyse their solution. The student is also free to change the problem at any time.

Each of the major modules is described in more detail in the following subsections.

3.2.1 Student Interface

The interface, shown in Figure 3.2, is the communication medium between the student and the system. Students are able to view problems, work on their solutions and receive feedback. The problem-solving area is the main part of the page, and its general layout is the same for all steps. First, there is a short description of the student's task for the current step. For example, for step two the task text reads "*Map all weak entities*". This is basically to remind the student what is required at this step, rather than be educational

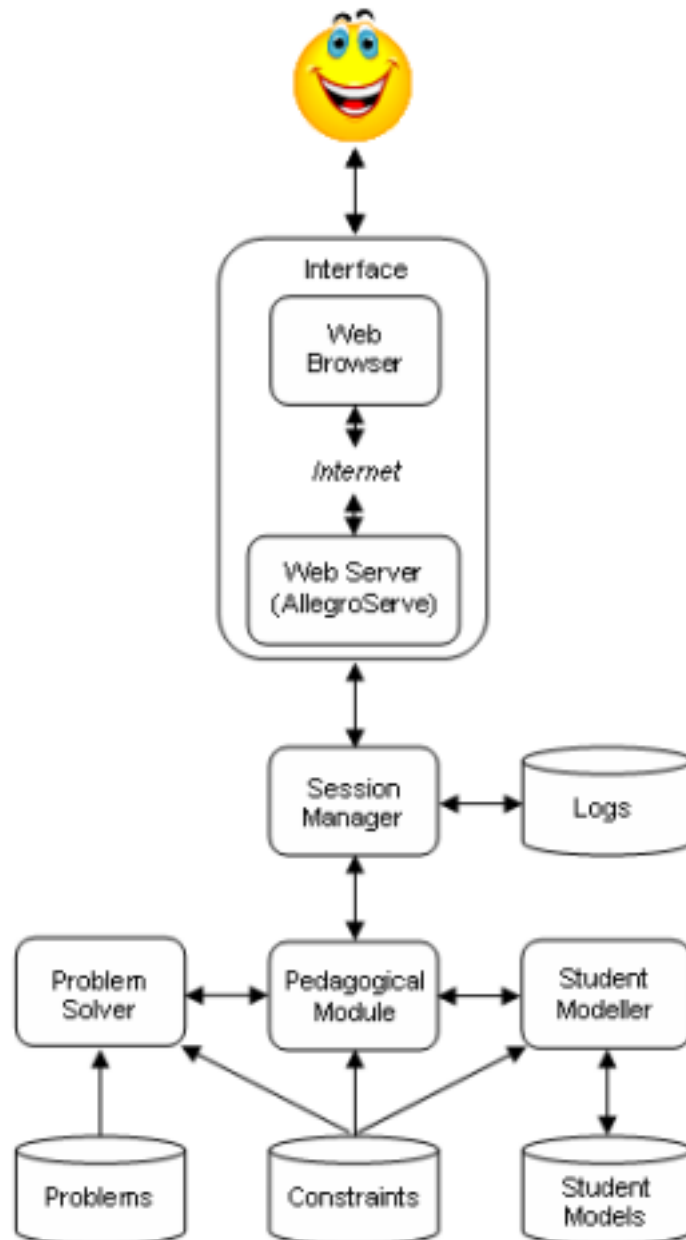


Figure 3.1: Architecture of ERM-Tutor

material in its own right. The problem is presented to the student as an ER diagram, but the student also has an option of seeing a textual description of the database, by clicking the Problem Text button. Underneath the diagram, brief instructions on what is expected in this step and how to use the input boxes to create or alter a table are presented. At any time, the student can view the solution developed so far by clicking the Completed Tables button; this pops up a window containing all the relations defined by the student for the current problem in the previous steps.

The screenshot shows the ERM-Tutor interface. At the top, there is a navigation bar with buttons for "Problem Text", "Completed Tables", "Change Problem", "Help", and "Logout". The main area is divided into a left sidebar and a central workspace.

Step 1: Map all the regular entity types

The central workspace displays an ER diagram with two entities: "STUDENT" and "HALL". "STUDENT" has an attribute "Number". "HALL" has attributes "Name" and "Address". They are connected by a relationship "LIVE_IN" with cardinalities "N" and "1".

Instructions: Choose the entity you want to map, then specify each attribute you want to add to the table you have created, use the checkboxes if the attribute is a key or foreign key.

Entity to map: [Input field] [Create table]

Table attribute: [Input field] [Add attribute]
 Key
 Foreign Key

Current table: HALL [Delete table name]

| | |
|--------|---------|
| name | address |
| Edit | Edit |
| Delete | Delete |

Feedback Level: [Simple Feedback] [Check table] [Check step] [Clear]

Feedback: 1. Choose an attribute of the entity to be a key for this relation.

Figure 3.2: Snapshot of the student interface in the original version of ERM-Tutor

The interface provides the student with the working area to create or alter one relation at a time. Each step of the algorithm is broken into subtasks. For example, in step

one, the student maps one regular entity type at a time, and the system checks the resulting relation before moving on to the next entity type. Figure 3.2 illustrates a situation when the student has mapped the HALL entity type, and has specified a relation, with the same name, with two attributes (Name and Address). For each attribute, the student can specify whether it is a primary or foreign key. When the student completes the relation, they can request the system to check the solution. If there are any mistakes in the solution, ERM-Tutor provides feedback to the student. In Figure 3.2, for example, the system informs the student that a primary key needs to be specified for the HALL relation. If the solution is correct, the student can move on to the next entity type, or to the following step of the algorithm. The interface displays to the student a record of each correct relation for the current step.

When submitting a solution, the student can also specify the desired level of feedback. There are five levels of feedback available, from the most general to the most specific. Simple Feedback simply indicates whether the submitted solution is correct or not. If it is incorrect, the number of errors is given. Hint gives a clue to remedy the first violated constraint. Detailed Hint provides more details on the first violated constraint. List All Errors provides the student with a list of errors (i.e. constraint violations), each with its detailed explanation. Finally, Solution displays the ideal solution.

For each new step in a problem, the pedagogical module (PM) starts the student at the Simple Feedback level. For each subsequent feedback request for the same step, the PM increments the feedback level until it reaches List All Errors, which it will reside on for the remaining attempts on that step. Ideal solutions are only available on student's request; PM will not reach the Solution level automatically. When the submitted solution is correct, regardless of the feedback level, the PM shows the message "*Well done! Your answer is correct.*"

The feedback/help area occupies the top right side of the screen. A help page is displayed to the student by default when a task page is first displayed. This provides a textual description of how to use the interface. When the student submits a solution, the help page is replaced by the appropriate feedback. The help page can be redisplayed any time by clicking on the Help button.

3.2.2 Problem Representation

The problems are stored in a problems definition file that is loaded at the initialisation of the system. Each problem has a unique number and name, the name of the image file associated with the problem, a textual description of the problem, and the textual representation of the problem. The textual representation of a problem is stored as four lists: entities, relationships, attributes and connections. Each of the lists is made up of entries where the beginning of each entry is delimited by a “@” followed by a unique identifier for its type. For example, the ER diagram in Figure 3.3 (also shown in Figure 3.2) is represented internally as illustrated in Listing 3.1.

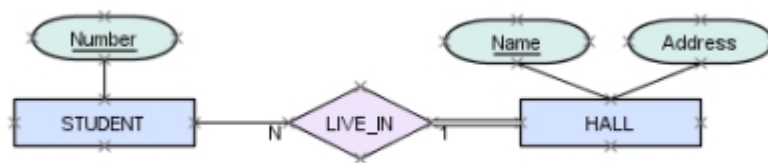


Figure 3.3: Graphical problem representation in ERM-Tutor

Currently there are 27 problems in ERM-Tutor. New problems can easily be added to the system by altering the problem definition file and supplying a corresponding image of the ER diagram. The current problems are taken from the EER-Tutor problem set.

```

1  (5 ;;unique identifier
   "Some students live in student halls. Each hall has a name (unique) and
     an address. Each student has a number (unique). Assume that there are
     students living in every hall." ;;textual description
3  (("ENTITIES" "@ E1 STUDENT regular @ E2 HALL regular")
   ("RELATIONSHIPS" "@ R1 LIVE_IN regular")
5  ("ATTRIBUTES" "@ E1K1 Number key simple E1 @ E1S1 Name simple simple E1 @
     E2K1 Name key simple E2 @ E2S1 Address simple simple E2")
   ("CONNECTIONS" "@ total 1 R1 E2 @ partial N R1 E1")) ;;textual representation
7  "5.jpg" ;;image filename
   "Student Halls" ;;problem name

```

Listing 3.1: Internal problem representation in ERM-Tutor

EER-Tutor is a tutoring system in which students learn to create an Entity-Relationship diagram from a textual problem description. The next logical step in database creation is to map the diagram into relational tables. As this is the focus of ERM-Tutor, it is a natural progression after using the EER-Tutor. Using the same problems means the students can see the progression of a database problem description through to relational database tables.

3.2.3 Problem Solver

As mapping ER schemas to relational ones is a well-structured, close-ended procedural task, it was possible to implement a problem solver to dynamically generate ideal solutions, rather than having to specify them manually for each problem. In the mapping algorithm, often there is more than one correct way of mapping the relations. For example, there may be more than one option for a student to choose as the primary key for a relation, and this is a choice that follows through and affects future steps in the algorithm when specifying foreign keys. For this reason, it is important to have a problem solver in the system that generates ideal solutions, rather than relying on pre-specified ideal solutions, which can not take into account earlier decisions of a student.

The problem solver comprises of fourteen different functions, two for each step of the algorithm. The first function deals with the creation or alteration of a relation. It typically depends on the student having specified an element of the ER diagram they are attempting to map. For example, in step one the student must specify the regular entity type they are working on. The problem solver then performs the mapping, which is used to diagnose the student's solution and is also shown to the student as the ideal solution when requested.

The second function determines all the elements from the ER diagram that need to be mapped. For example, for step 6 the problem solver will identify every multivalued attribute that needs a relation created for it. This is used to check whether a student's solution is complete or not.

3.2.4 Domain and Student Models

As ERM-Tutor is a constraint-based tutoring system, its knowledge base is represented as a set of constraints. The constraints are restrictions on correct solutions in the domain, and therefore represent the basic principles of the domain. Originally there were 121 constraints in the domain model; this was later extended as described in the next section. Syntactic constraints deal with the syntax of a student's solution independently of the problem. Semantic constraints, on the other hand, are concerned with the relationship between the student's and ideal solution, and therefore check the semantics of the problem. Although semantic constraints make up the majority of the constraints in ERM-Tutor, they are problem independent; they do not contain any elements of problems directly.

Each constraint consists of a relevance condition, a satisfaction condition and two feedback messages. For example, constraint 21, shown in Listing 3.2, is relevant for step

2 (mapping weak entity types), when there are some weak entity types in the problem (i.e. in the ideal solution, IS), and the student has produced a relation corresponding to a weak entity type, and an identifying relationship type. In that case, the student's solution is correct if there is a foreign key which corresponds to the key of the owner entity type. The hint message is given to the student when the feedback level is set to Hint, and the explanation is given for the Detailed hint and List all errors levels.

```

(21 ;;unique identifier
2  "Make sure you include a foreign key." ;;hint message
3  (and (equalp (current-task SS) 'step2) ;;relevance condition
4    (not (null (mapped IS)))
5    (not (null (current-table SS)))
6    (bind ?t (current-table SS))
7    (match '(?*d1 "@" ?tt ?t "weak" ?*d2)(entities SS) bindings)
8    (not (null (current-rel SS)))
9    (bind ?r (current-rel SS))
10   (match '(?*d3 "@" ?rt ?r "ident" ?*d4)
11     (relationships SS) bindings)
12   (match '(?*d5 "@" ?p1 ?c1 ?rt ?tt ?*d6)
13     (connections SS) bindings))
14 (not (null (current-fkey SS))) ;;satisfaction condition
15 "Step 2" ;;step number
16 "The key of the owner entity must be included as a foreign key in the new
    relation." ;;explanation

```

Listing 3.2: Example of a constraint

The short-term student model is the result of the solution diagnosis, and consists of a list of satisfied and a list of violated constraints. This model is used by the pedagogical module to present feedback to the student. ERM-Tutor also maintains a long-term student model, which records the history of each constraint as well as other useful information that can be used in visualising the student's state, such as a list of the attempted and solved problems.

3.3 Enhancements to ERM-Tutor

We modified ERM-Tutor extensively in order to prepare it for running our evaluation studies. From a user's perspective, the interface should be self explanatory, appealing and easy to use. Furthermore, it is vital that the system is robust, consistent and accurate as much as possible. This is important in order for us to have a *good* working system that will appeal to students to use and hence yield enough data for testing our hypotheses.

The modifications were based on what we experienced, as well as observed from watching students interact with the system during our evaluation studies. We also took into account the students' feedback from the questionnaires submitted from the evaluation studies.

In particular, it was apparent that although restricting the students to follow the step process was helpful in learning the steps of the algorithm (in fact students praised it in their comments), it took some time for the students to realise that each *concept* can only be mapped in its allocated step. Although the current step number, description of its task and a short set of instructions are always shown in the interface, as shown in Figure 3.2, the students did not attend to them and were often *lost* in the order of steps. Moreover, it was frustrating for students to discover that they have been mapping an incorrect concept for the step after spending sometime working on it, and to have to clear their input, navigate to the correct step and re-enter their solution. For this reason we decided to clarify following the step process, which is reflected in our modifications to the interface.

The modifications can be categorised into:

- The constraint set
- The interface

3.3.1 The Constraint Set

We enhanced the constraint set to include more constraints, making the system more robust, as well as addressing additional domain representations. For instance, we added constraints for testing whether the current step is necessary for the problem (i.e. whether there are any constructs to be mapped in that particular step), as opposed to just testing whether or not the entered label correctly matches that in the ideal solution. For example, if a student maps a construct in Step 3 but there are no 1:1 relationships in the current problem, the new feedback message they receive is *“For Step 3 of the algorithm, only map the 1:1 relationships. Check if there are any 1:1 relationships to map!”*, as opposed to receiving *“The relationship you map must have the same name as a relationship in the ER diagram.”* This provides more specific feedback to help the student in recognising and correcting their mistakes.

We also modified a number of constraints to be more specific in diagnosing the student solutions. For example, we added specific checks testing whether there are missing attributes and whether the entered attributes are the correct ones for the current step, as opposed to just checking whether the submitted attributes match those in the ideal solution. Moreover, we modified a number of the feedback messages to include the current step number, making it more obvious to the student which step they are currently working on. For example, if the student does not complete all tables for a particular step, they will receive a message similar to the message shown in Figure 3.4: *“Almost there - there are still one or more tables that need to be completed for this step (Step 4).”* Also, the *“Well done!”* messages have been modified to include the name of the particular construct that was successfully completed. For example, *“Well done! You have mapped the Hall regular entity correctly!”* is given as soon as the Hall regular entity is mapped successfully. These modifications make it easier for the

students to know where they are currently at in the problem-solving task and aids them in formulating what is required to successfully complete the mapping task.

3.3.2 The Interface

As shown in Figure 3.2, the original version of ERM-Tutor presented the students with all the input boxes required at all times for the current step. For example, for Step 1 the text-boxes along with the Create table and Add attribute buttons were presented at the same time although the student could only submit one entry (click on one button) at any one time. To clarify the process of specifying the various constructs we changed the interface to initially only show the input box to Create table, and only when a table is created the input box to Add attribute is shown to the student. The student is able to modify the entered labels by clicking on the Edit links provided. Figure 3.4 shows the modified interface for Step 4, where the student is expected to first enter the name of the relationship to be mapped, then the interface will present the next input box for entering the name of the table to be modified, followed by the attributes to be added.

Furthermore, we modified the navigation buttons displayed in the Feedback Level area, shown in Figure 3.2, to clarify the navigation process between steps. In particular, when a step's page is first displayed to the student, the student is presented with two navigation buttons: the Next step and Check problem buttons. The student is able to click on Next step if they believe that there are no constructs to be mapped for the current step. The Check problem is used when the student believes that there are no more steps required for solving the current problem, that is, rather than navigating through each step and clicking Next step, they are able to click once on Check problem. As the student is not currently mapping nor modifying any tables, Check table button is not shown. When a student starts working on a particular step the navigation area is updated to show the

ERM-TUTOR

| Problem Text | | Completed Tables | | Change Problem | | Help | | Logout |

Step 1 Step 2 Step 3 **Step 4** Step 5 Step 6 Step 7

Step: 4. Map all the binary 1:N relationship types

Instructions: Specify the relationship you want to map, then the table you will alter to do this. Specify each attribute you want to add to the table you have created, use the checkboxes if the attribute is a key or foreign key.

Relationship Relationship

Next step Check problem

Feedback
Almost there - there are still one or more tables that need to be completed in this step (Step 4).

Figure 3.4: A snapshot of the modified version of ERM-Tutor

Check table button. Once the table is successfully completed for the step, the page is refreshed to show its initial input boxes and navigation buttons, with the difference of showing the names of the successfully completed tables for the current step. Figure 3.4 shows the interface when a student clicks on the Next step button before mapping the 1:N relationship presented in the ER diagram.

We also included an additional graphical *cue* that will easily alert the students to the current step they are working through while solving a problem. We called it a *steps-line*. The *steps-line* was added to the top of the problem area to visually indicate to the students which step they are currently at. As shown in Figure 3.5, the current step is shown in bold on the steps-line with an arrow pointing towards it. The previous steps are shown greyed out to indicate that they have been solved and can not be revisited while solving this problem. The following steps are also shown to the students to keep track of where they are currently at, stressing the order of the seven procedural steps. Figure 3.5 shows the interface with our added *steps-line*.



Figure 3.5: Steps-line showing Step 4 as the current step

CHAPTER 4

Question Asking

The process of building and correcting knowledge structures is proposed to be driven by the questions we ask. Research has shown that *good* help-seeking behaviour, such as asking good questions, is a crucial meta-cognitive skill that plays a central role in learning. Although the act of asking questions has the potential to greatly facilitate the learning process, question generation by students is infrequent in traditional settings, such as a classroom [Graesser et al., 2005; Graesser and Person, 1994; Cohen et al., 1982]. When students do ask questions, the questions are about how to behave in the classroom rather than requests for meaningful explanations. Moreover, most of the questions directed toward the students by teachers are shallow level questions concentrating on factual information that can be memorised. Schools are oriented toward telling students answers and rewarding the repetition of answers, rather than rewarding them for good questions and allowing them to figure things out for themselves.

Research in the field of ITSs has shown that interactive computerised tutors are effective for enhancing learning [Anderson et al., 1995] as well as developing such meta-cognitive skills [Gama, 2004]. It is evident that meta-cognitive skills are mastered only through experience [Brown, 1987]. ITSs have the potential of engaging the students

more actively in learning, by allowing them to generate questions and develop their question-asking skill that fosters deep learning. This research incorporated a question-asking module into an ITS that tests the hypothesis that answering students' open-ended questions will clarify their understanding and result in higher performance.

This chapter presents the meta-cognitive skill of question-asking and details our approach in incorporating a question-asking module into ERM-Tutor. Section 4.1 elaborates on the importance of question generation to learning, as well as presents an overview of the nature of questions and highlights a number of question types. Section 4.2 describes the relevant research on question-asking in ITSs. Section 4.3 details our design decisions and implementation of the question-asking module, which enables students to type in their free-form questions and receive the system's answers.

4.1 Question Generation

Question-asking is still a young concept in the ITSs field. Previous research has yielded some promising results with plenty of scope for further exploration [Anthony et al., 2004]. Researchers in cognitive science, psychology and education have reported learning benefits for environments that encourage students to generate questions [Beck et al., 1997; Dillon, 1988; King, 1994; Miyake and Norman, 1979; Pressley and Forest-Pressley, 1985]. Question generation is believed to be a primary attribute of active learning which reveals how deeply the learner has mastered the material and even shifts the students' goals from performance-orientation toward learning-orientation [Graesser and Olde, 2003; Otero and Graesser, 2001; Scardamalia and Bereiter, 1985; Wisher and Graesser, 2005]. However, it is suggested that not all processes of generating questions achieve this.

Research suggests that only good questions facilitate the learning process. *Good* questions, also known as deep questions, are defined as those that require students to use higher-order thinking or reasoning skills. *Shallow* questions, on the other hand, concentrate on factual information that can be memorised. Sanders [1966, p. ix] states, “Good questions recognize the wide possibilities of thought and are built around varying forms of thinking. Good questions are directed toward learning and evaluative thinking rather than determining what has been learned in a narrow sense.”

While some studies and popular belief favour asking high-level-cognitive questions, other studies reveal the positive effects of asking low-level cognitive questions. Gall [1984, p. 41], for example, cited that “emphasis on fact questions is more effective for promoting young disadvantaged children’s achievement, which primarily involves mastery of basic skills; and emphasis on higher cognitive questions is more effective for students of average and high ability. . .”

Therefore, it is important for our question-asking module to accommodate a combination of low-level-cognitive and high-level-cognitive questions. The hope is that students’ question asking skills and the resulting learning will radically improve when they are immersed in learning environments that encourage question-asking and provide good answers.

4.1.1 Nature of Questions

Research indicates that question-asking plays a vital role in learning. Although there are different types of questions, which are generated in different ways, they share common characteristics. Questions are said to point to holes in our memory structures that we wish to fill. They provide a starting point for processing and integrating new information into memory, tie old information together in new ways, create new paths for retrieving

stored knowledge, and correct our faulty generalisations. Schank and Cleary [1995] have even suggested that until we ask a question, we are unable to integrate an answer into our memories. Further, the more questions we ask about an item, the more ways we index that item in our memories. Better indexing allows our memories to be more flexible. So the more questions we ask, the more easily we can recall the items that we require from our memories.

Classroom question generation between a teacher and students undoubtedly plays a major role in the domain of help seeking behaviours. Although research has shown question-asking to be a crucial meta-cognitive skill, question generation by students is infrequent in traditional settings, such as a classroom [Graesser and Person, 1994]. One possible reason that students perform more efficiently in traditional tutoring sessions is attributed to the quality of the questions that the tutor directs towards the student. However, it has been shown that only a small percentage of teacher-generated questions in the classroom are deep questions. Most of the questions directed to the students are shallow level questions designed to test the students' explicit memory. This indicates a need to further explore the types of good questions asked as well as how to encourage students to develop and utilise their question-asking skills.

Moreover, since question generation has been proven to be an integral aspect of learning, it would be beneficial for learners to be able to generate their own questions. In other words, question generation is more effective to learning than merely receiving the information. The act of generating questions by the learner requires a different set of skills as well as different mental processing. For this reason, presenting the learners with a glossary of terms or a list of Frequently Asked Questions (FAQ) will not contribute to their deep learning as well as them explicitly formulating their own questions.

4.1.2 Types of Questions

It is believed that effective questions open the door to knowledge and understanding. The art of questioning lies in knowing which questions to ask when. When a question is generated it is used purposefully to achieve well-defined goals. Through the art of thoughtful questioning the learners are able to connect concepts, make inferences, increase awareness, encourage creative and imaginative thought, aid critical thinking processes, and generally explore deeper levels of knowing, thinking, and understanding.

In general, all questions can be categorized into two main types: open and closed questions. Open questions elicit a wide range of answers. They do not invite any particular answer, but open up discussion or elicit a wide range of answers for creative problem solving. Closed questions on the other hand, are specific and must be answered with a yes or no response.

A common classifier of questions used is Bloom's Taxonomy [Bloom, 1956]. It is a hierarchical system of ordering thinking skills from lower to higher, with the higher levels including all of the cognitive skills from the lower levels. Below are the levels of the taxonomy, a brief explanation of each one, and examples of questions which require students to use thinking skills at each level. Below are the levels of the taxonomy.

Knowledge Remembering previously learned material, for instance, definitions, concepts, principles and formulas. For example, *What is the definition of x?*

Comprehension Understanding the meaning of remembered material, usually demonstrated by explaining in one's own words or citing examples. For example, *What does x mean?*

Application Using information in a new context to solve a problem, to answer a question, or to perform another task. The information used may be rules, principles, formulas, theories, concepts, or procedures. For example, *How does x explain y ?*

Analysis Breaking a piece of material into its parts and explaining the relationship between the parts. For example, *How does x compare to y ?*

Synthesis Putting parts together to form a new whole, pattern or structure. For example, *How are x and y related to z ?*

Evaluation Using a set of criteria, established by the student or specified by the instructor, to arrive at a reasoned judgment. For example, *How well does x measure up?*

A number of other classifications branched from Bloom's taxonomy. For instance, instead of referring to a specific level of the taxonomy people refer to *lower-level* and *higher-level* questions or behaviours. Lower-level questions are those at the knowledge, comprehension, and simple application levels of the taxonomy. Higher-level questions are those requiring complex application, such as analysis, synthesis, and evaluation skills.

As shown, each type of question has its purpose and value to knowledge construction. Whether the question asked is factual, convergent, divergent, evaluative or a combination, it still contributes to the learning process. As individuals benefit from different instructional methods, they also construct their own knowledge through asking different types of questions. Therefore, it is important to include a combination of such questions in a question-asking environment that aims to enhance the individual's learning experience and caters for the individual needs of each learner.

4.2 Relevant Research in ITSs

The nature of student questions in ITSs is largely unexplored. Four studies are detailed in this section. The first is ALPS [Anthony et al., 2004; Corbett et al., 2005], which is implemented in an environment where students are in control. The second is AutoTutor [Graesser et al., 2005], which gives the system full control, employing Socratic dialogues. The third study addresses the students behaviour after asking a question [Sullins et al., 2007]. The fourth and final is VTA [Heiner, 2007], a proposed system for answering students' questions while learning the JAVA programming language.

4.2.1 ALPS

ALPS [Anthony et al., 2004; Corbett et al., 2005], shown in Figure 4.1, is an Algebra Cognitive Tutor that was used in exploring question-asking patterns in ITSs. It allows the student to ask any question, to which the system replies with a pre-recorded video clip, using Synthetic Interview (SI) technology, which provides an illusion of a face-to-face interaction with an individual. The results show that students do ask questions, but the rate of unprompted questions is lower than in the case of one-on-one human tutoring. Furthermore, half of the questions are not related to problem-solving, but are rather social interactions. Of the remaining questions, there are many which are performance-oriented, and not deep questions that would facilitate learning.

4.2.2 AutoTutor

The Institute for Intelligent Systems at the University of Memphis recently has developed a version of AutoTutor, a physics tutor, to handle a range of student questions by extracting the answers from electronic textbooks [Graesser et al., 2005]. The ques-

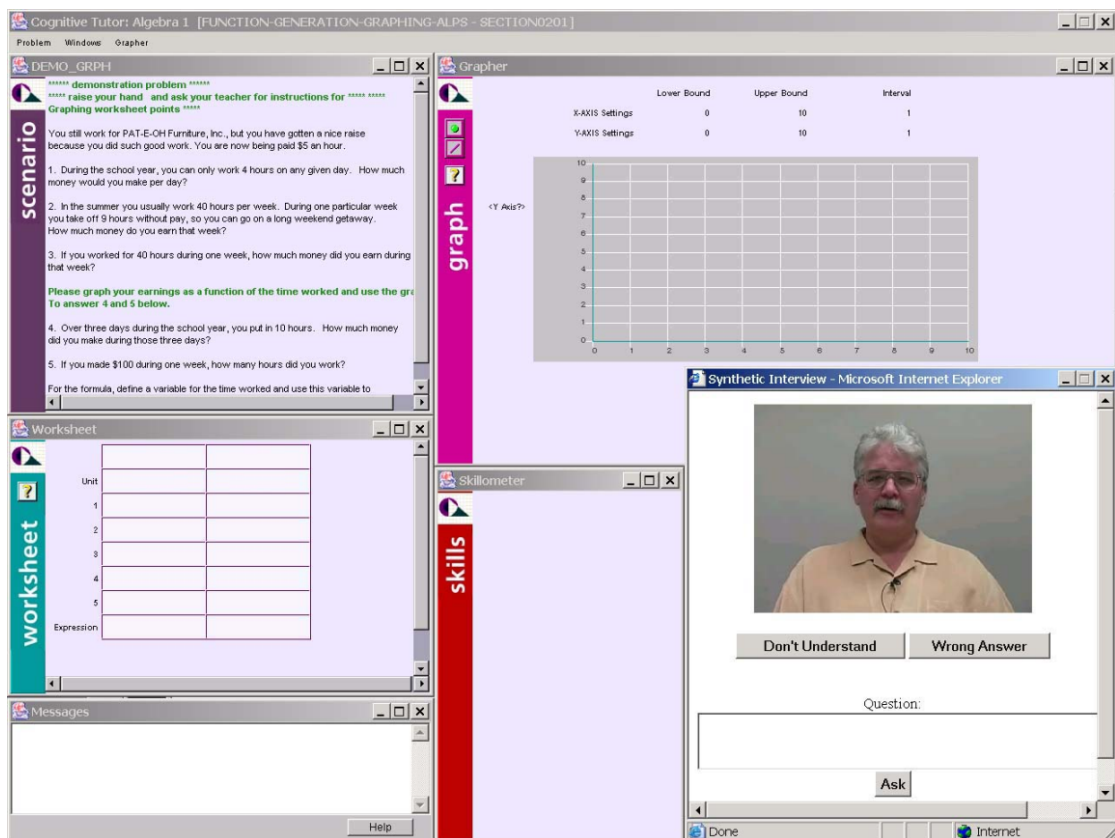


Figure 4.1: Snapshot of question-asking in ALPS (Figure from <http://www.cs.cmu.edu/alps/about4.html>)

tion answering system in AutoTutor produces answers to domain questions that are not hand-crafted by lesson planners who generate the curriculum script. More specifically, the answers are composed by first interpreting the question and then fetching a paragraph from the electronic textbook that includes an answer to the question. The question answering system classifies questions into 16 categories, including definition, comparison, and deep comprehension questions, for example why, how, and what-if questions. The evaluations involved students posing questions during learning and then rating how relevant or informative the information is in the paragraph that gets returned. Moreover, the results presented are related to the quality of the returned answers rather than the effectiveness of the question-asking process.

4.2.3 Biology Circulatory System

This study looked into the frequency and quality of student generated questions while using a Biology hypermedia learning environment [Sullins et al., 2007]. The study is a *think aloud* study with a human tutor present to prompt students to keep on verbalising their thoughts as well as to attend to any questions they ask. This initial study compares two groups; those who improved between the pre- and post- tests of 5 points or less, called low shifters, and those who improved by more than 5 points, called high shifters. The results show no difference in behaviour between the two groups being compared. Moreover, there no difference was found in the number of questions asked nor their quality; whether *deep* or *shallow*. The study reports a difference however, in the self-regulatory processes between the two groups after asking a question. The high shifters were better at *judgement of learning*, for example stating that they understand, and *feeling of knowing*, for example being aware that they have learnt something similar or related in the past. Suggesting an ideal situation where the tutor prompts the student

with the appropriate self-regulatory technique. Although this study presents an interesting concept from a meta-cognitive perspective, it still warrants further investigations.

4.2.4 VTA

VTA [Heiner, 2007], stands for Virtual Teaching Assistant, is a proposed question answering system that mediates the question answering process while learning the JAVA programming language. The software allows the students to type in their questions and upload their source-code. The system then sends the question and source-code submitted to a human expert, in this case a teaching staff member, who answers the question through the interface. VTA presents the answer to the student and stores it along with its question in a database. For subsequent questions, VTA checks whether a similar question exists in its database, and if so it will be retrieved and presented, bypassing the human expert. The authors propose to build a matrix classifying the questions to help retrieving them, but the details and results are yet to be published. Although this is still a proposed system, we suspect that by getting the students to upload their source code, they will be lead to think on low-level/event-specific details of solving their particular problem.

4.3 Our Question-Asking Module

The basic idea behind our question-asking module is portrayed in Figure 4.2. Our approach of how the module operates can be broken into five steps as follows. First, the students submit their free-form questions through the provided interface. Second, the server receives the question and sends it to the information retrieval system. Third, the information retrieval system parses the submitted question and retrieves the most ap-

propriate question-answer pair stored in the database. Forth, the information retrieval system returns the question-answer pair to the server. Finally, the interface displays the question-answer pair to the student.

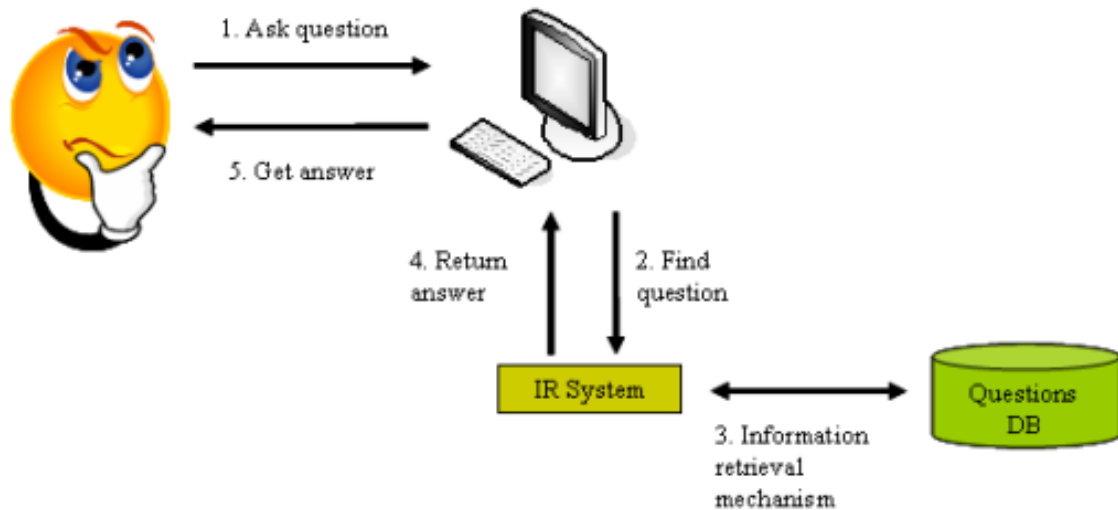


Figure 4.2: Basic idea of the question asking module

The ERM-Tutor was modified to incorporate the question-asking module. This module gives the students the opportunity to ask open-ended/free-form questions related to the domain and receive the most appropriate feedback. The modifications include the addition of the following:

- Questions Database
- Information Retrieval Mechanism
- Customisation of ERM-Tutor

In this section we present an overview of these additions. First we present the questions and answers database, followed by the implementation of the information retrieval mechanism, and lastly the *Questions* frame that allows the students to interact with the module.

4.3.1 Questions Database

We created a database of pre-defined questions that are used to respond to the students' submissions. We defined the questions based on our experiences with the mapping algorithm and other constraint-based tutors, as well as through consultations with an expert in teaching the domain and based on the results of a small think-aloud study.

Preliminary Think Aloud Study

A small think aloud study was conducted in May 2005 using the original version of the ERM-Tutor. There were three volunteer participants, all of whom were postgraduate students from the department of Computer Science and Software Engineering. The purpose of the study was to compile a set of questions that students may ask while using the ERM-Tutor. This was useful in building the question-answer database. The participants were asked to use the system for at least an hour while saying aloud what they were thinking as they interacted with the system and worked on the problems. Moreover, they were asked to say out loud any questions they were thinking of that would enhance their performance and clarify their understanding.

The results showed that the participants were very reluctant to ask questions. Rather they went through trial and error phases whenever they did not understand a concept. Overall the questions asked were classified as interface usage and terminology definition questions.

Question-Answer Pairs

The questions and answers are stored as text in a .def file, which we refer to as the database. There are 93 unique questions in the database. These can be categorised into

interface usage (e.g. What does the button Check Step do?), definitions of terms (e.g. *What is a foreign key?*), diagram notations (e.g. *How is an attribute represented in the ER diagram?*), mapping regulations (e.g. *How is a relationship mapped?*), and deeper questions (e.g. *Why are the steps arranged in this order?*). The database also includes a number of repeated questions but phrased differently. For example, the database stores “*What is an entity?*” as well as “*What does an entity mean?*”. On average, there are two versions of each question. This is necessary as the students’ questions are processed using a simple text retrieval mechanism, hence the system needs to cater for the various ways a question can be asked.

A typical question in the database consists of three parts; a unique integer representing the question’s id, question text in a string form, and answer text also in a string form. In contrast to ALPS, the answers to questions are textual. The following is an example of a question:

```
(1 ;;unique identifier
  "What is an Entity?" ;;question text
  "An Entity is a specific object or thing in the
  mini-world that is represented in the database.") ;;answer text
```

Listing 4.1: Example of a question

4.3.2 Information Retrieval Mechanism

The task of retrieving data from a user defined query has become common in recent years. This query retrieval can be loosely described as the task of searching a collection or corpus of data, be that text documents, databases, networks, etc., for specific instances of that data. Moreover, it is the task of searching for instances of that data that the query retrieval system considers relevant to what the user entered as the query.

The TFIDF (Term Frequency Inverse Document Frequency) vector weighting scheme [Salton and McGill, 1986] was chosen as the information retrieval mechanism, as is the case in ALPS. It is a scheme used to build an inverse index of words and then, based on the query calculations, retrieve the document with the highest weight. In other words, the document retrieved is the one that most accurately contains the query words. In this model, every word is a dimension and every document is a vector. It allocates a weight to each word dimension in a document vector, for example if D stands for *document* and w stands for *word* then $D_i = (w_1, w_2, \dots, w_3)$. Each word has only one weight and hence large documents accumulate more weight than smaller documents. The order of words in a document is not important, but repeated words gain more weight.

The first step in the algorithm is to calculate the frequency of each term, or word, used. The *term frequency* in the given document is simply the number of times a given term appears in that document. It is obtained by dividing n_i , the number of times the word appears in a document, by the number of total words in the document (4.1).

$$\text{tf}_i = \frac{n_i}{\sum_k n_k} \quad (4.1)$$

The *inverse document frequency* is then used to measure the general importance of the term. This is calculated by dividing the number of all documents by the number of documents containing the term, and then taking the logarithm of that quotient (4.2).

$$\text{idf}_i = \log \frac{|D|}{|\{d : t_i \ni d\}|} \quad (4.2)$$

Then the final step of the algorithm calculates the *term frequency inverse document frequency* (4.3).

$$\text{ifidf} = \text{tf} \cdot \text{idf} \quad (4.3)$$

In our system a document is equivalent to a question. Firstly, the questions are read from the database and separated into words. The weight of each question and word is calculated, and words are indexed in a hash table. When the student asks a question, or queries the system, the same calculations are applied to the query string: it is also broken-up into words and their weights are calculated. Each question is then allocated a query weight. Finally, the answer corresponding to the question with the highest query weight is returned to the student.

4.3.3 Customisation of ERM-Tutor

Figure 4.3 illustrates the updated architecture of ERM-Tutor with the newly added question-asking module. Initially the Pedagogical Module receives the students' question submissions from the Session Manager (corresponding to step 1 in Figure 4.2), and sends them to the Question Module where the information retrieval mechanism takes place (step 2). Upon receiving a question submission, the Question Module, applies the TFIDF calculations to it and based on the calculated weights, retrieves the most appropriate question-answer pair from the database (step 3) and returns it back the Pedagogical Module (step 4). The Pedagogical Module then sends the question-answer pair to the interface to be displayed to the student (step 5). Finally, the Pedagogical Module notifies the Student Modeller to update the student's model with the submitted question and the returned question-asking pair, as well as the student's rating for the system's response.

We added a new frame to the interface underneath the feedback panel, called "Questions", shown in Figure 4.4. This frame is the communication medium between the students and the question-asking module, where the students can submit their free-form questions and receive the system's answers. It includes a textarea for the students to

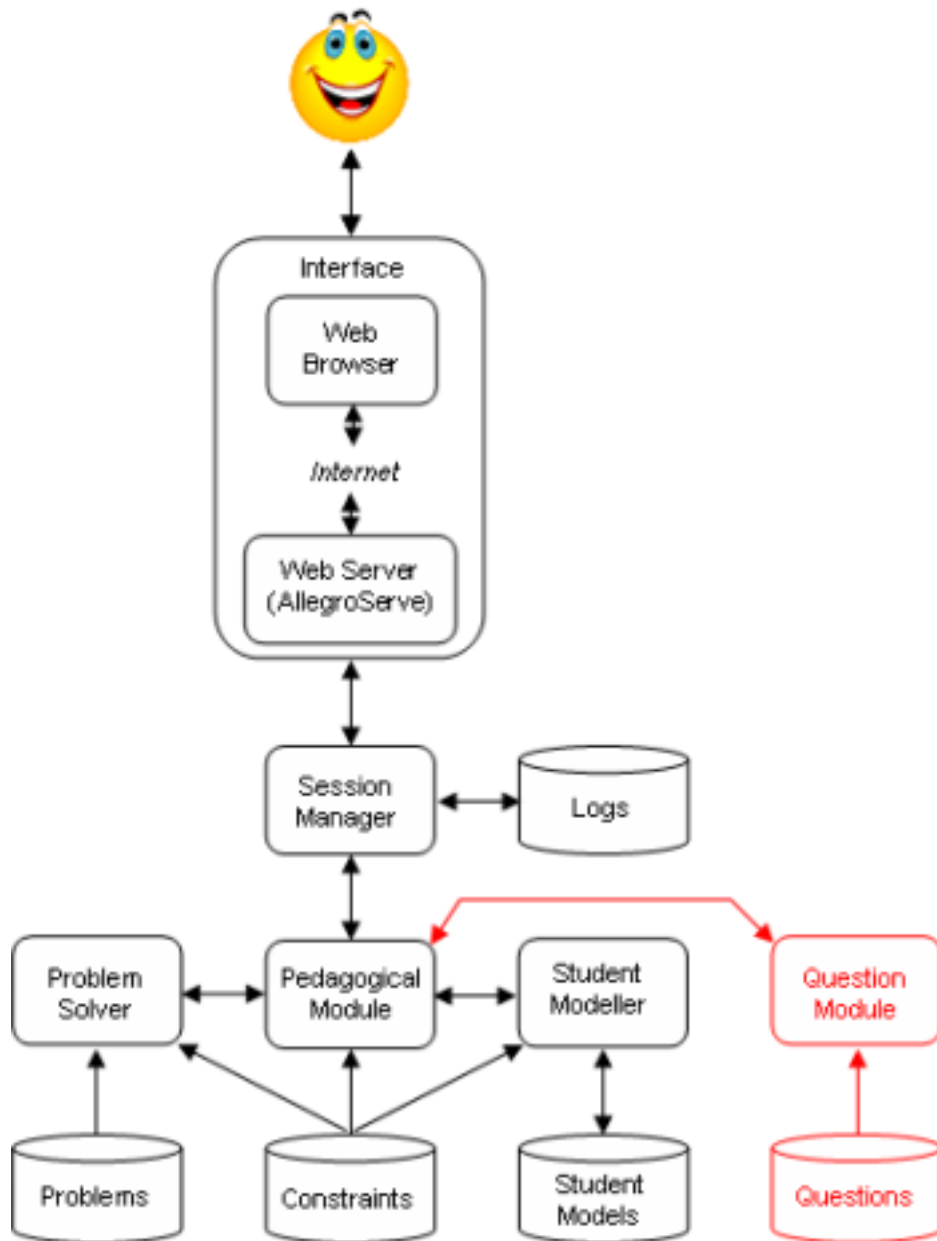


Figure 4.3: Architecture of ERM-Tutor with question-asking module

type in their questions; typing the questions will force the students to give some thought to formulating the questions, which is proposed to enhance deep learning. Once the student clicks on the Submit question button, the information retrieval mechanism is invoked and the query is processed. The returned question, including the question text as well as answer text, is then presented in the same frame, as shown in Figure 4.5. To evaluate the subjective relevance of each of the answers from the students, a select menu is presented along with the returned question with five rating levels. The students are encouraged to submit their ratings; however, the system does not enforce it to avoid mode errors and distractions from the problem solving task.

ERM-TUTOR

| Problem Text | | Completed Tables | | Change Problem | | Help | | Logout |

Step 1 Step 2 Step 3 Step 4 Step 5 Step 6 Step 7

Step: 1. Map all the regular entity types

Instructions: Choose the entity you want to map, then specify each attribute you want to add to the table you have created, use the checkboxes if the attribute is a key or foreign key.

Table attribute: Add attribute
 Key
 Foreign Key

Current table: COURSE Delete table name

| | |
|--------|--------|
| title | code |
| Edit | Edit |
| Delete | Delete |

Feedback: List All Errors Check table Clear

Feedback

1. Choose an attribute of the entity you are mapping to be a key for this relation.

Questions

Enter your question:

 Submit question

Figure 4.4: ERM-Tutor interface with Questions module

Questions

Enter your question:

Submit question

Q. What is an entity?

A. An entity is a specific object, thing, person or an event in the mini-world that is represented in the database. Entities can be thought of as nouns.

Please rate the tutor's response to your question:

very relevant
relevant
neutral
irrelevant
very irrelevant

Submit rating

input
textarea

question text

answer text

rating select
menu

Figure 4.5: Close up on the Questions module

CHAPTER 5

Spatial Ability

ITSs are effective learning tools due to the adaptive pedagogical assistance they provide. They make decisions about the timing and content of teaching actions and feedback to each student based on their individual state. Students differ in their strategies, approaches, and capabilities for learning and processing cognitive information. Although it is evident that such personal characteristics play a vital role in the learning process and in developing meta-cognitive skills, only a small number of studies have investigated the effects of accounting for them in ITSs. For example, Conati and Maclaren [2004] use the Five Factor personality traits (openness to experience, conscientiousness, extraversion, agreeableness and neuroticism) in representing different personality types and goal priority in a Dynamic Bayesian Network. This network is then used to maintain an assessment of the students current emotional state. In contrast, EDUCE [Kelly and Tangney, 2004] uses the Multiple Intelligence learning characteristics (logical/mathematical, verbal/linguistic, visual/spatial and musical/rhythmic) in order to provide a customised learning path.

In this chapter, we focus on *spatial ability*, a psychometric construct [Jensen, 1999] essential to activities related to spatial reasoning, such as the ability to manipulate im-

ages or spatial patterns into other arrangements [Carroll, 1993]. Learners with high spatial abilities perform better with graphic or spatially-oriented content than those with low spatial ability. It is worth noting, however, that a low spatial ability score is not a deficit; there is evidence that it can be improved through training and practice [Baeninger and Newcombe, 1989; Vicente and Williges, 1988]. Nevertheless, changing ITSs to accommodate low spatial ability learners could be more practical and beneficial for the system/domains problem solving task. That is, learners with different spatial abilities should receive different types of content.

This chapter presents an approach to support the learners' spatial ability in ERM-Tutor (described in detail in Chapter 3). We start by presenting the *Cognitive Theory of Multimedia Learning* in Section 5.1. Section 5.2 gives an overview of spatial ability and the tests used to measure it. Section 5.3 discusses the modifications made to ERM-Tutor to reflect our proposed solution.

5.1 Cognitive Theory of Multimedia Learning

Personal characteristics are a major factor in learning. Many theories exist regarding how individuals process and encode information differently, such as Richard Mayer's theory of multimedia learning [Mayer, 1997; Moreno and Mayer, 1999; Mayer, 2001; Mayer et al., 2003; Mayer and Moreno, 2003]. Mayer defines multimedia as the presentation of material using both verbal and pictorial forms, such as both words and pictures. He proposes that presenting verbal explanations alone in instructional situations is less conducive to learning for some students than presenting verbal explanations in conjunction with pictures [Mayer, 1997]. Subsequently, he defines a multimedia instructional message as communication that makes use of our dual learning channels [Baddeley, 1986; Paivio, 1986] which is intended to foster learning.

Subsequently, Mayer defined the *Cognitive Theory of Multimedia Learning* [Mayer, 2001]. There are three assumptions underlaying the theory. They are [see Mayer, 2001, chap. 3]:

1. Dual channels
2. Limited capacity
3. Active processing

The first assumption states that humans possess separate channels, referred to as the dual channels, for processing visual and auditory information. Figure 5.1 shows a representation of the dual channels. One channel is dedicated to processing visually represented material, for example printed words and pictorial forms, and the other is for processing auditory represented material, for example speech. Information enters the human information system via one of the channels depending on its representation type. The learners may then convert the representation, or form a corresponding representation in the other form, for processing in the other channel. For instance, when a learner hears a statement they initially process it via their auditory channel, but they may also form a corresponding mental image that is processed in the visual channel. Such cross-channel representations of the same information play a vital role in the dual-coding theory [Paivio, 1986].

According to the second assumption, humans are limited in the amount of information that they can process in each channel at one time. The learner is able to hold only a few items in working memory at any one time. This idea of limited capacity in consciousness has long been established in the field of cognitive psychology, and is used as the foundation for a number of theories, such as the *Theory of Working Memory* [Baddeley, 1986], the *Cognitive Load Theory* [Chandler and Sweller, 1991] and Miller's commonly known theory of *The Magical Number Seven, Plus or Minus Two* [Miller,

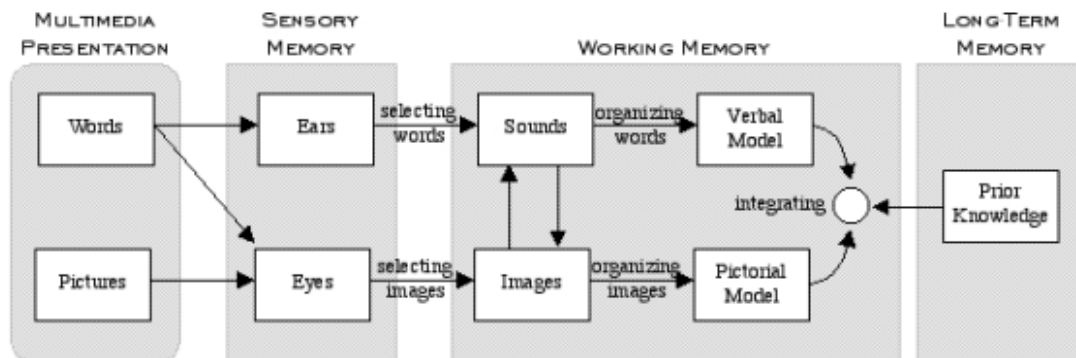


Figure 5.1: The dual learning channel, from presentation to sensory memory then working memory and finally to long-term memory [Mayer, 2001, Figure 3.2, p. 44]

1956]. There are a number of proposed tests to measure the cognitive capacity, such as the *Memory Span Test* [Miller, 1956]. Along with the limited cognitive capacity there are limited cognitive resources. These limitations force us to make decisions about which incoming information to pay attention to, the degree to which we process selected pieces of information, and the degree to which we connect processed pieces of information with our existing knowledge. Moreover, this assumption plays a central role in modern theories of intelligence [Sternberg, 1990].

The third assumption is that, in order to learn effectively, humans actively engage in attending to relevant incoming information, organising selected information and integrating mental representations with other knowledge. Active learning is therefore, viewed as a process of model building, which activates knowledge in long-term memory and bridges it into working memory. A mental model, or knowledge structure, is a representation of the key elements of the presented material and their relations.

Based on these three assumptions, the *Cognitive Theory of Multimedia Learning* states that learning occurs when learners attend to relevant incoming information (sensory memory), select and organise important information and integrate it with their prior knowledge (working memory) into mental representations (long-term memory). Mayer

argues that making use of both visual and auditory channels when presenting learning instructions aids in deep, or meaningful, learning, indicated by good retention and transfer performance. His rationale is that when presenting a message combining an image and text, as illustrated in Figure 5.2, the information is effectively being perceived and processed twice (once through each channel). Moreover, the words and pictures complement each other, aiding the learner to mentally encode and integrate the information.

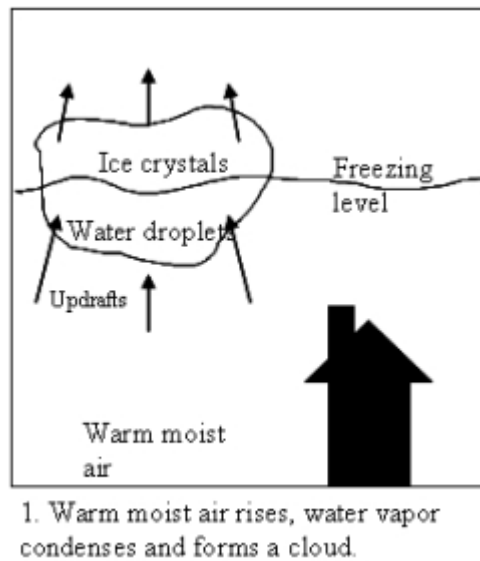


Figure 5.2: An annotated illustration [Mayer, 2001, Figure 2.1, p. 25]

It is evident, however, that learners differ in the way they process information and build their knowledge structures. For instance, as presented in Section 2.3, individuals have different cognitive styles and abilities. Some people learn better with visual methods of instruction, whereas others learn better with verbal methods of instruction. The question that rises is whether presenting the same instructional information is beneficial for both groups of people. Or does it overload the mental processing of some people or even confuse them? More importantly, if learners process information differently, then how can an instructional environment be tailored to better suit their individual needs? Is it actually beneficial to customise digital instructional environments?

These are also some of the questions that Mayer considered. As a result, he documented a number of principles for designers of instructional environments to follow in order to make the maximum use of the learners' dual channels. Table 5.1 illustrates the seven basic principles for the design of multimedia presentations. Mayer proposes that learners will benefit from incorporating these *good* design principles into multimedia presentations. For instance, the *Coherence Principle* states that students learn better when extraneous words, pictures and sounds are excluded rather than included, that is, presentations should be as clear and concise as possible to minimise mental processing overload.

- | |
|--|
| <ol style="list-style-type: none">1. <i>Multimedia Principle</i>: Students learn better from words and pictures than from words alone.2. <i>Spatial Contiguity Principle</i>: Students learn better when corresponding words and pictures are presented near rather than far from each other on the page or screen.3. <i>Temporal Contiguity Principle</i>: Students learn better when corresponding words and pictures are presented simultaneously rather than successively.4. <i>Coherence Principle</i>: Students learn better when extraneous words, pictures, and sounds are excluded rather than included.5. <i>Modality Principle</i>: Students learn better from animation and narration than from animation and on-screen text.6. <i>Redundancy Principle</i>: Students learn better from animation and narration than from animation, narration, and on-screen text.7. <i>Individual Differences Principle</i>: Design effects are stronger for low-knowledge learners than for high-knowledge learners and for high-spatial learners rather than for low-spatial learners. |
|--|

Table 5.1: Seven research-based principles for the design of multimedia messages [Mayer, 2001, Figure 11.1, p. 184]

The principle that is of most interest to us however, is the *Individual Differences Principle*, which states that “[multimedia] design effects are stronger for low-knowledge learners than for high-knowledge learners and for high spatial learners rather than for low spatial learners” [Mayer, 2001, p. 161]. Mayer proposes that

although processing information via the dual-channel is useful for all learners, other factors must be accounted for such as the level of domain knowledge. This principle suggests that it is particularly important to implement good multimedia design for not only high spatial but also low knowledge learners. This is because high-knowledge learners are able to use their prior knowledge to compensate for the cognitive processing needed to integrate the information received by the dual-channel. Moreover, they are able to create and use mental models on their own, without the additional benefits of well-designed multimedia presentations. In contrast, low-knowledge learners may need to have pictorial representations supplied to them and therefore are more likely to benefit from multimedia presentations. Additionally, low-spatial learners require so much mental energy to hold images in their working memory that they do not have enough capacity left over to mentally integrate the words and pictures, that is, they must devote so much cognitive capacity to mentally integrate the information. In contrast, high-spatial learners have sufficient mental energy to hold images and coordinate them with verbal representations. Therefore, it is the combination of the learners' spatial ability and level of knowledge that influences their meaningful/deep learning.

5.2 Spatial Ability

Spatial ability is a psychometric construct, or cognitive attribute, generally defined as the ability to generate, maintain, and manipulate mental visual spatial information, i.e. images [Carroll, 1993]. It is the ability to engage in spatial cognition that is important in multimedia learning and information processing [Mayer, 2001]. In accordance with the information processing through the dual learning channel concept, spatial cognition can be thought of having three main attributes; ability to encode spatial information from sensory memory, ability to maintain an internal representation of the information

in working memory, and ability to perform spatial transformations in order to integrate the information in long term memory.

Similar to other cognitive attributes, there has been an interest in finding a correlation between individuals' spatial ability level and their gender and age. Studies investigating such correlation, for example testing spatial memory and spatial navigation through a novel environment, showed a male advantage for spatial performance, suggesting that spatial ability is one of the most reliable of all cognitive gender differences in humans (e.g. [Moffat et al., 1998]), as well as an age related decline in performance (e.g. [Moffat et al., 2001]).

Other studies have demonstrated that the learners' spatial ability level and the type of content representation directly affect the learners' cognitive load, level of concentration and motivation. For instance, Steinke et al. [2004] investigated the presence of 3D models in a hypermedia learning system on plant and animal cell biology. They found that participants with a high spatial ability level spent more time on task relevant content than those with a low spatial ability level, whereas those with a low spatial ability level spent more time with the 3D models. Those with a low spatial ability level experience more difficulties in using 3D models and are more easily distracted from task relevant content. More interestingly, those with a high spatial ability level had a more positive attitude towards 3D content, thus confirming that a high subjective involvement results in a positive influence on the knowledge gain.

Cognitive load and level of prior knowledge have an influence on the learners' actions and overall experience. Vicente and Williges [1988] looked into learners' and their navigational behaviour. They found that individuals with a low spatial ability level generally have longer mean execution times and more first try errors than those with a high level. These studies also suggest the difficulties experienced by individuals with a

low level were specifically related to system navigation issues. In particular, low spatial ability users often report being “*lost*” within hierarchical menu systems.

5.2.1 Measuring Spatial Ability

There are a number of well established psychometric tests that are used in classifying students as high or low spatial ability learners (e.g. the kit of factor referenced cognitive tests [Ekstrom et al., 1997], the tube figures test [Stumpf and Fay, 1983], and the Purdue Spatial Visualisation Tests [Guay, 1977]). Most of these tests are paper-and-pencil tasks requiring inspecting, imagining or mentally transforming small shapes or manipulable objects at the figural scale of space [Hegarty et al., 2006]. These tests do not provide a discrete value on the spatial ability scale, but rather it is the relative position within the sample group that determines the high or low classifications. A median split is the method commonly used in classifying students as high or low spatial ability learners.

We have explored short versions of two classic tests for measuring spatial ability from the battery of cognitive tests developed by Ekstrom et al. [1997]: a ten-item *Paper Folding Test* intended to evaluate a component of spatial ability called *visualisation*, and an eighty-item mental *Card Rotation Test* intended to evaluate a component of spatial ability called *spatial orientation*. Each test had a three-minute time limit and is suitable for grades 9-16 (ages 13-18). The tests are explained below.

Paper Folding Test

The kit of factor-referenced cognitive tests defines visualisation as “the ability to manipulate or transform the image of spatial patterns into other arrangements.” [Ekstrom et al., 1997, p. 173] The Paper Folding Test is intended to measure the individual’s visualisation ability through their performance in a set of multichoice questions.

For each question in this test, the student assumes that we fold a sheet of paper one or more times, then punch one or more holes in it, and then unfold the paper back into its original form, as illustrated in Figure 5.3. The student then selects from a set of five possible answers the alternative that corresponds to how the punched sheet would appear when fully reopened. There is only one correct solution/choice for each question. Figure 5.4 shows an example of such a question, to which the correct answer is C. A copy of the test and its instructions are attached in Appendix D.

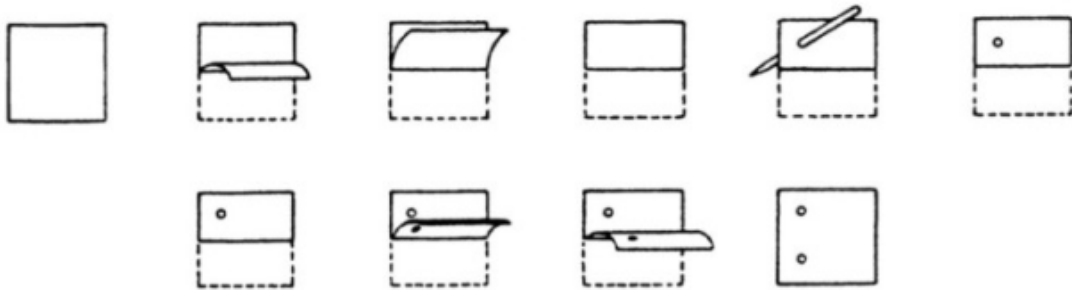


Figure 5.3: The process of folding, hole punching and unfolding a sheet of paper [Ekstrom et al., 1997, p. 176]

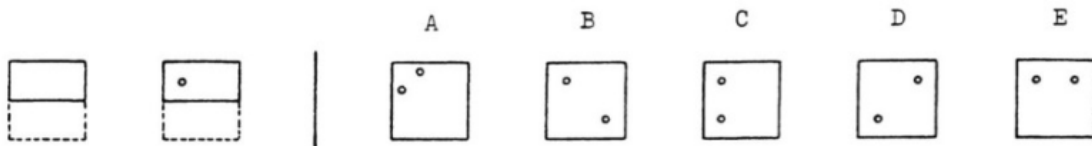


Figure 5.4: An example of a Paper Folding Test question [Ekstrom et al., 1997, p. 176]

Card Rotations Test

The Card Rotations Test is designed to measure the spatial orientation. This is defined in the same kit as “the ability to perceive spatial patterns or to maintain orientation with respect to objects in space.” [Ekstrom et al., 1997, p. 149]

In this test, each problem gives a drawing of a card cut into an irregular shape. To its right are eight other drawings of the same card, sometimes merely rotated, for example 5.5a, and sometimes turned/flipped over to its other side, for example 5.5b. The student indicates whether or not the card has been rotated by indicating whether it is the same (s) or different (d) to the original card. As the student needs to make a judgement about each of the eight drawings, each drawing is considered as a question, making eighty questions in total. Figure 5.6 shows a solved example. A copy of the test and its instructions are attached in Appendix D.

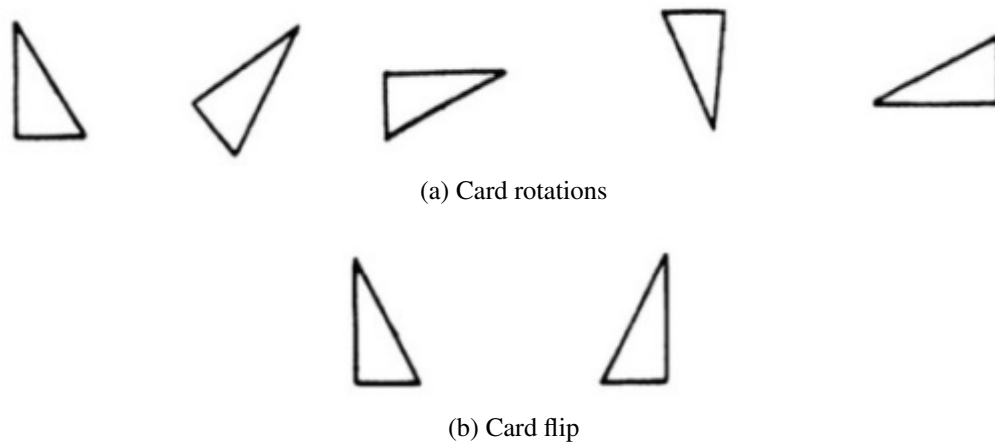


Figure 5.5: Example of the difference in a card being rotated and flipped [Ekstrom et al., 1997, p. 151]



Figure 5.6: An example of a Card Rotations Test question [Ekstrom et al., 1997, p. 151]

5.3 Our Spatial Ability Module

Influenced by Mayer's work, we decided to modify ERM-Tutor in order to cater for the learners' spatial ability and incorporate good multimedia presentations. We decided to create an alternative multimedia presentation of the system's feedback messages. Ideally, it would be beneficial to customise all the system's instructional messages, including the feedback messages as well as the question-asking module responses. However, due to time constraints it was decided that just the feedback messages will be sufficient for the scope of this project. This is because all students will be exposed to feedback messages when interacting with the system and submitting their solutions for evaluation, whereas only a subset of students chooses to use the question-asking module.

In this section we highlight the key modifications we made to ERM-Tutor in order to cater for the learners' spatial ability and incorporate good multimedia presentations. We also prepared ERM-Tutor for use in our evaluation studies in order to test the effectiveness of the multimedia messages and their impact on learning. The modifications can be categorised into:

- Feedback messages
- Customisation of ERM-Tutor

5.3.1 Feedback Messages

The original ERM-Tutor only provided text-based feedback messages in response to evaluating the student's solution. Based on the feedback level chosen (hint, explanation, list all errors, or full solution), the system informed the student whether their solution was correct or incorrect and provided hints to aid the student in solving their errors when appropriate. Following the multimedia theory, we decided to incorporate a pictorial aspect in the messages. For each explanation feedback message, we created a graphically annotated version, in accordance with the theory's good design principles.

The decision to use pictures and words for the multimedia messages as opposed to other presentation modes, such as speech or animation, was because of the following. Our evaluation studies were conducted in a controlled computer lab environment, and due to limitations in technical resources, having verbal speech was not possible. Furthermore, considering the nature of the domain, we felt animated messages will not give any additional benefit, that is, static arrows would convey the same information. Again, due to technical constraints in the lab environment and implementation purposes, the use of animated messages was discarded.

Since ERM-Tutor is a constraint-based tutor, each feedback message in the system is associated with one constraint. In other words, each constraint has a feedback message which is returned when the constraint is violated. Consequently, each message provides a hint on how to satisfy its particular constraint. As described in Section 3.2, each constraint in ERM-Tutor has two levels of feedback; a short hint and an explanation. For the purposes of our research and evaluation studies, we chose to create a multimedia representation of just the explanation messages, and we disabled the hint messages from ERM-Tutor for the duration of the studies.

To make the text-based (textual) and the newly created messages comparable, we kept the text identical in both versions/representations. The only difference is the addition of a pictorial representation of the text in the new version. Moreover, we only used *single letters* as labels of constructs shown in the pictorial representations, to prevent any additional help the context might give to students receiving the multimedia messages over the textual messages. Listing 5.1 shows an example of a constraint including its explanation textual feedback message (“*For this step you only need to specify the foreign keys from the owner entities.*”), and its corresponding multimedia representation is shown in Figure 5.7. This particular constraint checks whether the student solution includes two foreign keys, when a ternary identifying relationship is being mapped in Step 2.

```

(215
 2  "Check you have the correct number of foreign keys!"
 4  (and (equal (current-task SS) 'step2)
 6  (not (null (mapped IS)))
 8  (not (null (current-table SS)))
10  (bind ?t (current-table SS) bindings)
12  (match '(?*d1a "@" ?tt ?t "weak" ?*d2a)(entities SS) bindings)
14  (match '(?*d1 "@" ?rt ?r ?*d2) (relationships SS) bindings)
16  (match '(?*d3 "@" ?p1 ?c1 ?rt ?e1 ??role1
        ?*d4 "@" ?p2 ?c2 ?rt ?e2 ??role2
        ?*d5 "@" ?p3 ?c3 ?rt ?e3 ??role3 ?*d6) (connections SS) bindings
    )
    (not (null (current-fkey SS)))
    (not (< (length (current-fkey SS)) 2)))
(= (length (current-fkey SS)) 2)
"Step 2"
"For this step you only need to specify the foreign keys from the
  owner entities.")

```

Listing 5.1: Example of a constraint

A total of 125 JPG images were created in accordance to Mayer’s principles of good multimedia design. For example, we adhered to the *Spatial Contiguity Principle* by

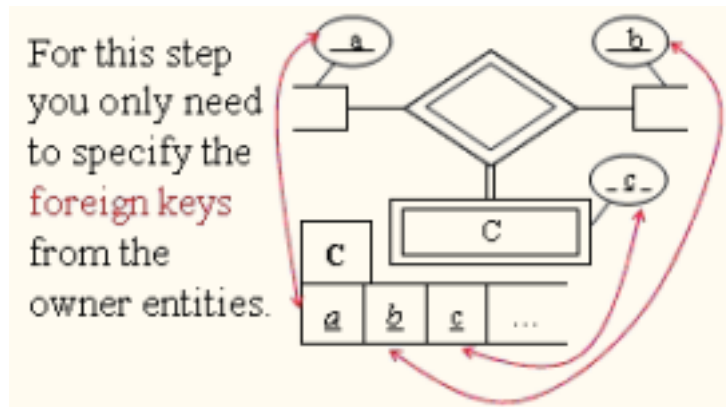


Figure 5.7: Example of a constraint in multimedia representation

presenting the text near its corresponding pictorial representation. Furthermore, each image corresponds to a single feedback message. ERM-Tutor was modified to cater for both versions of the messages and prepared for the evaluation studies. We outline the main changes to ERM-Tutor in the following subsection.

5.3.2 Customisation of ERM-Tutor

Figure 5.8 illustrates the updated architecture of ERM-Tutor with the newly added spatial ability module. The multimedia feedback messages are stored in a database as a set of JPG images. Each image has a unique name corresponding to its constraint, and is used in retrieving it for presentation to the student through the interface. The Spatial Module communicates with the Pedagogical Module to administer the spatial ability tests (the paper folding and card rotations tests) and evaluate the students' responses to the tests. The Pedagogical Module communicates with the Student Modeller to update the student models with the students' test scores, as well as the allocated feedback presentation mode (textual versus multimedia). Subsequently, with every student submission, during their problem-solving interactions with ERM-Tutor, the Pedagogical

Module consults the student model about the allocated feedback presentation mode, retrieves the appropriate mode and returns it to the interface to be presented to the student.

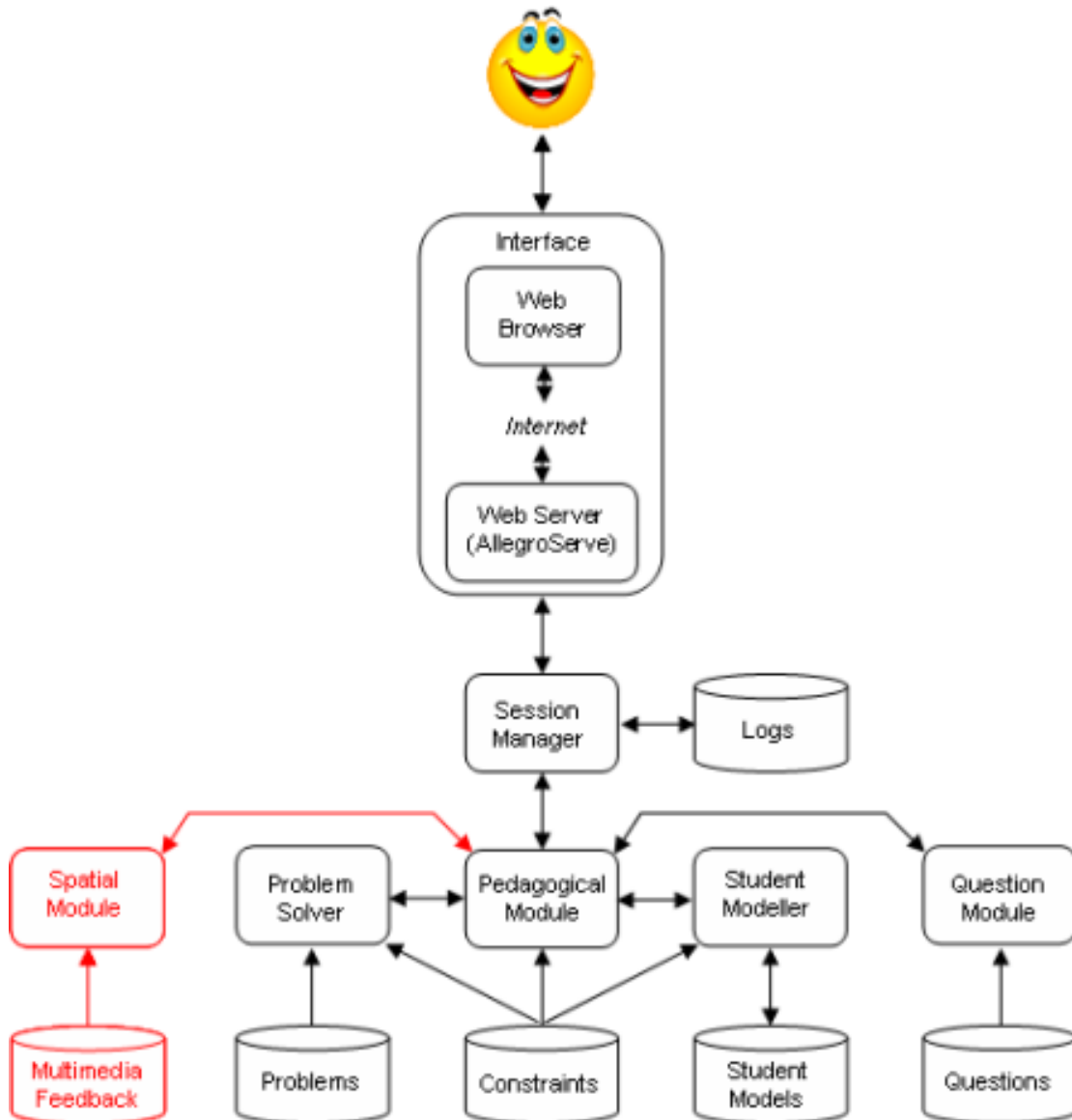


Figure 5.8: Architecture of ERM-Tutor with feedback presentation reasoning/module

It was necessary to automate the spatial ability tests in order to determine each student's spatial ability level and subsequently allocate the appropriate feedback presentation mode. We implemented both tests online and automated the marking/scoring pro-

cedure. Each test is administered through a set of interactive HTML webpages, where the student can read the instructions and solve the questions. Before sitting a test, the student is presented with a page of instructions explaining the purpose of the test and how to solve its problems, as well as a sample problem similar to Figures 5.4 and 5.6 (the sets of instructions and questions for both tests are attached in Appendix D). Additionally, for each test, the students were asked to rate their own ability for the particular skill being tests on a Likert scale of 1 (lowest) to 5 (highest) before sitting the test. This is used to compare the students perception of their ability to their *scored* level.

Each test has a time-limit of three minutes to solve its problems. We included a timer in each webpage administering the tests, as well as a Finished test button for exiting the tests before the time is up. Once the tests are completed, or the time is up, the webpage sends the student's responses back to the server for evaluation and the results are displayed to the student in a new webpage. Once both tests are completed and the student's feedback presentation mode is allocated, they are able to use ERM-Tutor in the usual manner, with the only difference being the feedback mode presented. Figure 5.9 shows a snapshot of ERM-Tutor serving a multimedia feedback message.

ERM-TUTOR

| Problem Text | | Completed Tables | | Change Problem | | Help | | Logout |

Step 1 **Step 2** Step 3 Step 4 Step 5 Step 6 Step 7

Step: 2. Map all the weak entity types

Instructions: Choose the entity you want to map, then specify each attribute you want to add to the table you have created, use the checkboxes if the attribute is a key or foreign key.

Table attribute: Add attribute
 Key
 Foreign Key

Current table: MEETING Delete table name

| | | |
|-----------|--------|-------------|
| <i>id</i> | timing | description |
| Edit | Edit | Edit |
| Delete | Delete | Delete |

Relationship: MEETS Delete relationship

Feedback: List All Errors Check table Clear

Feedback

For this step you need to specify all the foreign keys from the owner entities.

Questions

Enter your question:

Submit question

Figure 5.9: Snapshot of ERM-Tutor with a multimedia feedback message

CHAPTER 6

Evaluation

Evaluation is essential for accurately appraising hypotheses and proposed solutions. We investigated the implication of providing a question-asking module that allows students to ask free-form questions. We also looked into the potential of tailoring the feedback messages towards students' spatial ability. These goals were tested through two hypotheses. The first hypothesis is that answering students' open-ended questions will clarify their understanding and result in higher performance. The second hypothesis is that presenting the system's responses tailored to the students' spatial ability (textual vs multimedia) will lead to more effective learning and higher learning gain.

We enhanced ERM-Tutor, a constraint-based ITS that teaches logical database design (i.e. the algorithm for mapping conceptual to logical database schemas), by (a) implementing a question-asking module and (b) incorporating a multimedia representation of the system's feedback messages. We then conducted a series of evaluation studies to evaluate the effectiveness of the developed environment and to test our hypotheses. All our studies were conducted with tertiary students enrolled in an introductory database course (COSC 226)¹ offered by the department of Computer Science

¹<http://www.cosc.canterbury.ac.nz/open/teaching/>

and Software Engineering² at the University of Canterbury. This chapter outlines these evaluation studies. They are:

1. Preliminary Evaluation (2005)
2. Evaluation Study (2006)
3. Evaluation Study (2007)

The following section describes the preliminary evaluation (2005) and its results. Section 6.2 details the first evaluation study (2006) and Section 6.3 details the second evaluation study (2007). We also discuss the findings of all the studies in Section 6.4.

6.1 Preliminary Evaluation (2005)

We enhanced ERM-Tutor with a question-asking module where students are able to ask for additional clarifications by asking free-form questions. The system processes the submitted question and returns the answer with the highest relevance weight, using the TFIDF weighting scheme (presented in detail in Chapter 4).

In order to investigate the usage of the newly added question-asking module, a preliminary evaluation study of ERM-Tutor was carried out with students enrolled in an introductory database course (COSC 226) at the University of Canterbury in 2005. The results showed an indication of how students would respond to the proposed question-asking module and the modifications needed to enhance the module as well as minimise certain confounding factors for the evaluation study.

²<http://www.cosc.canterbury.ac.nz>

6.1.1 Procedure (2005)

The aim of the experiment was to investigate the system's effectiveness as well as the usage of free-form questions. Therefore, there was only one version of ERM-Tutor offered to the participants. This version included the question-asking module as well as instructions describing it and encouraging the participants to use it. As described in Section 4.3, the module also included a rating option.

There were 89 students enrolled in the course, who were invited to use ERM-Tutor. The students had attended lectures on ER mapping and had some practice during tutorials prior to the evaluation. They were free to use the system at any time from October 10, 2005, for as long as they wanted. Students sat a pre-test the first time they logged onto the system. The participants worked individually, solving problems at their own pace. The system recorded all student actions in logs. The students were given a post-test the first time they logged into ERM-Tutor on or after November 4, 2005; this date was decided based on the course examination date being November 5th, this is because from previous findings a large number of students log into the systems offered (e.g. SQL-Tutor, EER-Tutor) just before the course's examination date. The pre/post tests were used to evaluate the student's performance before and after using the system. There were two tests (A and B) of the same complexity, consisting of four multichoice questions each (copies of these tests are included in Appendix B, the version used in this study included only questions 1-4). About half the participants were randomly assigned version A as the pre-test and version B as the post-test. The remaining participants were assigned version B as the pre-test and version A as the post-test.

6.1.2 Results (2005)

A total of 29 students logged into ERM-Tutor at least once, but five students used it for less than two minutes and so their logs were excluded from analyses. The average interaction time was under one hour (mean=54min, sd=64min), ranging from several minutes to 4.5 hours over several weeks. The number of sessions ranged from one to four (mean=1.7, sd=1.0)³. On average, students attempted 4.6 problems and completed 25% of them. As the study was voluntary, only four students sat the post-test, and therefore we cannot compare the pre/post test results. Table 6.1 shows a summary of the statistics.

| | Mean (s.d.) |
|--------------------------------------|--------------------|
| Time spent on problem solving (min.) | 54:27 (63:59) |
| Number of sessions | 1.7 (1.0) |
| Number of attempted problems | 4.6 (4.7) |
| Completed problems (%) | 24.8% (21.9%) |
| Time spent per problem (min.) | 11:30 (7:30) |
| Pre-test (%) | 47.9 (29.4) |
| Post-test (%) | 50.0 (28.9) |

Table 6.1: Mean system interaction details (standard deviation) (2005)

Only eight students asked questions, with a total of 24 questions submitted. The number of questions per student ranged from one to five. The questions can be categorised into the following groups (also illustrated in Table 6.2): task-focused (50%), definition-focused (8%) and phatic⁴ questions (42%). Task-focused questions ask directly for help solving the problem (e.g. “How could I solve this table?”), and in most cases for resolving the errors identified in the submitted solution. For instance, three students copied the feedback messages, added a question mark at the end or a “How to”

³Numbers are rounded to one decimal place due to the low number of students.

⁴*Phatic*: *adj.* Of, relating to, or being speech used to share feelings or to establish a mood of sociability rather than to communicate information or ideas. (<http://www.thefreedictionary.com>)

at the start, and submitted them as the questions. Definition-focused questions ask for definition of terms. There were only two such questions submitted: “What is foreign key?” and “What is multivalued?” Phatic questions establish a sense of social mood. For example, questions included “What is your name?”, “How are you?” and “How do you answer questions?”, as well as some expressive statements such as “I can’t solve this?”

| Question Type | Example | Percentage |
|----------------------|--------------------------------------|-------------------|
| Task-focused | Help on directly solving the problem | 50% |
| Definition-focused | Definition of terms | 8% |
| Phatic | Establish a social mood | 42% |

Table 6.2: Percentage of question types asked (2005)

There were 14 questions (excluding phatic questions) that were relevant for students’ actions. Five of these questions were answered correctly, and for two of these, the students specified highest relevance. The answer could not be found for one question. The remaining questions received answers which were related to the query, but were not useful to students. This happened when the students did not formulate questions well, but instead copied a part of the feedback message, adding a question mark at the end (e.g. “Make sure the relationship is 1:1?”).

The log files also showed a number of situations that were not being addressed by the constraints set at the time, which have been attended to for the following study.

These results were published as a short paper (a camera ready copy is included in Appendix E) and presented at the 8th International Conference on Intelligent Tutoring Systems held in Jhongli, Taiwan, June 2006 [Milik et al., 2006].

6.1.3 Summary (2005)

We enhanced ERM-Tutor with a question-asking module, which allows the student to ask free-form questions, which the system processes and returns the answer with the highest relevance weight, using the TFIDF weighting scheme. We conducted a preliminary study to investigate the students' reactions toward the question-asking module and subsequently the module's effectiveness. Our preliminary study showed some evidence that students welcome the idea of asking free-form questions and are willing to use it.

The results show an indication of the types of questions the students are inclined to ask. We classified half of the questions submitted as task-focused, which were requesting help on directly solving the current problem the students were working on. The rest of the questions were either definition-focused, acquiring about the meaning of domain terms/keywords, or phatic questions, making *small talk* with the system to establish a social mood or express their perception of the task.

The results of this study confirmed the need for eliciting deeper questions as well as various techniques to encourage students to use the question-asking module. Moreover, we added all the questions that the students asked and were not found to our questions database, which will improve the effectiveness of the question-asking module.

6.2 Evaluation Study (2006)

We modified ERM-Tutor to provide not only textual feedback messages, but also messages containing combinations of text and pictures, we refer to them as multimedia messages, in accordance with the multimedia theory of learning [Mayer, 2001]. To test the effects of customising the learning environment in accordance to the student's spatial ability, we preformed an evaluation study with students enrolled in the same in-

troductory database course, COSC 226, at the University of Canterbury in March 2006. Our hypothesis is that matching the presentation mode (textual vs multimedia) towards the students' spatial ability will lead to more effective learning and higher learning gain. In particular, students with a high spatial ability level will benefit more from multimedia feedback than students with a low spatial ability, given the same background knowledge.

A total of 74 students were enrolled in the course. As each student's spatial ability level (low or high - as opposed to the actual value) is determined relatively to the sample group, it was decided to compute it after the experiment was conducted in a post-hoc manner. The students were randomly allocated to one presentation mode of the system's feedback, providing either textual or multimedia feedback. That is, about half of the students were only given the feedback messages in textual form, whereas the remaining half of students were given the feedback messages in multimedia form. Other than the feedback presentation mode, the system was identical and behaved in the exact same manner towards all participants. The assumption is that each group will ultimately include students with low and high spatial abilities. Therefore, the experiment allows for a 2x2 comparison as shown in Table 6.3: textual messages for low (LT) and high spatial ability students (HT), and multimedia messages for low (LM) and high spatial ability students (HM). Furthermore, all students had access to the question-asking module and were encouraged to use it as much as possible.

| Feedback Messages | Spatial Ability | |
|-------------------|-----------------|------|
| | Low | High |
| Textual | LT | HT |
| Multimedia | LM | HM |

Table 6.3: Evaluation study experimental design (2006)

After gaining an approval from the Human Ethics Committee at the University of Canterbury to run the experiment, the evaluation study was conducted on the 27th and

28th of March, 2006. In contrast to the preliminary study, this study was conducted during the course's scheduled tutorials on ER mapping, straight after students attended lectures on the topic. There were two 2-hour lab streams; we refer to them as sessions. The students sat the study during their regular lab time. Again, the participants worked individually, solving problems at their own pace. The system also recorded all student actions in logs.

6.2.1 Procedure (2006)

At the start of each session, the students were given an information sheet describing the study, a consent form, and a paper-based pre-test consisting of four multichoice questions. To make the results of the pre-test comparable, similar to the preliminary study, two tests were used; students in the first session were given version A as the pre-test and students in the second session were given version B as the pre-test (the version of the tests used in this study is the same as the preliminary study; questions 1-4 of the tests in Appendix B).

The first time a student logged onto the system, they sat the spatial ability tests. Participants were presented with a set of instructions explaining the spatial ability tests used, and a sample problem similar to Figures 5.4 and 5.6 (more details about the spatial tests are presented in Chapter 5). Additionally, for each test, they were asked to rate their own ability for the particular skill being tested on a scale of 1 (lowest) to 5 (highest) before sitting the test. They had three minutes to solve the problems in each test. Once the tests were completed, or their time was up, the students were randomly assigned to one of the two feedback presentation modes. They were asked to use the system, solving as many problems as they would like, while making use of the question-asking module. At the end of each session, students were asked to fill in a post-test and a questionnaire

about the system. The questionnaire included questions on a Likert scale with five responses ranging from 1 to 5, as well as questions requiring free-form responses. The questionnaire is included in Appendix C. Lastly, the students were encouraged to use the system at any time until the end of the course.

6.2.2 Results (2006)

A total of 57 students logged into the system. However, due to an internal fault in the system, there were a number of problems in the first session which prevented many students from using the system. Nevertheless, their pretest results were computed and stored. The system was fixed for the second session, in which a total of 25 students logged onto the system.

Spatial Ability (2006)

55 students from both sessions completed both spatial tests; the paper fold and card rotation tests. Table 6.4 shows a summary of the test scores for the 55 students. Before completing each test, the students were asked to rate their own ability of the spatial skill they are being tested for. For the paper fold skill, the students gave themselves a mean rating of 6.6 ($sd = 2.0, median = 6$). This was close to the actual test score which had a mean of 6.9 ($sd = 2.0, median = 7$) out of a possible 10, with correlation coefficient of 0.51. The students' personal rating for the card rotation skill had a mean of 7.6 ($sd = 2.2, median = 8$). As explained in Chapter 5, the total possible score for the card rotation test is 80. We computed the total card rotation test score by dividing the score by 8 giving a range of 1-10, and the students scored a mean of 6.4 ($sd = 2.0, median = 6.3$), with correlation coefficient of 0.06. To compute the spatial ability of each student, we added both test scores giving a possible range of 1-20. Using a median split, we

classified the students as either low or high spatial. A total of 28 students scored above the median and were classified as high spatial, and the other 27 students were classified as low spatial.

| | Paper Fold | | Card Rotation | | Both Tests |
|--------|-------------------|------------|----------------------|------------|-------------------|
| | Rating | Test Score | Rating | Test Score | Test Score |
| Mean | 6.6 | 6.9 | 7.6 | 6.4 | 13.3 |
| S.D. | 2.0 | 2.0 | 2.2 | 2.0 | 3.4 |
| Median | 6.0 | 7.0 | 8.0 | 6.3 | 13.1 |

Table 6.4: Spatial ability test scores (2006)

The number of students allocated to each of our four groups is presented in Table 6.5. As mentioned above, due to a technical problem, the logs from the first session could not be used. Therefore, Table 6.5 shows two figures for each group; the first number reported for each condition shows the total number of students allocated to it, while the number in brackets shows the number of valid logs for the same group.

| Feedback Messages | Spatial Ability | | Total |
|--------------------------|------------------------|------------|----------------|
| | Low | High | |
| Textual | LT: 15 (7) | HT: 13 (5) | 28 (12) |
| Multimedia | LM: 12 (7) | HM: 15 (6) | 27 (13) |
| Total | 27 (14) | 28 (11) | 55 (25) |

Table 6.5: Participants assignment to groups (2006)

Pre and Post Tests (2006)

The pre and post tests were used to evaluate the student's performance before and after using the system. The pre-test was collected at the start of the session, while the post-test was administered after two hours of interaction with the system. Both tests contained four multichoice questions, with a possible score of 0-4, and were completed on pa-

per. Only 13 students from the second session completed both tests, scoring a mean of 1.9 ($sd = 1.0$) on the pre-test and 3 ($sd = 1.2$) on the post-test, as shown in Table 6.6, resulting in significant improvement of their performance ($t = 3.1, p < 0.001$). As mentioned above, the numbers of participants in each of the four groups (LT, HT, LM and HM) were too small for comparing the performance on the tests between the different groups. Nevertheless, the numbers are presented in Table 6.7. The *No.* columns show the number of students who have submitted both tests in that particular group and the other numbers show the mean scores for the tests and their standard deviations between brackets.

| | Pre-Test | Post-Test |
|------|-----------------|------------------|
| Mean | 1.9 | 3.0 |
| S.D. | 1.0 | 1.2 |

Table 6.6: Pre-test and post-test scores (2006)

| Feedback | Low Spatial | | | High Spatial | | |
|-----------------|--------------------|-----------|-----------|---------------------|-----------|-----------|
| | No. | Pre-test | Post-test | No. | Pre-test | Post-test |
| Textual | LT: 4 | 1.5 (1) | 3.5 (0.6) | HT: 3 | 2 (1) | 2.3 (0.6) |
| Multimedia | LM: 2 | 1.5 (0.7) | 3.5 (0.7) | HM: 4 | 2.5 (1.3) | 2.8 (1.9) |

Table 6.7: Mean (*sd*) pre-test and post-test scores for groups (2006)

We have also calculated the correlation between the pre-test scores and spatial ability of students to be 0.4. The correlation between the post-test scores and spatial ability of students was 0.3. These numbers are small suggesting that there is little correlation between the students' spatial ability level and how well they scored on the pre- and post-test.

These preliminary results (although with small numbers) seem to refute Mayer's prediction that high spatial learners will benefit most from multimedia messages. However, it does seem that the subsets of participants from the HT and HM groups who

completed both tests started with higher pre-existing knowledge, and therefore Mayer's individual differences principle may be more pertinent in that low knowledge individuals will have a higher gain. Of course, with such low numbers of submitted tests, we might expect a lot of error and therefore further investigation is warranted.

System Interactions (2006)

Out of those who logged onto and used ERM-Tutor, 17 students used it for more than 10 minutes, and only 13 students completed both the pre and post tests. On average, students attempted 3.4 problems and completed 33% of them. Only six students logged into the system after the study, four of which logged in a day before the course's exam. The numbers of valid logs in each group (LT, HT, LM, and HM) are too small, and we are therefore unable to closely analyse the effect of the students' spatial ability on their performance. For this reason, we decided that it was necessary to run another evaluation study at the start of the following year (March 2007) with COSC 226, with the hope that the system will be enhanced, allowing more students to use it for longer time. The results from the 2006 evaluation study was used as a guideline for the following study.

Question-Asking Module (2006)

13 students asked a total of 32 questions. As this evaluation study was carried out during a normal lab session with all the *human tutors* being there, many students asked questions through the conventional way. Many students felt more comfortable asking the human tutors for clarifications than the system.

Most of the questions were task-focused (e.g. *how to map a multivalued attribute?*). Some questions were asking about the interface usage (e.g. *how do I add a second table?*). It was promising to see some good questions (e.g. *Why does a multivalued*

attribute have two key values?). We enhanced our questions database to make sure it caters for all the questions asked.

It would also be interesting to find out the students' motivation behind using the question-asking module as opposed to asking the human tutor. Whether it is just experimenting with the tool, not wanting to admit not knowing to a tutor, wanting an easy/-faster way of finding the answer, filling in time while the tutor is attending to another student, or something else.

Subjective Results (2006)

As predicted, some of the comments from the first session were not very supportive due to the system problems in that session. Nevertheless, there were some encouraging comments from the second session. Some students liked the idea of the tutor and indicated that they would use it when it becomes available.

Analyses of the questionnaires showed that students who received multimedia feedback rated the overall quality of the feedback messages 25% higher (mean of 4 out of a possible 5) than those who received textual feedback (mean of 3 out of a possible 5).

Overall the students from the second session gave positive comments towards the system. Students appreciated the problem solving environment. Some of the comments were "*quicker than using pen and paper*", "*very good practice...*", "*I liked the step by step process*".

The results of this evaluation study were published and presented as a full paper at the 5th New Zealand Computer Science Research Student Conference held in Hamilton, New Zealand, April 2007 [Milik et al., 2007a], and as a short paper at the 13th International Conference on Artificial Intelligence in Education held in Los Angeles, United States, July 2007 [Milik et al., 2007b] (camera ready copies are included in Appendices G and F).

6.2.3 Summary (2006)

We looked at the potential of accounting for spatial ability in ERM-Tutor. The study evaluated the effectiveness of the type of feedback representation, whether textual only or multimedia, to the learner's spatial ability level. We hypothesised that students with a high spatial ability will benefit more from multimedia feedback than students with a low spatial ability.

The results presented show an overall improvement in the students' domain knowledge level after interacting with ERM-Tutor for the duration of the study (2 hours). We could not however report any findings on the correlation between spatial ability, content representation and the learning experience due to a technical problem. Although the amount of data collected was small, the results show a promising indication for further explorations. We therefore made the decision to use this study as the basis for another evaluation study testing the same hypothesis.

6.3 Evaluation Study (2007)

We performed our final evaluation study with students enrolled in the same course, COSC 226, at the University of Canterbury in 2007. 78 Students were invited to use ERM-Tutor during their scheduled tutorials on ER mapping, straight after they had attended lectures on the topic. There were two 2-hour lab streams. The study was conducted on the 28th and 29th of March, 2007, in a similar manner to the previous year's study. We used the same experimental design from the 2006 study with the difference of allocating equal number of students to each of our four groups (as shown in Table 6.3).

6.3.1 Variations in contrast to the 2006 Study

There are a number of variations between the 2006 and the 2007 evaluation studies.

Two significant changes are outlined here, they are:

- Allocation of Students to Groups
- Pre- and Post- Tests

Allocation of Students to Groups

One of the issues that we faced in the 2006 evaluation was the number of students in each of our four groups; textual messages for high and low spatial ability students and multimedia messages for high and low spatial ability students, as shown in Table 6.3. As the spatial ability is computed in reference to the students' median score, in the 2006 study the students were randomly assigned to their feedback group. The number of students in each of our four groups was then calculated in a post-hoc manner. To avoid having an unbalanced number of students in the 2007 evaluation study, we planned on allocating equal number of students to each of our four groups, with the hope that it will guarantee comparable groups.

In order to allocate students to the four groups, we need to know their spatial ability level, whether low or high, which is computed in reference to their median score. As the study was conducted during the scheduled lab sessions for the course, it was difficult to collect all the spatial ability scores and compute the median, either at the start of the sessions or at a different time, prior to giving students access to ERM-Tutor. We therefore, made the assumption that the two samples of participants (the 2006 and 2007 participants) are comparable, and used the median obtained in the previous study (13.1) as the threshold to classify students as having high/low spatial abilities. Based on the

students' classification, they were then spread evenly between the text and multimedia feedback groups.

Another experimental design decision that we were faced with was accounting for the prior domain knowledge level of students in each group. The students prior knowledge of the domain is assessed through the pre-test they sit before using the system. The effectiveness of ERM-Tutor is then inferred through examining the students' learning gain by comparing the pre- and post- tests scores after interacting with the system. As with the spatial ability tests, it was not possible to administer the pre-test prior to the lab sessions. Moreover, it was not feasible to balance the pre-test mean score as well as the number of participants in each of the four groups in real time, because of the variations in the students' levels. We therefore made the decision to only balance the number of students based on their spatial ability.

ERM-Tutor was modified to keep count of how many students were in each of the four groups. Based on each student's spatial ability (low vs high) and the number of students in their spatial ability feedback groups, they can be allocated to the appropriate feedback mode (text vs multimedia). The hope is that this will result in a comparable number of students in each of the groups. If it became apparent however, that the spatial test scores are not comparable between the two student populations, then some shifting of students between the two vertical columns (low vs high spatial) would take place without affecting the experiment. This is because we are comparing the performance of those students who were given a presentation mode according to their spatial ability (referred to as *matched*) and those who were not (referred to as *unmatched*).

Pre- and Post- Tests

As we mentioned above, we have published a short paper at the International Conference on Artificial Intelligence in Education describing the 2006 study. One of the comments

we received from the conference reviewers was regarding the pre- and post- tests having a maximum score of 4 marks. The reviewer was concerned that a 4-point scale might not be large enough to adequately capture differences between the scores. We have therefore added a mapping question to both tests with a score of 4 marks, giving a maximum score for the tests of 8 marks each.

The newly added questions, shown in Figure 6.1, present an ER diagram similar to the diagrams used as problems in ERM-Tutor, with the difference of using letters as labels as opposed to context describing words. The students are required to map the diagram into relational schemas as they would would using the ERM-Tutor. As mentioned above, the new question is worth 4 marks, giving the tests an 8-point scale.

Map the following ER diagram into its appropriate relational schemas

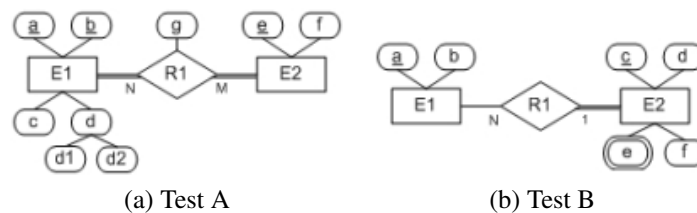


Figure 6.1: The newly added question to tests A and B

6.3.2 Procedure (2007)

At the start of each session, the students were given an information sheet describing the study, a consent form, and a pre-test on paper (with a maximum score of eight marks) consisting of four multichoice questions and a mapping question (see Appendix B, the version used in this study included all questions, 1-5). Similar to our previous studies, two tests were used; students in the first session were given version A as the pre-test and version B as the post-test, while students in the second session were given the reverse.

The first time a student logged onto the system, they were presented with a set of instructions explaining the spatial ability tests used, with sample problems. Additionally, for each test, students were asked to rate their own ability on a scale of 1 to 5 before sitting the tests. They had three minutes to solve the problems in each test. In contrast to the 2006 evaluation study, upon completing the spatial ability tests, or if the time was up, the students were *allocated* to the appropriate feedback presentation mode as explained above. Again, other than the feedback presentation mode, the ERM-Tutor was identical and behaved in the exact same manner towards all participants. This allows for the same 2×2 comparison used in the previous study as shown in Table 6.3, with the difference this time of distributing the number of participants across the different groups.

The participants were then asked to use the system, solving as many problems as they would like, while making use of the question-asking module. As in the previous studies, the participants worked individually, solving problems at their own pace while the system recorded all their actions in logs. At the end of each session, students were asked to fill in a post-test and a questionnaire about the system on paper. The questionnaire was also the same version used previously (see Appendix C).

6.3.3 Results (2007)

A total of 50 students logged into ERM-Tutor (23 students in the first session and 27 in the second session), of which 43 students completed both the pre-test and post-test (19 from the first session and 24 from the second session) and 45 students completed the questionnaire. 23 students used the question-asking module. The analyses of the collected data are presented in the following subsections. We start by looking at the spatial ability scores, followed by an analysis of the pre- and post-test scores. The log

files have been distilled to produce the data for the interactions with the system. We also present the mastery of constraints analysis, as well as an analysis of the interactions with the question-asking module and the questionnaire's subjective results.

Spatial Ability (2007)

Table 6.8 shows a summary of the test scores for the 50 students who logged into ERM-Tutor. As with the previous study, the students were asked to rate their own ability of the spatial skill they are being tested for before completing each test. For the paper fold skill, the students gave themselves a mean rating of 6.2 ($sd = 1.7, median = 6$). The actual test score had a mean of 5.6 ($sd = 2.3, median = 6$) out of a possible 10. The students' personal rating for the card rotation skill had a mean of 7.6 ($sd = 1.2, median = 8$). The students scored a mean of 5.6 ($sd = 2.2, median = 5.7$) out of a possible 10 in the actual card rotation test. To compute the spatial ability of each student, we added both test scores giving a possible range of 1-20. The spatial ability score in this study had a mean of 11.2 ($sd = 3.4$) and a median of 11.81.

| | Paper Fold | | Card Rotation | | Both Tests |
|--------|------------|------------|---------------|------------|------------|
| | Rating | Test Score | Rating | Test Score | Test Score |
| Mean | 6.2 | 5.6 | 7.6 | 5.6 | 11.2 |
| S.D. | 1.7 | 2.3 | 1.2 | 2.2 | 3.4 |
| Median | 6.0 | 6.0 | 8.0 | 5.7 | 11.8 |

Table 6.8: Spatial ability test scores (2007)

This study's spatial ability median of 11.8 is lower than the previous study's of 13.1. We therefore shifted some participants who were marked as low spatial to high spatial. Table 6.9a shows the number of participants in each of the four groups using the original median, whereas Table 6.9b shows the division based on this year's median. As shown

in Table 6.9, only the total count of the spatial columns have changed due to the shift of participants.

| Feedback Messages | Spatial Ability | | Total |
|-------------------|-----------------|-------|-----------|
| | Low | High | |
| Textual | LT: 17 | HT: 8 | 25 |
| Multimedia | LM: 17 | HM: 8 | 25 |
| Total | 34 | 16 | 50 |

(a) Based on 2006 spatial ability median

| Feedback Messages | Spatial Ability | | Total |
|-------------------|-----------------|--------|-----------|
| | Low | High | |
| Textual | LT: 11 | HT: 14 | 25 |
| Multimedia | LM: 14 | HM: 11 | 25 |
| Total | 25 | 25 | 50 |

(b) Based on 2007 spatial ability median

Table 6.9: Participants assignment to groups (2007)

The rest of this section presents the analysis of the collected data in light of the four groups of students presented in Table 6.9b. To further examine the data we categorised the students based on three additional levels, illustrated in Figure 6.2. They are:

Textual vs Multimedia Firstly, to test whether the type of the feedback presented, textual versus multimedia, had an effect on the results, we merged the groups of students into the two horizontal (rows) groups. The *Textual* classification is for all those who were presented with the textual feedback (LT and HT), whereas the *Multimedia* classification covers all the students presented with the multimedia feedback (LM and HM) (shown in Figure 6.2a).

Low vs High Secondly, as shown in Figure 6.2b, in some analyses we compared the students just on the basis of their spatial ability, that is, we compared the students across the two vertical (columns) groups. This is to test whether or not students

who are low spatial (LT and LM) (*Low*) differ to those who are high spatial (HT and HM) (*High*).

Matched vs Unmatched Thirdly, to examine the effects of presenting the feedback type in accordance to the student's spatial ability, we further grouped the students who are low spatial and were presented with textual feedback (LT), with the students who are high spatial and were presented with multimedia feedback (HM) into the *Matched* classification. Subsequently, we grouped the students who are low spatial and were presented with multimedia feedback (LM), with the students who are high spatial and were presented with textual feedback (HT) into the *Unmatched* classification (shown in Figure 6.2c).

| Feedback Messages | Spatial Ability | | |
|-------------------|-----------------|------|---|
| | Low | High | |
| Textual | LT | HT | — Text — Multimedia |
| Multimedia | LM | HM | |

(a) Text vs Multimedia feedback type

| Feedback Messages | Spatial Ability | | |
|-------------------|-----------------|------|--|
| | Low | High | |
| Textual | LT | HT | — Low — High |
| Multimedia | LM | HM | |

(b) Low vs High spatial ability

| Feedback Messages | Spatial Ability | | |
|-------------------|-----------------|------|---|
| | Low | High | |
| Textual | LT | HT | — Matched — Unmatched |
| Multimedia | LM | HM | |

(c) Matched vs Unmatched spatial ability with feedback type

Figure 6.2: Groupings of participants

Pre and Post Tests (2007)

A total of 43 students submitted both pre and post tests. As shown in Table 6.10, the mean score for all students on the pre-test was 4.3 ($sd = 2.2$) and on the post-test was 5.2 ($sd = 2.2$), out of a possible score of 8. We performed a *paired t-test* analysis to evaluate the students' performance before and after using the system. The analysis indicated that there is a statistically significant difference between the means of the pre-test and post-test scores ($t_{43} = 3.5, p < 0.001$). This suggests that all students significantly improved in performance on the post-test compared to the pre-test after interacting with ERM-Tutor for the 2-hours duration of the study, indicating that all students *learned* from ERM-Tutor. Moreover, a correlation analysis shows that the pre-test is a significant factor for the post-test performance with a correlation of 0.7.

| | Pre-Test | Post-Test |
|------|----------|-----------|
| Mean | 4.3 | 5.2 |
| S.D. | 2.2 | 2.2 |

Table 6.10: Pre-test and post-test scores (2007)

Table 6.11 shows the mean and standard deviation scores for the pre- and post- tests for each of our four groups. Again, the *No.* columns show the number of students who have submitted both tests in that particular group and the standard deviations are shown between brackets. The data is also illustrated in Figure 6.3.

| Feedback | Low Spatial | | | High Spatial | | |
|------------|-------------|-----------|-----------|--------------|-----------|-----------|
| | No. | Pre-test | Post-test | No. | Pre-test | Post-test |
| Textual | LT: 8 | 5.8 (1.1) | 5.9 (2.5) | HT: 12 | 4.2 (2.4) | 5.7 (1.8) |
| Multimedia | LM: 12 | 3.6 (2.1) | 4.5 (2.1) | HM: 11 | 4.2 (2.3) | 5.1 (2.5) |

Table 6.11: Mean (sd) pre-test and post-test scores for groups (2007)

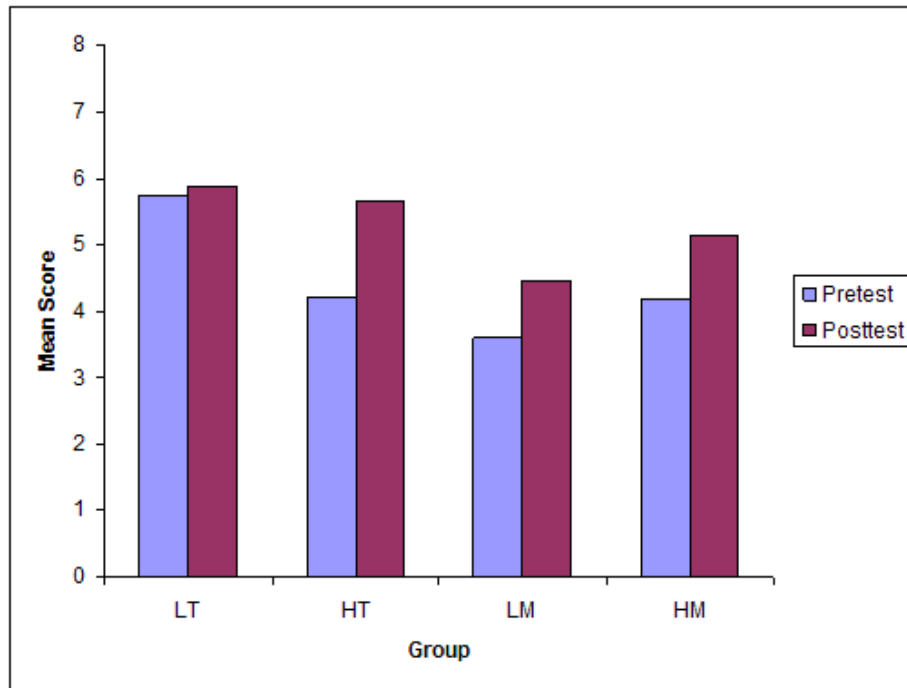


Figure 6.3: Pre- and post-test scores for the four groups

To analyse the difference in performance between the pre- and post-test for each of our four groups, we used a *paired two sample for means t-test*, shown in Table 6.12. The analysis indicated that there was statistically significant difference between the students' performance in the pre- and post- tests by those in the HT ($t_{12} = -3.4, p < 0.005$), LM ($t_{12} = -2.0, p < 0.05$) and HM ($t_{11} = -1.8, p = 0.0553$) groups. However, there was no significant difference for the LT ($t_8 = -0.2, p = 0.4365$) group. A closer look at the LT group shows that its students have a higher pre-test score, with a mean of 5.8 ($sd = 1.1$), and hence they improved the least in comparison with the other groups, scoring means of HT: 4.2 (2.4), LM: 3.6 (2.1) and HM: 4.2 (2.3). As we mentioned above, although we hoped for an ideal setting of comparable groups, this imbalance in prior knowledge between the four groups was unavoidable.

| Classification | t-test: Paired Two Sample for Means | | | | |
|----------------|-------------------------------------|-----------|-----------|--------|---------|
| | No. | Pre-test | Post-test | t-Stat | P-value |
| LT | 8 | 5.8 (1.1) | 5.9 (2.5) | -0.2 | 0.4365 |
| HT | 12 | 4.2 (2.4) | 5.7 (1.8) | -3.4 | 0.0029 |
| LM | 12 | 3.6 (2.1) | 4.5 (2.1) | -2.0 | 0.0385 |
| HM | 11 | 4.2 (2.3) | 5.1 (2.5) | -1.8 | 0.0553 |
| Textual | 20 | 4.8 (2.2) | 5.8 (2.0) | -2.2 | 0.0185 |
| Multimedia | 23 | 3.9 (2.2) | 4.9 (2.2) | -2.7 | 0.0070 |
| Low | 20 | 4.5 (2.1) | 5.0 (2.2) | -1.4 | 0.0840 |
| High | 23 | 4.2 (2.3) | 5.4 (2.2) | -3.6 | 0.0008 |
| Matched | 19 | 4.8 (2.2) | 5.4 (2.5) | -1.4 | 0.0954 |
| Unmatched | 23 | 3.9 (2.2) | 5.1 (2.0) | -3.8 | 0.0005 |

Table 6.12: Mean (*sd*) pre-test and post-test scores for all classifications (2007)

To examine whether there is a significant difference between the post-test across the four groups, we first checked to make sure that the groups were not significantly different at the pre-test mean scores, by performing a one-way ANOVA between-groups analysis. The result does not indicate a significant difference between the pre-test mean scores across the four groups ($F_{3,39} = 1.8, p = 0.1722$). Subsequently, we performed an ANOVA analysis across the post-test mean scores of the four groups, however it did not yield any significant difference either ($F_{3,39} = 0.9, p = 0.4618$). The result of this ANOVA analysis indicates that all four groups of students improved in a similar manner regardless of the feedback mode presented nor their spatial ability, suggesting that the different presentation modes have similar influence on performance regardless of the spatial ability; that is, all students, whether low or high spatial, improved in performance regardless of the feedback mode they were given. We suspect however, that although the ANOVA analysis on the pre-test did not indicate a significant difference, the higher pre-test in the LT group has an influence on these statistical tests.

A key component of the evaluation study focused on whether or not the students in the matched groups did significantly better than the students in the unmatched groups.

Since the ANOVA analysis did not address the difference in pre-test performance between the four groups, we decided to use another between-groups statistic, the one-way ANCOVA⁵ analysis. This analysis neutralises the effect of a continuous independent variable in the experiment, the pre-test in our case, and is therefore valuable in this situation. We used the post-test scores as the dependent variable, the group as the classifier and the pre-test as the covariate. Similar to the ANOVA results however, the results of the ANCOVA, shown in Table 6.13, indicate that when controlling for the pre-test scores, there was no significant difference among post-test scores of the four groups ($F_{3,38} = 0.6, p = 0.6020$).

| Source | Type III SS | df | Mean Square | F | Sig. |
|-----------------|-------------|----|-------------|-------|----------|
| Corrected Model | 101.89 | 4 | 25.47 | 9.51 | 1.98E-05 |
| Intercept | 31.95 | 1 | 31.95 | 11.93 | 0.001 |
| PreTest | 89.04 | 1 | 89.04 | 33.24 | 1.19E-06 |
| Group | 5.04 | 3 | 1.68 | 0.63 | 0.60 |
| Error | 101.78 | 38 | 2.68 | | |
| Total | 1381 | 43 | | | |
| Corrected Total | 203.67 | 42 | | | |

Table 6.13: One-way ANCOVA for testing of between-subjects effects of post-test scores (2007)

Table 6.12 also shows the *paired two sample for means t-test* for the other groupings. The tests show either a statistically or marginally significant difference in students performance between the pre- and post- tests scores for all groupings. The p-values produced show that unmatched groupings scored a higher significant confidence than the matched groupings. A closer look at the figures show that the matched groupings had a higher pre-test mean score of 4.8 ($sd = 2.2$) than the unmatched groupings

⁵ANCOVA, or analysis of covariance is a general linear model with one continuous explanatory variable and one or more factors. ANCOVA is a merger of ANOVA and regression for continuous variables. ANCOVA tests whether certain factors have an effect after removing the variance for which quantitative predictors (covariates) account. The inclusion of covariates can increase statistical power because it accounts for some of the variability. (<http://en.wikipedia.org/wiki/ANCOVA>)

($mean = 3.9, sd = 2.2$). This difference was verified as marginally significant using a *two-sample assuming unequal variances t-test* ($t_{19,24} = 1.5, p = 0.0759$). A further test within the matched groupings, comparing the pre-test scores between the LT and HM groups indicated a significant difference ($t_{8,11} = 2.0, p < 0.05$).

We also examined the tests gains across all the different groupings. The students' gains are calculated by subtracting the pre-test scores from the post-test scores. The tests did not produce a statistically significant difference, indicating that all students gained comparable amount of knowledge from interacting with ERM-Tutor.

Furthermore, we examined the correlation between the students' spatial ability and their performance on the pre- and post-test. The correlation between their spatial ability and pre-test score was 0.2, which is lower than the previous year's of 0.4. The correlation between their spatial ability and post-test score was 0.2, this is close to the previous year's correlation of 0.3.

We also computed the effect size and power to determine the effects and validity of the experiment and its results. The effect size examines the total population variance due to the experimental treatment. It is used to compare the results of the control versus the experimental conditions. This is calculated by subtracting the control's condition learning gain mean score from that of the experimental's, followed by dividing the result by the standard deviation of the gain scores of the control condition [Bloom, 1984]. In our case, we used the unmatched groupings as the control condition and the matched as the experimental, and we calculated the learning gain as the difference between the pre- and post- tests scores. The calculation produced $(0.6 - 1.2)/1.5 = -0.4$, indicating a negative relationship, that is, on average the unmatched group will do better than the matched group. This low effect size can be attributed to the low number of students and short duration of the study.

Power, or sensitivity, is used to measure how easily the experiment can detect differences. Power is measured as the fraction of experiments that for the same design, the same number of participants and the same effect size would produce a given significance. The calculation produced a power of 0.2 at a significance of 0.05. This is quite low compared to the recommended power of 0.8 [Chin, 2001]. Again, this could be due to the low number of students in each of the groups.

Interaction with the System (2007)

A total of 50 students logged onto and used ERM-Tutor, all of whom used it for more than 10 minutes, with a mean of 66 minutes ($sd = 31.7$). On average, students attempted 7.3 problems and completed 64% of them. These figures are higher than the previous study's and this can be attributed to the improvements made in ERM-Tutor that made it more appealing for students to use.

Table 6.14 presents a summary of the mean and standard deviation of the total time spent interacting with the system, number of attempted problems, number and percentage of solved problems and total number of attempts/solutions submitted for students in each of the four groups.

| | LT | HT | LM | HM |
|--------------------|--------------|--------------|-------------|--------------|
| Total time | 62 (27.3) | 72.7 (37.3) | 57.7 (29.8) | 72.2 (31.8) |
| Attempted problems | 6.9 (6.3) | 8.8 (5.0) | 4.8 (3.0) | 9.1 (5.0) |
| Solved problems | 4.4 (5.0) | 6.5 (4.6) | 3.2 (3.8) | 7.2 (4.5) |
| % solved problems | 61.2 (36.8) | 69.4 (26.1) | 53.4 (43.2) | 72.9 (21.3) |
| Total attempts | 118.4 (97.7) | 130.1 (73.2) | 69.7 (46.6) | 148.5 (90.7) |

Table 6.14: Summary of means (sd) of system interaction results (2007)

Figure 6.4 shows the number of attempted problems and the number of solved problems for the four groups. Students in the HM group attempted the highest number of

problems, with a mean of 9.1 ($sd = 5.0$) and solved 72.9% of them, whereas students in the LM group attempted the least number of problems, with a mean of 4.8 ($sd = 3.0$) and solved 53.4% of them. Although an ANOVA analysis did not show a significant difference across the four groups, there is a significant difference between the HM and LM groups ($t_{11,14} = 2.5, p = 0.01$). Moreover, ANOVA analyses for the other interaction data did not yield any significant results, however there was a marginally significant difference for the number of total attempts ($F_{3,46} = 2.5, p = 0.0740$) as well as number of problems solved ($F_{3,46} = 2.3, p = 0.0948$).

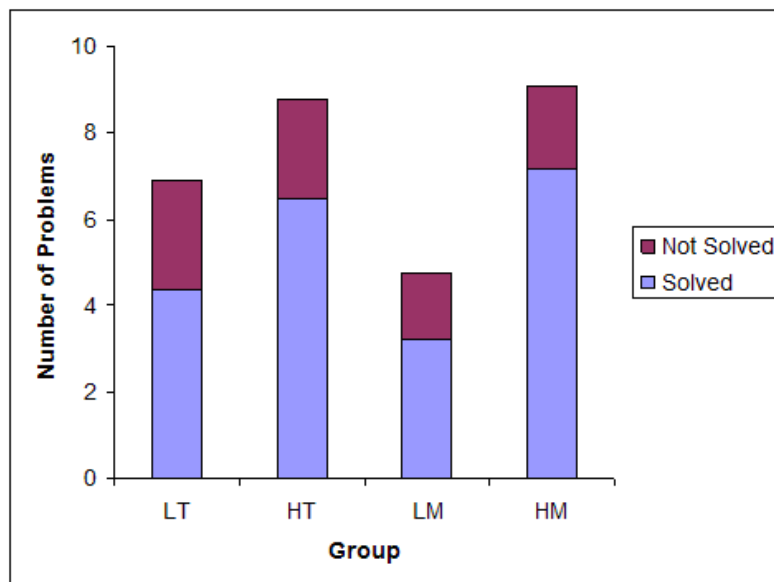


Figure 6.4: Number of attempted problems by the four groups

Mastery of Constraints (2007)

ERM-Tutor is a constraint-based ITS, that is its domain knowledge is represented as a set of constraints. Each constraint represents a piece/unit of domain knowledge, in other words it addresses one concept of the domain knowledge. A common way of measuring how well the domain concepts were learnt by students is to plot a learning

curve [Martin et al., 2005; Martin and Mitrović, 2005]. In our case, the learning curve is plotted using the probability of violating a constraint against its number of occurrences. The data points are then approximated using a power curve. A closely fitted, smooth power curve with a decreasing trend indicates a good learning rate.

We evaluated the constraints histories inside the student models to see whether or not students learnt the domain concepts. For each student, we calculated the probability of violating each constraint on the first occasion of being relevant, then the second occasion and so on. These probabilities are then averaged across all the constraints in order to obtain an approximate probability of violating a given constraint on a given occasion. Finally, the resulting probabilities are then averaged across all students and plotted as a function of the number of occurrences when a constraint was relevant.

Figure 6.5 illustrates the probability of violating a constraint plotted against the number of its occurrences (i.e. the number of times it was relevant), averaged over all students. The regular decrease in the plotted data points indicates that the probability of violating the same constraint decreases as the number of its occurrences increases, that is as students are more exposed to it. A power curve is used to approximate the data points with an equation of $y = 0.1839x^{-0.2506}$ and a good fitness of $R^2 = 0.79$.

Figure 6.6 shows the learning curves for the four different groups. All four sets of data points are approximated using a power curve, indicating that all four groups learned the domain concepts from interacting with ERM-Tutor. As shown by the slope, the LM group had the highest learning rate with an equation of $y = 0.1934x^{-0.3423}$ and fitness of $R^2 = 0.79$, indicating that the students with low spatial ability who received feedback in multimedia mode had the highest learning rate than the other groups.

Similar to the pre- and post- tests analyses, we looked at the learning curves with respect to the different groupings of students, illustrated in Figure 6.7. When we compared the students based on the feedback mode they received, the learning rate for

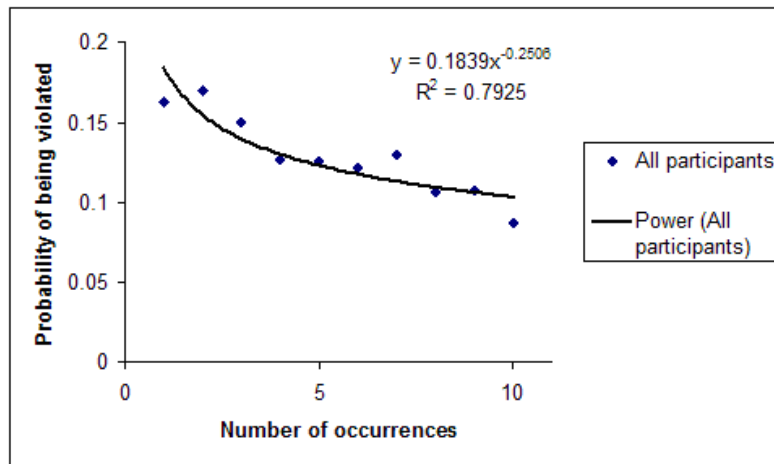


Figure 6.5: Learning curve for all participants

those who received the multimedia feedback is slightly higher, with an equation of $y = 0.1975x^{-0.2759}$ and fitness of $R^2 = 0.75$, than those who received the textual feedback ($y = 0.172x^{-0.2281}$, $R^2 = 0.76$), as shown in Figure 6.7a. This suggests that, although not statistically significant, the multimedia feedback messages helped students learn the domain concepts covered by the constraints slightly better than the textual feedback messages.

There was no noticeable difference in the learning rate between students with low ($y = 0.1899x^{-0.2744}$, $R^2 = 0.80$) and high ($y = 0.1788x^{-0.2318}$, $R^2 = 0.67$) spatial abilities (shown in Figure 6.7b). Similarly, the learning rate for the matched groups ($y = 0.1949x^{-0.2238}$, $R^2 = 0.74$) was close to the unmatched groups ($y = 0.1756x^{-0.2849}$, $R^2 = 0.75$) (shown in Figure 6.7c). It is worth noting however, that the matched groups had a consistently higher probability of violating a constraint than the unmatched groups.

Moreover, we examined the number of constraints learnt by students with respect to the four groups. For each student, we calculated whether a constraint was learnt or not by inspecting the ratio of the constraint being stratified over its last five occurrences. As illustrated in Figure 6.8, the HM group learned the most number of constraints

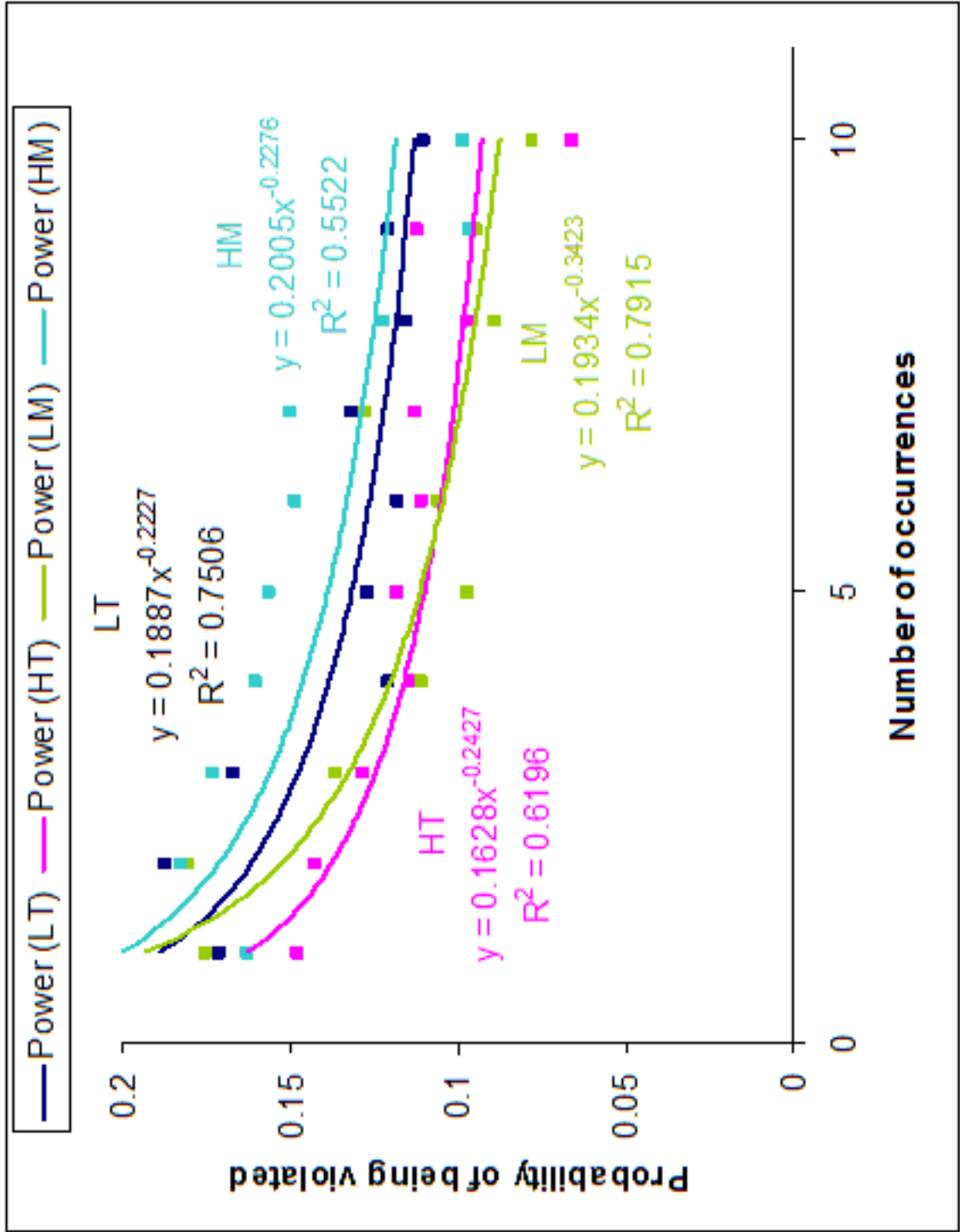
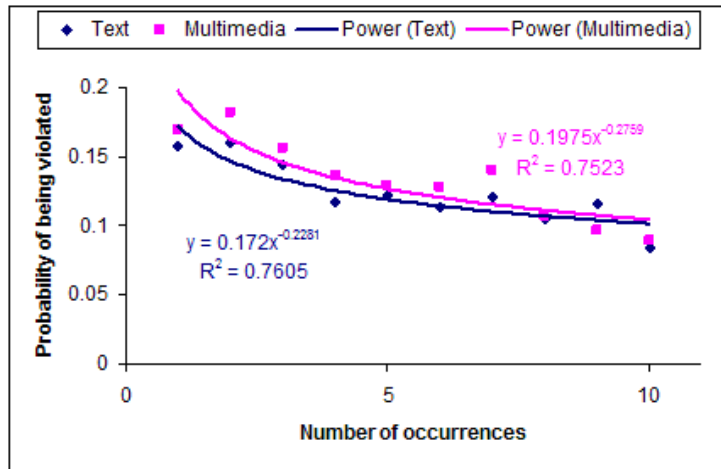
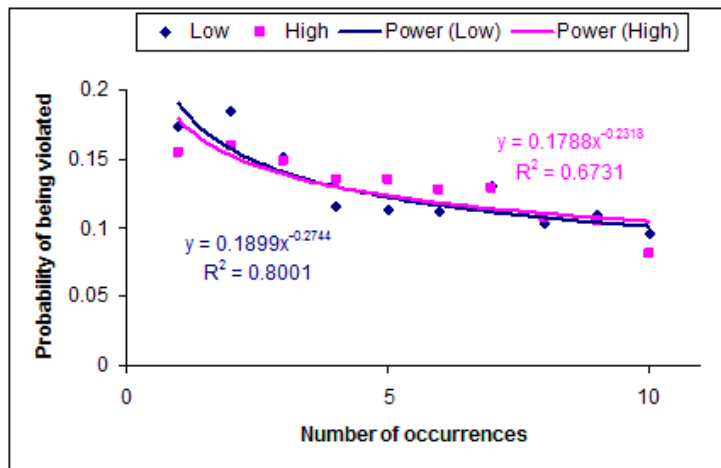


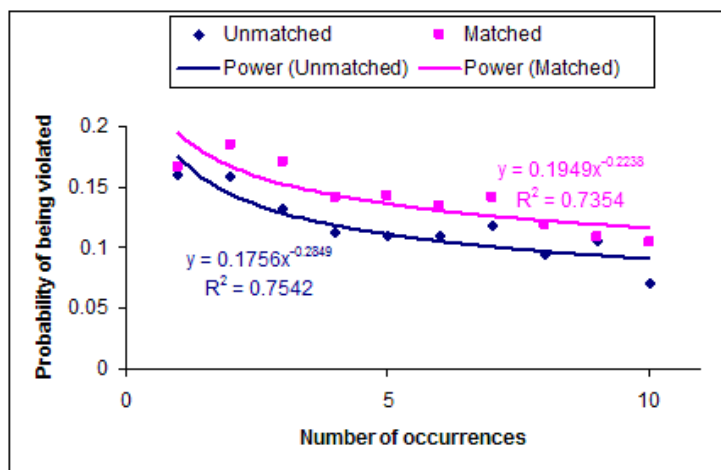
Figure 6.6: Learning curves for the four groups



(a) Text vs Multimedia feedback type



(b) Low vs High spatial ability



(c) Unmatched vs Matched spatial ability with feedback type

Figure 6.7: Learning curves for the different groupings of participants

($mean = 5.1, sd = 3.3$), followed by the LT group ($mean = 4.7, sd = 4.9$) and the HT group ($mean = 3.7, sd = 2.8$). The LM group learned the least number of constraints ($mean = 3.4, sd = 2.7$). The difference in number of constraints learned, although not statistically significant (ANOVA: $F_{3,46} = 0.7, p = 0.5772$), is inline with Mayer's findings with respect to the cognitive theory of multimedia learning. More specifically, the students with high spatial ability learned more domain concepts when the feedback they received was in multimedia form, while the students who have low spatial ability learned the least when presented with multimedia feedback. This suggests that the multimedia feedback was not as effective for students with low as with high spatial ability. Moreover, these findings are consistent with the findings from the number of attempted and solved problems for the groups presented above.

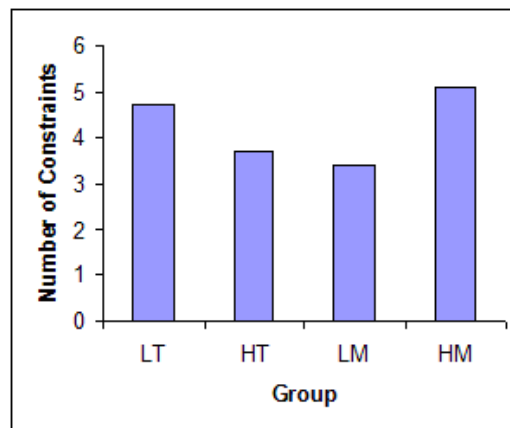


Figure 6.8: Number of constraints learnt for the four groups

Question-Asking Module (2007)

A total of 23 students used the question-asking module with a total of 71 questions submitted. On average the number of questions asked per student was 3.1 ($sd = 2.6$), ranging from one to ten questions per student. Similar to the previous studies, we categorised

the questions into the three categories. Table 6.15 reports the percentages of question in each category: task-focused (73%), definition-focused (17%) and phatic questions (10%). Therefore, 90% of the submissions (in the task-focused and definition-focused categories) were relevant for the current problem, and 78% of these questions received relevant answers.

| Question Type | Number of Submissions | Percentage |
|----------------------|------------------------------|-------------------|
| Task-focused | 52 | 73% |
| Definition-focused | 12 | 17% |
| Phatic | 7 | 10% |

Table 6.15: Percentage of question types asked (2007)

As we mentioned earlier, the task-focused questions are oriented towards solving the current problem. That is, these questions request help on directly solving the current difficulty/error that the student is faced with. There were 52 such questions asked, making up 73% of the total questions submitted. A closer inspection of the *wording* and *structure* of these questions lead us to further divide them into the following two sub-classifications; *problem-specific* and *task-specific*.

Problem-specific questions included *nouns* that were directly taken from the ER problem diagram. In other words, the question asked could not be applied/asked in another situation, that is when faced with the same error in another problem. Examples of these questions include “*Why Colour isn’t an attribute?*” where *colour* is the name of a multivalued attribute in the given ER diagram, “*Are there any attributes for the table LIVE_IN?*” where *LIVE_IN* is the name of a binary 1:N relationship, and even just the noun with a question mark such as “*car?*” There were nine such questions submitted. These questions indicate that the students were looking for the correct answer rather than an understanding of the constraint or domain concept that was violated. Moreover, these questions require low cognitive effort to formulate them and do not indicate that

the students have conceptualised the ‘big-picture’ of the domain knowledge. Since our database of questions does not include problem specific nouns, the relevance of the returned answers was low. For instance, the question “*Why Colour isn’t an attribute?*” was matched to and received the answer for “What is an attribute?” This is a limitation in our module; our module does not consult the student’s model when responding to their submissions. Ideally, it would be effective to examine the current problem and step the student is working on, matching any problem-specific nouns used in their submission, in order to provide a good response.

On the other hand, task-specific questions are independent of the specific problem the student is working on, i.e. they are related to the problem-solving task only. Ideally these would be worded in general terms using keywords from the domain, indicating that the students had given some thought to generating such questions, and would include a combination of both low-level-cognitive (shallow) and high-level-cognitive (deep) questions. Examples of the submitted questions include, “*How can we deal with multivalued attributes?*” “*Can one table have two attributes with the same name?*” “*How many keys can there be in one table?*” and “*What is the order of mapping?*” Nevertheless, we also observed the same behaviour as in the previous studies of submitting the system’s feedback messages as the question, with the addition of either “*Why is it saying*” or “*How to*” as well as a question mark. There were five such questions submitted, including a submission where the student had copied the current step’s instruction from the interface. The answers returned to those five questions were relevant to the domain keywords used. For example, the question “*Why is it saying Specify the identifying relationship that relates the weak entity to its owner entity?*” received the response for “*What is an identifying (weak) Relationship?*”, which defines the *identifying relationship* keyword.

There were twelve definition-focused questions. Submissions included well structured questions such as “*What’s an entity*” and “*What is a foreign key?*” as well as just keywords from the domain, either with or without the question mark, such as “*identifying relationship*” “*many to many?*” and “*weak entity*”. Furthermore, four students used the question-asking module in making phatic questions and expressions. Examples include “*hello*” and “*it didn’t let me put in what I wanted*”. These submissions are consistent with the questions submitted in the previous studies, confirming the need for addressing them in ERM-Tutor.

We performed a correlation test between the total number of questions asked by students and their learning gain, indicated by their post-test score minus their pre-test score, for the 23 students who used the question-asking module. The correlation test yielded a value of -0.3, indicating a negative relationship between the number of questions asked and the learning gain. However, we can not make any conclusions based on this statistic as the number, as well as quality of questions asked were lower than expected.

Again, we suspect that the question-asking module was not fully utilised by the participants in this study as it was carried out during their normal lab sessions with all the *human tutors* being there. A closer examination of the submitted questions seems to indicate that students were trying out the module or filling in time while the human tutors attend to their questions.

Subjective Results (2007)

All students were asked to complete a questionnaire at the end of their session interacting with ERM-Tutor. A total of 45 students completed the questionnaire, and their data were collated to determine their perception of ERM-Tutor. Table 6.16 shows the mean and standard deviation of the responses to the 1 to 5 Likert scale questions in respect

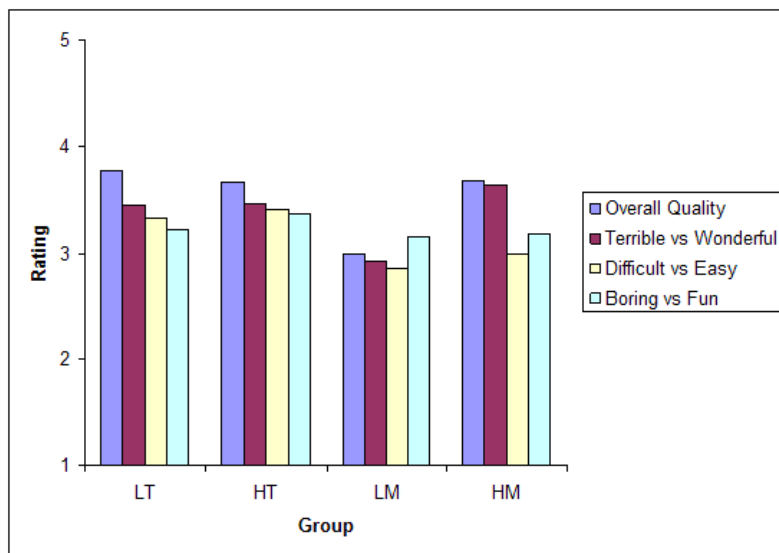
to our four groups, where 1 represents the most negative response and 5 represents the most positive response.

| | LT | HT | LM | HM |
|--------------------|-----------|-----------|-----------|-----------|
| Overall quality | 3.8 (0.7) | 3.7 (0.7) | 3.0 (1.1) | 3.7 (0.8) |
| Terrible-Wonderful | 3.4 (0.9) | 3.5 (0.9) | 2.9 (0.8) | 3.6 (0.7) |
| Difficult-Easy | 3.3 (1.0) | 3.4 (0.8) | 2.8 (1.1) | 3.0 (0.8) |
| Boring-Fun | 3.2 (1.0) | 3.4 (0.9) | 3.3 (0.8) | 3.2 (0.8) |
| Feedback messages | 3.4 (0.7) | 3.4 (1.2) | 3.1 (1.1) | 3.2 (0.9) |
| Question-asking | 3.4 (0.8) | 3.3 (0.8) | 3.5 (0.5) | 3.1 (0.7) |

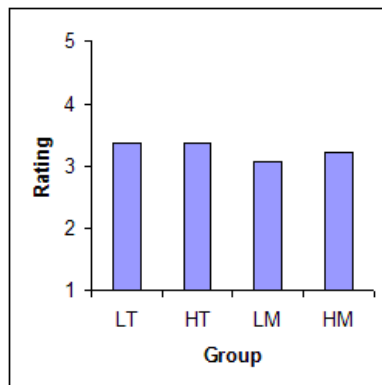
Table 6.16: Summary of means (*sd*) of subjective results for ERM-Tutor (2007)

The figures presented in Table 6.16 seem to be approximately equal, suggesting that all students, regardless of their spatial ability or the feedback mode presented to them, gave similar ratings to the questions. As illustrated in Figure 6.9, taking a closer look at the figures shows that the LM group gave slightly lower ratings for the *Terrible-Wonderful* scale, scoring a mean of 2.9 ($sd = 0.8$), and the *Difficult-Easy* scale, scoring a mean of 2.8 ($sd = 1.1$). This suggests that the students who were classified as low spatial and given the multimedia feedback found ERM-Tutor slightly more *terrible* and *difficult* than the other groups.

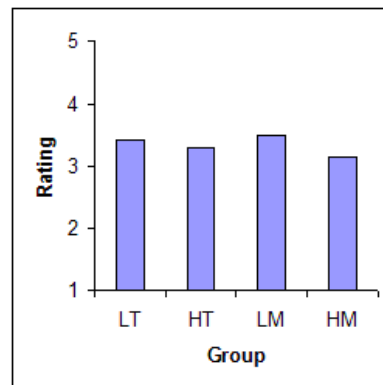
We used the *Kruskal-Wallis* analysis to examine whether or not the difference in mean scores for the LM group compared to the other groups is statistically significant. The *Kruskal-Wallis* analysis is used when there is one independent variable with more than two levels, and independent groups in each level. It is a parametric test equivalent to the non-parametric *one-way ANOVA* analysis. The results indicate that the Likert scale ratings between the four groups were not significantly different for either the *Terrible-Wonderful* scale (*Kruskal-Wallis* Test corrected for ties, $H = 7.8, df = 3, LT = 9, HT = 12, LM = 13, HM = 11, p = 0.119$), or the *Difficult-Easy* scale (*Kruskal-Wallis* Test corrected for ties, $H = 7.8, df = 3, LT = 9, HT = 12, LM = 13, HM = 11, p = 0.295$).



(a) ERM-Tutor ratings



(b) Feedback ratings



(c) Question-asking module ratings

Figure 6.9: Subjective ratings from the four groups

Moreover, we conducted the Kruskal-Wallis analysis on the rest of the rating questions and none produced significant difference. This indicates that the subjective opinions of the students did not differ significantly across the different groups, that is the students had similar experiences with ERM-Tutor.

We also examined the responses to the open questions. Overall, the comments were very positive and supportive of ERM-Tutor equally from the four groups. We present a range of examples of the submitted comments in Table 6.17. In general, the students appreciated and praised the procedural steps process in ERM-Tutor. A number of students wanted more specific feedback that would directly help them resolve their errors, others did not favour having to type the names of the constructs shown in the ER diagram and 7 out of 27 students, who responded to the question on the question-asking module, did not feel the need to use the module.

6.4 Discussion

We conducted a series of evaluation studies to test the two hypotheses of this research: firstly, that question-asking will have a positive effect on the students' performance, and secondly, the students who receive feedback mode matched to their spatial ability level will benefit more than the unmatched students.

Not surprisingly, the results showed an overall improvement in students' performance and level of domain knowledge after interacting with ERM-Tutor. Moreover, the improvements we made to ERM-Tutor for the 2007 study, meant that students could use it for a longer period, attempting as well as solving more problems. Nonetheless, the duration of our studies was quite short, which limited the amount of collected data. The studies ran during scheduled lab sessions, which were 2 hours long, including the

| Question | Examples of responses |
|--|---|
| What did you like about ERM-Tutor? | <p>It really helped me learn about ER mapping</p> <p>It has quite good error recognition and produces useful feedback on what the problem is. It also is good to teach the 7 steps.</p> <p>I like the steps of working 7 step. Honestly I still did not read much until now but ERM-Tutor is very helpful and guide me where to read on text book.</p> <p>It forced you to follow the steps in order</p> <p>How it split up the seven steps of mapping</p> <p>It helped me understand the steps more fully</p> <p>Once used to how it works, it is an excellent tool</p> <p>Very helpful hints</p> <p>It has good feedback</p> <p>The question section was useful</p> <p>Giving you clear answers to your questions</p> <p>I learnt something</p> |
| What changes would you like to see in ERM-Tutor? | <p>Error messages don't repeat</p> <p>Nothing</p> <p>More complex problems</p> <p>Quickly accessible notes on the current question to help when the hints can't</p> <p>Clickable names of entities and attributes etc to save typing</p> <p>Click-drag entity/attributs into the tables</p> <p>Very easy to misspell things making it frustrating. So clicking on an entity/attribute to fill in name would be good</p> <p>Attributes should change colour in diagram when added to table</p> <p>Interactive tutorial (I don't want to read text)</p> |
| What changes would you like to see in ERM-Tutor's feedback messages? | <p>Don't repeat!</p> <p>More information</p> <p>Still quite vague/maybe a little less initiative as big technical words were used</p> <p>Be more informative</p> <p>A little bit more precise if possible</p> <p>Show a comparison of what was entered and what is required in the solution</p> <p>It should tell where exactly the mistake is, preferably showing it on the graphical ER diagram</p> <p>I think they are pretty good</p> |

Table 6.17: Examples of students subjective comments (2007)

| Question | Examples of responses |
|--|--|
| What made you use/not use the questions-answering component? | <p>I got stuck so I used it</p> <p>Seems very useful tool for reminding learning about aspects of ERM</p> <p>If they have a list of question, I will use it</p> <p>I couldn't think of any questions to ask</p> <p>Didn't really use it, perhaps you could have an example about it (to show what it's used for). ie I thought the box would email the tutor</p> <p>I guess I knew what I was doing, so I didn't need to ask any question</p> <p>It was hard to phrase questions to the computer</p> <p>The feedback checking the answer was exactly the answers to the questions I had</p> <p>Why when I get what I need from help/hints</p> <p>The problems were so easy. Didn't require to use this feature</p> |

Table 6.17: Examples of students subjective comments (2007) continued

time to complete the consent forms, domain knowledge tests, spatial ability tests and the questionnaire.

First we look at the question-asking module. Due to the low number of students who used the module and subsequently the low number of questions submitted, we are unable to make any inferences on the impact of the question-asking module on the students' performance/learning. We suspect that the module was not fully utilised by the students because of the presence of human tutors during the evaluation studies. We felt that since the studies were carried out during the students' regular lab sessions, they were more inclined to turn to their regular tutors as they had been throughout the course. On the other hand, comments in the questionnaire indicate that a tutorial explaining the purpose of the module and how to use it, as well as presenting examples of frequently asked questions, would be beneficial for students.

We examined the questions submitted by the students and classified them into three categories; task-focused, definition-focused and phatic questions. The types as well as the quality/depth of questions asked resemble a typical student behaviour of wanting the correct solution as opposed to a deeper understanding of the material/domain. It is therefore, worth investigating various techniques to encourage students to use the question-asking module as well as *generate* questions, such as prompting students to ask more questions and even suggesting a question to ask based on their student model. In particular, there is a need for techniques that elicit deeper questions.

An interesting finding however, is the number of students who interacted with the module in a social manner by their phatic questions and expressions, as if there was a human responding to their submissions. We suspect that some of the motivations behind submitting those phatic questions might include '*testing the waters*'/'*breaking the ice*' before asking a domain specific question, testing the module or being curious of how it would respond, doing something different during a short break, wanting a social conversation or even just expressing their affective state (for example, being frustrated or bored as a result of using the system). It would be interesting to find out the motives behind those phatic submissions as well as investigate various techniques of dealing with them in an appropriate manner. For instance, to encourage and motivate the students to not only keep on using ERM-Tutor but also to ask questions.

Another finding worth noting is the students copying and pasting the system's feedback messages into the question-asking module. This indicates that they either did not fully understand the messages or did not know how to resolve the error themselves. This is reflected in the questionnaire's comments, for instance the students stated that they wanted "*more specific*" feedback as well as highlighting the particular error in their submission. Moreover, the addition of a question mark and/or the words "*Why is it*

saying” or “*How to*” indicates that the students perceive the question-asking module to only respond to questions.

Subsequently, we hoped that customising the system’s feedback messages with respect to the students’ spatial ability would give the students a greater understanding of the messages. This was proven to be necessary especially after analysing the types of questions the students submitted into the question-asking module. We classified the students as having either low or high spatial ability based on their spatial ability test scores, and examined their performance based on the feedback mode, either textual only or multimedia, they were presented with. We compared the data across four groups LT, HT, LM and HM, as well as across different combination of them; low versus high, textual versus multimedia and matched versus unmatched.

The results do not show a statistically significant difference between the pre- and post- tests scores across the four groups. We suspect this is, at least partially, due to the imbalance of prior domain knowledge before using ERM-Tutor and sitting the pre-test; the LT group had a higher pre-test score and hence improved the least, although this difference was not significant. Another possible reason for not finding a significant difference is that the students may be too familiar with the domain images/representation. In other words, although the multimedia messages have the potential of clarifying the text via graphics, there is a possibility that since the domain is already graphical in nature and the students are familiar with the constructs being used, the multimedia messages did not provide additional insights in comparison to the textual messages. Therefore, it is worth further evaluating whether or not students perceive a difference between the textual and multimedia feedback messages.

We did find, however, an interesting trend in the data after analysing the students’ log files and their interaction with the system. We looked at the total time interacting the system, number of attempted problems, number of solved problems, percentage

of solved problems and the total number of attempts/student solutions submitted. We found that the HM group had a consistently higher mean for all these types of interactions, followed by the HT group, then the LT group and lastly the LM group with the lowest mean. The four groups came out in the same order for all the types of interactions we examined. Although the difference in numbers is quite small and statistically insignificant, this trend is in line with Mayer's theory that high spatial students will benefit more from multimedia presentation. In other words, the high spatial students appreciated and used the system more when they received multimedia feedback messages, whereas the low spatial students were less inclined to use the system when they received the multimedia feedback messages.

Moreover, the numbers show that the high spatial students interacted more with the system than the low spatial students. This can be attributed to the graphical nature of the domain; the high spatial students were more comfortable with the mapping task. However, we did not find a correlation between the pre-test and spatial ability. It is also worth noting that although there was a noticeable difference between the high spatial students who received different feedback modes, the students who received the text feedback had similar means regardless of their spatial ability.

The same trend is reflected by the students' perception of the system indicated by their subjective results. Again although the difference is not statistically significant, it seems that the LM group consistently reported the lowest ratings for the system, finding it more difficult and boring than the other groups. An interpretation of this could be that because the LM group spent more cognitive effort processing the feedback messages and hence enjoyed ERM-Tutor the least.

In contrast, analyses of the mastery of constraints showed that the LM group had the highest learning rate compared to the other groups, indicating that they were less likely to re-violate a domain concept than the other groups. Although this indication was not

expected, it could be attributed to their lower interactions with the system (e.g. number of attempts per problem and number of attempted and solved problems) and lower prior domain knowledge shown in their pre-test mean score.

CHAPTER 7

Conclusions

Learning is an integral aspect of our daily life. People constantly face the challenge of acquiring new skills and knowledge. Such learning tasks are easier for those who are able to accurately assess their own knowledge, use the available resources, ask questions that clarify their understanding and solicit help when needed. Personal differences play a vital role in developing such skills. Research in the field of ITSs has shown that interactive computerised tutors are effective for enhancing learning and developing such meta-cognitive skills with plenty of scope for expansions and improvements.

Based on the belief that the mastery of question-asking skills has a significant influence on learning outcomes, and that accommodating for different psychometric measures facilitates the learning process, the presented research tested two hypotheses. First, answering students' open-ended questions will clarify their understanding and result in higher performance. Second, presenting the system's responses tailored to the students' spatial ability will lead to more effective learning and higher learning gain.

In this chapter we present our research conclusions that complete this thesis. The next section (Section 7.1) presents a summary of our research and outlines our contributions, followed by an overview of future directions of this research in Section 7.2.

7.1 Research Contribution

We are interested in enhancing the research field of ITSs through contributions to the meta-cognitive skill of question-asking and the spatial ability psychometric measure. As portrayed in Section 1.3, this research has four main objectives. First, development of an environment that engages learners in question-asking during their problem-solving learning experience. Second, evaluation of the effectiveness of the question-asking module and an analysis of the learners' behaviour towards it. Third, incorporation of a multimedia representation of the system's feedback messages and subsequently presentation of the feedback mode in accordance to the the students' spatial ability. Forth, evaluation of the effectiveness of tailoring the feedback messages towards the students' spatial ability levels. We have achieved our objectives.

In Chapter 3, we described ERM-Tutor, the test-bed ITS chosen for implementing our research framework. ERM-Tutor is a constraint-based tutoring system for teaching logical database design. It breaks the procedural task of mapping ER diagrams into relational schemas into the seven steps of the mapping algorithm and leads the student through them sequentially. The students are given feedback on their solutions after each component of the ER diagram they map, as well as at the end of each step of the algorithm. We enhanced the functionality of ERM-Tutor, furbished its interface and extended its domain knowledge (i.e. its constraints set). ERM-Tutor is now more robust and the students' are explicitly aware of their current state, or where exactly they are in problem-solving task. These modifications were shown to be effective by the increase in the length of interaction time as well as the number of attempted and solved problems.

We also modified ERM-Tutor to include our approaches for testing our hypotheses. Although we faced a number of limitations in our evaluation studies due to time constraints and number of participants involved, the results show evidence of the plau-

sibility of our research. The contributions of this research can be divided into two main parts, illustrated in detail below. The first main contribution covers our first and second objectives through our question-asking module (presented in Chapter 4), which enables students to ask free-form questions. The second contribution covers the third and fourth objectives through our approach of accounting for spatial ability in ERM-Tutor, as presented in Chapter 5.

7.1.1 Question-Asking Module

We successfully implemented the question-asking environment. We enhanced ERM-Tutor by incorporating the question-asking module, modifying the interface and extending the system's architecture. The module operates as follows. First, the students submit their questions through the provided interface. The module then processes the submission/query and uses the TFIDF information retrieval mechanism to retrieve the most appropriate question-answer pair from the database. Finally, the question-answer pair are displayed in the interface for the students. In contrast to a number of studies published in literature in the ITSs field, our question-asking module is fully automated, with no human tutor involvement.

We conducted a series of evaluation studies in which the students were able to use the question-asking module to test its effectiveness. The low number of students who participated in the studies as well as the short duration of the studies posed as considerable limitations in collecting sufficient data to effectively evaluate the module. Also due to time constraints, we chose to conduct our evaluation studies in a controlled environment during the students' usual lab sessions for the course. The presence of human tutors during the studies however, also imposed limitations on our results. We feel these limitations have influenced the low number of questions submitted by the students into

the question-asking module. As a consequence, we were unable to evaluate the effectiveness of the question-asking module on learning.

From the limited results we were able to examine the types of questions the students are willing to ask. The majority of questions submitted were task-focused, directly requesting help on the specific errors the students are faced with. Surprisingly, a number of students submitted phatic questions and expressions into the module, however the module did not include phatic entries and as a consequence the students' received a "*Not found*" message. Moreover, our question-asking module required students to phrase their submissions using domain problem-independent terminology, as well as the use of correct spelling; these are limitations in our module. Currently, our question-asking module is a stand-alone module that does not utilise the available information inside the student models, which limited its potential to effectively respond to the students questions.

Our results however, confirmed the need for addressing students' questions inside an ITS environment. Moreover, these results are an indication of the need for encouraging, as well as prompting, the students to ask questions.

Current research on meta-cognition in this field, focuses on self explanation, help seeking behaviour and self regulating learning. Our contribution is an initial step towards addressing the meta-cognitive skill of asking-questions and recognising the need for regular prompting. Moreover, our analyses show that students need more explanation on how to use the question-asking module. There is also a need to prompt the students to ask more questions.

7.1.2 Spatial Ability Module

We designed and created a multimedia representation of all the feedback messages in ERM-Tutor. ERM-Tutor was modified to measure the students' spatial ability level and present the multimedia feedback mode. This is the first implementation of multimedia feedback presentation mode in a constraint-based ITS.

Our evaluation studies presented students with one of the two feedback presentation modes, either textual or multimedia. We analysed the students' performance with respect to their spatial ability level and feedback mode they received. Again, the low number of students who participated and the short interaction times were limitations of our evaluation studies. The low number of students made it difficult to find statistically significant results especially when we categorised the students into various groups for further comparisons, as described in Chapter 6.

The results indicate that all students improved in their domain knowledge after interacting with ERM-Tutor. However, we did not find statistically significant difference between the pre- and post- tests scores across the four groups (LT, HT, LM and HM). Although we allocated equivalent number of students in each of the groups, we were unable to control for the students' prior existing knowledge that influence their gain scores between their pre- and post- tests.

Although we did not find conclusive results to support our hypothesis, we observed a number of trends in the collected data. In particular, there was a tendency for students with high spatial ability who received multimedia feedback to interact the most with the system. On the other hand, there was a tendency for students with low spatial ability to interact the least with the system. Moreover, there was no noticeable difference between students receiving textual feedback regardless of their spatial ability. These findings indicate that, in terms of interactions with the system (e.g. total time spent interacting

with the system and number of attempted and solved problems), the textual feedback had the same effect on students, whereas the multimedia messages had a greater effect on the high spatial students than on the low spatial students.

There is an increasing interest in the ITSs field to customise the learning environment towards the students' individual differences. Although our contributions towards accounting for the students' spatial ability lack statistically significant measures, there is evidence that matching the presentation of instruction towards the students spatial ability has an influence on their perception of the system and motivation to use it, more than their learning gain.

7.2 Future Work

Our contributions take the ITS research a step further and opens the doors for further research that will enrich the field, maximising the effectiveness of ITSs. We present a number of venues for future explorations for both the question-asking and spatial ability aspects of our research.

We gain a lot of insights about effective learning through analysing the behaviour and qualities of human tutors. An influential quality on learning is the ability to address each student individually according to their needs. In particular, if two students ask the same question, it is highly likely that the tutor will respond in a different manner to each student. This could be through the terminology used to answer the question, or the pedagogical strategy used to answer it, such as the use of examples or prompting the students to figure out the answer. Our current implementation of the question-asking module does not adapt towards individual students, that is, students who ask the same question will receive the same question-answer pair regardless of their knowledge state

or learning style. Therefore, it would be beneficial to incorporate different pedagogical strategies in answering the questions based on the student models.

In addition, the specific answer in each strategy should be based on the student's student model. For example, if the strategy used is to explain the answer using a worked example, then the example could reference a similar situation that the student was faced with while interacting with the system. It is also possible to integrate self explanation techniques into the module.

It is apparent that there is a need to encourage students to use their meta-cognitive skills. The skill of question-asking is no different, that is, it is worth investigating not only how to develop the students' question-asking skills, but also how to get students to use their question-asking skills. Our results showed a need for prompting the students to ask questions. Again, customisation of the prompts, such as the timing or frequency, would then be beneficial for students. This is because performance-oriented students may perceive such prompts as an overhead that overloads their cognitive capacity during problem solving. Similarly, the more domain competent students may perceive them as a burden or a distraction from their task.

Moreover, since research suggests that the quality of questions has an effect on learning, the system needs an automated mechanism or measure of quantifying the student questions. Questions are said to point to wholes in the learner's knowledge structure. Subsequently, in order for a question to be effective for a particular student it must for instance address a missing piece of information from their student model.

Our results showed an interesting tendency of submitting phatic questions as well as statements into the question-asking module. Addressing such submissions appropriately is important as it might affect the students' motivation, or even trust, to further use the module. In addition to exploring natural language processing (NLP) to address the variance in the question submissions, incorporating an AI *chatbot* and testing its

effectiveness is an interesting venue to explore. Moreover, it would be interesting to evaluate whether the presence of a pedagogical agent or avatar would have an effect on the perceived credibility and usefulness of the module or the motivation of the students, especially its effects on phatic submissions.

Another enhancement to the question-asking module would be to include an interactive tutorial. This tutorial could include an explanation of the purpose of the module and how to use it, as well as a number examples of effective questions to ask. Furthermore, recent research [Chi and VanLehn, 2007] suggests that explicit instruction of a problem-solving strategy improves students' performance in the domain. On the same grounds, we suspect that explicitly instructing students about the question-asking skill will teach student the skill as well as result in an increase in usage and higher quality of questions.

As for the spatial ability aspect of this research, our analyses suggest that although students have a range of spatial ability skills, their preferences could be different than their ability levels. It is therefore, worth further investigating whether students have a differing preference to their capabilities. If this is evident, then we suspect that allowing the students to choose their preferred feedback presentation mode would increase their motivation and influence a positive affective state.

Furthermore, the Spatial Module could be enhanced to utilise the students' interaction information stored in their student models. For instance, another evaluation study could investigate the effects of presenting an alternative presentation mode for subsequent violations of the same constraint; the first time a student violates a constraint they receive the textual representation for example, and if the same constraint is violated again they receive the multimedia representation.

We feel that the nature of our domain had an effect on the students' perception of the multimedia messages. In other words, since the students were familiar with the

ER constructs used in the messages and could constantly *see* them in the ER problem diagram presented on the screen, the multimedia messages might not have given any additional resources to the students. Therefore, it is worth investigating the effects of the instructional presentation mode with respect to the individual's spatial ability on learning in different types of domains. We suspect that there are limitations on certain domains or applications.

We were faced with the limitations of having a limited number of participants and short durations for the evaluation studies. Therefore, an evaluation study with more participants and a longer duration will yield more data that could result in statistically significant measures, as well as uncover a number of trends in students behaviour and perception towards the system.

Another direction for future research could be investigating the effects of good multimedia presentations on the transfer of learning. In particular, research could investigate whether students with varying spatial abilities encode knowledge differently based on the presentation modes they receive. If knowledge is encoded differently with different presentation modes, then how does that affect the transfer of learning?

Our research focused on investigating the effects of addressing question-asking and spatial ability in a constraint-based ITS; ERM-Tutor. The meta-cognitive skill of question-asking as well as the psychometric measure of spatial ability are domain-independent. We believe it is vital for future research to concentrate on domain-independent aspects that benefit individuals in becoming more effective learners.

References

- Aleven, V., McLaren, B., Roll, I., and Koedinger, K. R. (2004). Toward tutoring help seeking - applying cognitive modeling to meta-cognitive skills. In *Proceedings of 7th International Conference on Intelligent Tutoring Systems*, pages 455–467, Brazil. Springer.
- Alpert, S. R., Singley, M. K., and Fairweather, P. G. (1999). Deploying intelligent tutors on the web: An architecture and an example. *International Journal of Artificial Intelligence in Education*, 10(2):183–197.
- Anderson, J. R., Corbett, A. T., Koedinger, K. R., and Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of Learning Sciences*, 4:167–207.
- Anthony, L., Corbett, A. T., Wagner, A. Z., Stevens, S. M., and Koedinger, K. R. (2004). Student question-asking patterns in an intelligent algebra tutor. In *Proceedings of 7th International Conference on Intelligent Tutoring Systems*, volume 3220/2004, pages 455–467, Brazil. Springer.
- Baddeley, A. D. (1986). *Working Memory*. Oxford University Press, Oxford, England.
- Baenninger, M. and Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles*, 20(5):327–344.

- Baghaei, N., Mitrović, A., and Irwin, W. (2005). A Constraint-Based Tutor for Learning Object-Oriented Analysis and Design using UML. *International Conference on Computers in Education ICCE*, pages 11–18.
- Beck, I. L., McKeown, M. G., Hamilton, R. L., and Kucan, L. (1997). Questioning the author: An approach for enhancing student engagement with text. *Delaware: International Reading Association*.
- Bloom, B. S. (1956). Taxonomy of educational objectives: The classification of educational goals.
- Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6):4–16.
- Bransford, J. D., Brown, A. L., and Cocking, R. R. (2000). How people learn: Brain, mind, experience, and school.
- Brown, A. L. (1987). Metacognition, motivation, and understanding.
- Butcher, J. N. (1992). *MMPI-A: Minnesota Multiphasic Personality Inventory-Adolescent: Manual for Administration, Scoring, and Interpretation*. University of Minnesota Press: distributed by National Computer Systems, Inc.
- Carroll, J. (1993). *Human cognitive abilities*. Cambridge University Press, England.
- Chandler, P. and Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction*, 8(4):293–332.
- Chen, P. P. I. N. S. (1976). The Entity-Relationship Model-Toward a Unified View of Data. *ACM Transactions on Database Systems*, 1(1):9–36.

- Chi, M. and VanLehn, K. (2007). The impact of explicit strategy instruction on problem-solving behaviors across intelligent tutoring systems. In *Proceedings of the 29th Annual Conference of the Cognitive Science Society*. Erlbaum.
- Chi, M. T. H., de Leeuw, N., Chiu, M., and LaVanher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18:439–477.
- Chin, D. N. (2001). Empirical Evaluation of User Models and User-Adapted Systems. *International Journal of User Modeling and User-Adapted Interaction*, 11(1):181–194.
- Cohen, P. A., Kulik, J. A., and Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, 19:237–248.
- Conati, C. and Maclaren, H. (2004). Evaluating a probabilistic model of student affect. In *Proceedings of 7th International Conference on Intelligent Tutoring Systems*, volume 3220/2004, pages 55–66, Brazil. Springer.
- Corbett, A., Wagner, A., Yu Chao, C., Lesgold, S., Stevens, S., and Ulrich, H. (2005). Student questions in a classroom evaluation of the ALPS learning environment. In *Proceedings of 12th International Conference on Artificial Intelligence in Education*, volume 125, pages 780–782, Amsterdam, The Netherlands. IOS Press.
- Corbett, A. T. (2001). Cognitive computer tutors: Solving the two-sigma problem. In *Proceedings of User Modeling*, pages 137–147.
- Crocker, L. and Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando.

- de Vicente, A. and Pain, H. (2002). Informing the detection of students' motivational state: An empirical study. In *Proceedings of 6th International Conference on Intelligent Tutoring Systems*, pages 933–943, Berlin. Springer.
- Dillon, T. J. (1988). *Questioning and teaching: A manual of practice*.
- Ekstrom, R., French, J., and Harman, H. (1997). *Manual for kit of factor referenced cognitive tests*.
- Elmasri, R. and Navathe, S. B. (2000). *Fundamentals of database systems*. Addison-Wesley Reading, Mass.
- Fancher, R. E. (1985). *The Intelligence Men: Makers of the IQ Controversy*. Norton.
- Gall, M. (1984). Synthesis of research on teachers' questioning. *Educational Leadership*, 42:40–47.
- Gama, C. (2004). Metacognition in interactive learning environments: The reflection assistant model. In *Proceedings of 7th International Conference on Intelligent Tutoring Systems*, pages 668–677.
- Graesser, A. C., Hu, X., Person, P., Jackson, T., and Toth, J. (2005). Modules and information retrieval facilities of the human use regulatory affairs advisor. *International Journal on eLearning*.
- Graesser, A. C. and Olde, B. A. (2003). How does one know whether a person understands a device? the quality of the questions the person asks when the device breaks down. *Journal of Educational Psychology*, 95:524–536.
- Graesser, A. C. and Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, 31:104–137.

- Guay, R. B. (1977). Purdue spatial visualization test: Rotations. *Purdue Research Foundation, West Lafayette, IN.*
- Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., and Lovelace, K. (2006). Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence*, 34:151–176.
- Heiner, C. (2007). Towards A Virtual Teaching Assistant to Answer Questions Asked by Students in Introductory Computer Science. In *Proceedings of 13th International Conference on Artificial Intelligence in Education*, pages 689–690, Los Angeles. IOS Press.
- Jensen, A. R. (1999). The G Factor: the Science of Mental Ability. *Psicothema*, 11(2):445–446.
- Karabenick, S. A. (1998). Strategic help seeking. implications for learning and teaching.
- Kelly, D. and Tangney, B. (2004). Predicting learning characteristics in a multiple intelligence based tutoring system. In *Proceedings of 7th International Conference on Intelligent Tutoring Systems*, pages 678–688, Brazil. Springer.
- King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal*, 31:338–368.
- Koedinger, K. R., Anderson, J. R., Hadley, W. H., and Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, 8(1):30–43.
- Kort, B. and Reilly, R. (2002). A pedagogical model for teaching scientific domain knowledge. *Frontiers in Education, 2002. FIE 2002. 32nd Annual*, 1.

- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Lawrence Erlbaum.
- Marshall, M. (2004). ERM tutor and ER-to-relational mapping tutor. Honours report, University of Canterbury.
- Martin, B. (2001). Intelligent tutoring systems: The practical implementation of constraint-based modelling. Phd thesis, University of Canterbury.
- Martin, B., Koedinger, K., Mitrović, A., and Mathan, S. (2005). On Using Learning Curves to Evaluate ITS. In *Proceedings of 12th International Conference on Artificial Intelligence in Education*, pages 419–426, Amsterdam. IOS Press.
- Martin, B. and Mitrović, A. (2005). Using Learning Curves to Mine Student Models. In *Proceedings of the 10th International Conference on User Modelling*, volume 3538, pages 79–88, Edinburgh. Springer.
- Martin, B. and Nicholas, A. (2007). Studying model ambiguity in a language ITS. In *Proceedings of 11th International Conference on User Modeling*, pages 425–429, Corfu, Greece.
- Mayer, R. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, 32:1–19.
- Mayer, R. (2001). *Multi-media Learning*. University of California.
- Mayer, R. and Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38(1):43–52.
- Mayer, R., Sobko, K., and Mautone, P. (2003). Social cues in multimedia learning: Role of speaker's voice. *Journal of Educational Psychology*, 95(2):419–425.

- Mayo, M. and Mitrović, A. (2001). Optimising ITS Behaviour with Bayesian Networks and Decision Theory. *International Journal of Artificial Intelligence in Education*, 12(3).
- McCrae, R. and Costa Jr, P. (1987). Validation of the five-factor model of personality across instruments and observers. *J Pers Soc Psychol*, 52(1):81–90.
- Milik, N., Marshall, M., and Mitrović, A. (2006). Responding to free-form student questions in ERM-tutor. In Chan, M. I., Ashley, K. D., and Tak-Wai, editors, *Proceedings of 8th International Conference on Intelligent Tutoring Systems*, pages 707–709, Taiwan. Springer.
- Milik, N., Mitrović, A., and Grimley, M. (2007a). Fitting spatial ability and free-formed questions into Intelligent Tutoring Systems development. In *Proceedings of New Zealand Computer Science Research Student Conference*, Hamilton.
- Milik, N., Mitrović, A., and Grimley, M. (2007b). Fitting Spatial Ability into Intelligent Tutoring Systems Development. In *Proceedings of 13th International Conference on Artificial Intelligence in Education*, pages 617–619, Los Angeles. IOS Press.
- Miller, G. A. (1956). *The Magical Number Seven, Plus Or Minus Two: Some Limits on Our Capacity for Processing Information*. Bobbs-Merrill.
- Mitrović, A. (1998). Experiences in Implementing Constraint-Based Modeling in SQL-Tutor. *Proceedings of the 4th International Conference on Intelligent Tutoring Systems*, pages 414–423.
- Mitrović, A. (2002). NORMIT, a web-enabled tutor for database normalization. *Proceedings of the International Conference on Computers in Education*, pages 1276–1280.

- Mitrović, A. (2005). The Effect of Explaining on Learning: a Case Study with a Data Normalization Tutor. In *Proceedings of 12th International Conference on Artificial Intelligence in Education*, volume 125, pages 499–506, Amsterdam, The Netherlands. IOS Press.
- Mitrović, A., Martin, B., and Mayo, M. (2002). Using Evaluation to Shape ITS Design: Results and Experiences with SQL-Tutor. *International Journal of User Modeling and User-Adapted Interaction*, 12(2–3):243–279.
- Mitrović, A., Mayo, M., Suraweera, P., and Martin, B. (2001). Constraint-based tutors: A success story. *IEA/AIE*, pages 931–940.
- Mitrović, A. and Ohlsson, S. (1999). Evaluation of a constraint-based tutor for a database language. In *Proceedings of International Conference on Artificial Intelligence in Education*, pages 3–4.
- Mitrović, A., Suraweera, P., Martin, B., and Weerasinghe, A. (2004). DB-Suite: Experiences with Three Intelligent, Web-Based Database Tutors. *Journal of Interactive Learning Research*, 15(4):409–433.
- Miyake, N. and Norman, D. A. (1979). To ask a question one must know enough to know what is not known. *Journal of Verbal Learning and Verbal Behavior*, 18:357–364.
- Moffat, S. D., Hampson, E., and Hatzipantelis, M. (1998). Navigation in a virtual maze: Sex differences and correlation with psychometric measures of spatial ability in humans. *Evolution and Human Behavior*, 19(2):73–87.
- Moffat, S. D., Zonderman, A. B., and Resnick, S. M. (2001). Age differences in spatial memory in a virtual environment navigation task. *Neurobiology of Aging*, 22(5):787–796.

- Moreno, R. and Mayer, R. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, 91(2):358–368.
- Ohlsson, S. (1994). Constraint-based student modeling. In *Student Modeling: the Key to Individualized Knowledge-based Instruction*, pages 167–189, Berlin. Springer.
- Ohlsson, S. (1996). Learning from Performance Errors. *Psychological Review*, 103(2):241–62.
- Otero, J. and Graesser, A. C. (2001). PREG: Elements of a model of question asking. *Cognition & Instruction*, 19:143–175.
- Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. Oxford University Press.
- Pittenger, D. J. (1993). The utility of the Myers-Briggs Type Indicator. *Review of Educational Research*, 63(4):467–488.
- Pressley, M. and Forest-Pressley, D. (1985). The psychology of questions.
- Rasch, G. (1980). Probabilistic models for some intelligence and attainment tests.
- Roll, I., Aleven, V., McLaren, B., and Koedinger, K. (2007). Can Help Seeking Be Tutored? Searching for the Secret Sauce of Metacognitive Tutoring. In *Proceedings of 13th International Conference on Artificial Intelligence in Education*, pages 203–210, Los Angeles. IOS Press.
- Salton, G. and McGill, M. J. (1986). *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc. New York, NY, USA.
- Sanders, N. M. (1966). Classroom questions: What kinds?

- Scardamalia, M. and Bereiter, C. (1985). Fostering the development of self-regulation in children's knowledge processing. *Thinking and learning skills*, 2:563–577.
- Schank, R. and Cleary, C. (1995). *Engines for Educators*. Lawrence Erlbaum Associates, Publishers, Hillsdale, NJ.
- Self, J. A. (1990). Bypassing the Intractable Problem of Student Modeling. In Frason, C. and Gauthier, G., editors, *Intelligent Tutoring Systems: at the Crossroad of Artificial Intelligence and Education*, pages 107–123. Ablex Publishing Corporation, Norwood, New Jersey.
- Snow, C. (2002). Reading for understanding: Toward an R & D program in reading comprehension.
- Steinke, M., Huk, T., and Floto, C. (2004). The influence of cognitive abilities and the presence of 3D models on the use of task relevant content in hypermedia learning systems. *Proceedings of ED-MEDIA*.
- Sternberg, R. J. (1990). *Metaphors of Mind: Conceptions of the Nature of Intelligence*. Cambridge University Press.
- Stumpf, H. and Fay, E. (1983). Schlauchfiguren-ein test zur beurteilung des rumlichen vorstellungsvermgens. Gttingen.
- Sullins, J., Azevedo, R., Trousdale, A., and Scott, J. (2007). The Influence of External-Regulation on Student Generated Questions during Hypermedia Learning. In *Proceedings of 13th International Conference on Artificial Intelligence in Education*, pages 187–194, Los Angeles. IOS Press.

- Suraweera, P. and Mitrović, A. (2004). An intelligent tutoring system for entity relationship modelling. *International Journal of Artificial Intelligence in Education*, 14(3):375–417.
- Thorndike, R. L., Hagen, E. P., and Sattler, J. M. (1986). *Stanford-Binet Intelligence Scale*. Riverside Chicago, Ill.
- Tyler, L. E. (1965). *The Psychology of Human Differences*. Appleton-Century-Crofts New York.
- Vicente, K. J. and Williges, R. C. (1988). Accommodating individual differences in searching a hierarchical file system. *International Journal of Man-Machine Studies*, 29(6):647–668.
- WebCT, I. and Lane, S. K. (2002). Leveraging Technology to Transform the Educational Experience. *WebCT Inc*.
- Webster (2004). Webster's new millennium dictionary of english. <http://dictionary.reference.com/search?q=metacognition>.
- Wisher, R. A. and Graesser, A. C. (2005). Toward a science of distributed learning and training. chapter Question asking in advanced distributed learning environments. American Psychological Association, Washington, D.C.

APPENDIX A

Information and Consent Form

Adhering to the human ethics regulations in conducting evaluation studies, all participants received the following information describing both the procedure and purpose of the study. The participants were also free to consent to the study.

ERM-Tutor

Thank you for participating in this evaluation study. The aim of the study is to investigate the effectiveness of ERM-Tutor on learning. You are expected to work individually, solving problems at your own pace. You are encouraged to ask the system as many questions as you can.

Before you login, make sure you have answered the pre-test questions given to you, and at the end of this session, after you have used the ERM-Tutor, make sure you have answered the post-test questions.

This evaluation is not assessing your competence or intelligence in any way. All the data reported on this study will be anonymous. You are free to stop the session at any time, and also to require that your session is not used in the study.

This project is carried out by Nancy Milik – a M.Sc student at the Department of Computer Science and Software Engineering, University of Canterbury. She can be contacted through email at nmi14@student.canterbury.ac.nz. She will be pleased to discuss any concerns you may have about participating in the project.

Consent Form for Study on ERM-Tutor

I have read and understood the description of the above-named project. On this basis, I agree to participate in this project, and I consent on the publication of the results of the project with the understanding that anonymity will be preserved. I understand also that I may at any time withdraw from the project, including withdrawal of any information I have provided.

Signed: Date:

Usercode:

APPENDIX **B**

Domain Knowledge Tests

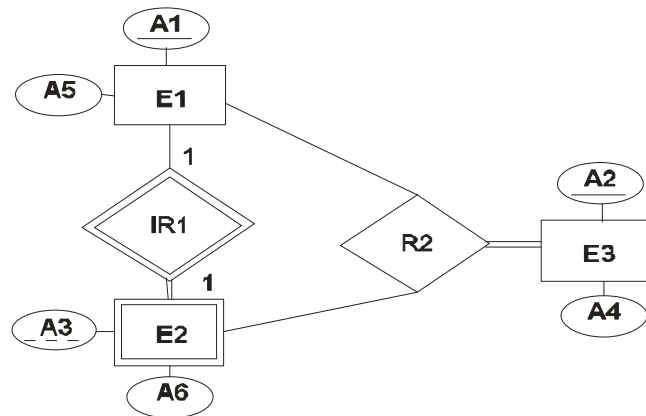
B.1 Test Version A

ER-Mapping Pre-test

userid:

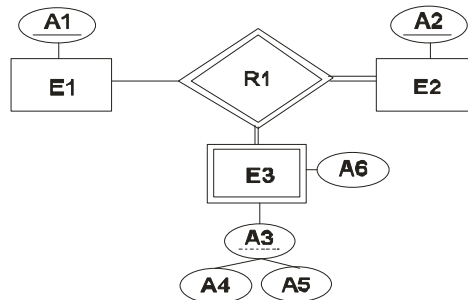
1. When mapping an ER schema into a relational schema, in what situation a new relation is NOT added to the database:
 - a. for each M:N binary relationship type
 - b. for each higher degree relationship
 - c. for each multivalued attribute
 - d. for each weak or regular entity type
 - e. for each 1:1 or 1:N binary relationship type

2. Which of the following two relational schemas correspond to the given ER schema?
 - a. E1(A1, A5)
E2(A1, A3, A6)
E3(A2, A4)
R2(A1, A3, A2)
 - b. E1(A1, A5)
E2(A1, A3, A6)
E3(A2, A4)
R2(A1, A2)
 - c. Both of a and b are correct
 - d. Neither a nor b are correct



3. When mapping a binary 1:N relationship type, a foreign key is added to:
 - a. the relation corresponding to the entity type on the 1 side of relationship type
 - b. the relation corresponding to the entity type on the N side of relationship type
 - c. both relations
 - d. none of the above

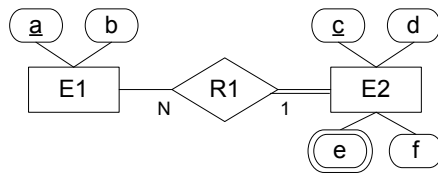
4. Table R3 corresponds to the weak entity type E3 from the given ER diagram.



R3(A1, A2, A3, A4, A5, A6)

True False

5. Map the following ER diagram into its appropriate relational schemas (tables).



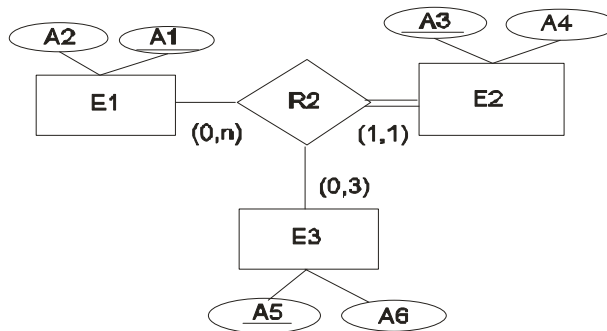
B.1.1 solution**B.2 Test Version B**

ER-Mapping Post-test

usercode:

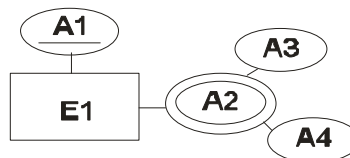
1. Which of the following options is true for mapping a multivalued attribute that belongs to a relationship type?
 - a. A foreign key is added to the relation corresponding to the entity type that participates fully in the relationship type.
 - b. There is a new relation created for the multivalued attribute.
 - c. A foreign key is added to the relation corresponding to the entity type on the N side of the relationship type.
 - d. There is a new relation created for the relationship type, which also contains the multivalued attribute.

2. Which of the following two relational schemas correspond to the given ER schema?
 - a. E1(A1, A2)
E2(A3, A4)
E3(A5, A6)
R2(A1, A3, A5, A6)
 - b. E1(A1, A2)
E2(A3, A4)
E3(A5, A6)
R2(A1, A3, A5, A6)
 - c. Both of a and b are correct
 - d. Neither a nor b are correct



3. When mapping a binary M:N relationship type, a foreign key is added to:
 - a. the relation corresponding to the entity type on the 1 side of relationship type
 - b. the relation corresponding to the entity type on the N side of relationship type
 - c. both relations
 - d. none of the above

4. Table R2 corresponds to attribute A2 from the given ER diagram.

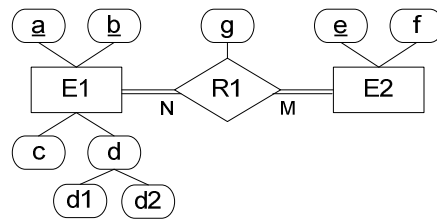


R2(A1, A2, A3, A4)

True

False

5. Map the following ER diagram into its appropriate relational schemas (tables).



APPENDIX C

Questionnaire

The following questionnaire was used in both evaluation studies.

Questionnaire

usercode:

1. How would you rate the overall quality of ERM-Tutor?

1 ----- 2 ----- 3 ----- 4 ----- 5
Poor Excellent

2. Rate your impression of ERM-Tutor:

1 ----- 2 ----- 3 ----- 4 ----- 5
Terrible Wonderful

1 ----- 2 ----- 3 ----- 4 ----- 5
Difficult Easy

1 ----- 2 ----- 3 ----- 4 ----- 5
Boring Fun

3. What did you like about ERM-Tutor?

4. What changes would you like to see in ERM-Tutor?

5. How would you rate the overall quality of the feedback messages from ERM-Tutor?

1 ----- 2 ----- 3 ----- 4 ----- 5 I haven't
Poor Excellent used it

6. What changes would you like to see in ERM-Tutor's feedback messages?

7. How would you rate the overall relevance/quality of the answers given by the system for your questions?

1 ----- 2 ----- 3 ----- 4 ----- 5 I haven't
Poor Excellent used it

8. What made you use/not use the questions-answering component?

Other comments about ERM-Tutor

APPENDIX D

Spatial Ability Tests

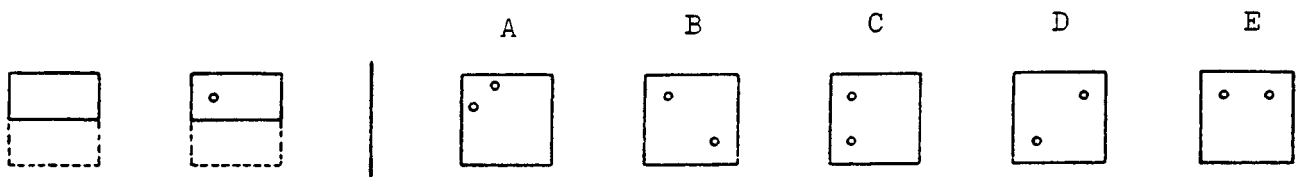
These tests were taken from Ekstrom, R., French, J., and Harman, H. (1997). Manual for kit of factor referenced cognitive tests

D.1 Paper Fold Test

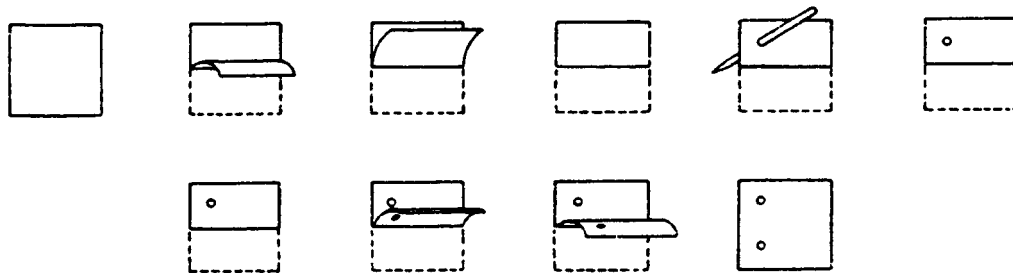
PAPER FOLDING TEST — VZ-2

In this test you are to imagine the folding and unfolding of pieces of paper. In each problem in the test there are some figures drawn at the left of a vertical line and there are others drawn at the right of the line. The figures at the left represent a square piece of paper being folded, and the last of these figures has one or two small circles drawn on it to show where the paper has been punched. Each hole is punched through all the thicknesses of paper at that point. One of the five figures at the right of the vertical line shows where the holes will be when the paper is completely unfolded. You are to decide which one of these figures is correct and draw an X through that figure.

Now try the sample problem below. (In this problem only one hole was punched in the folded paper.)



The correct answer to the sample problem above is C and so it should have been marked with an X. The figures below show how the paper was folded and why C is the correct answer.

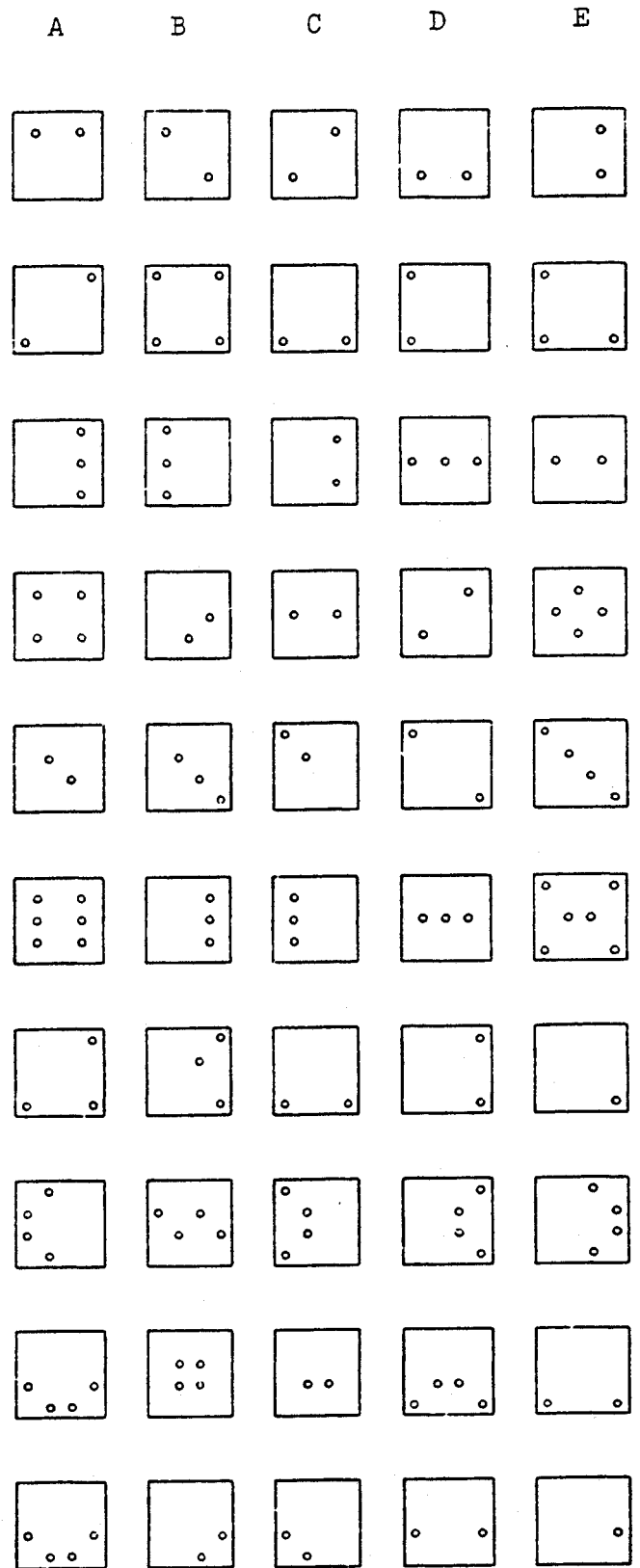
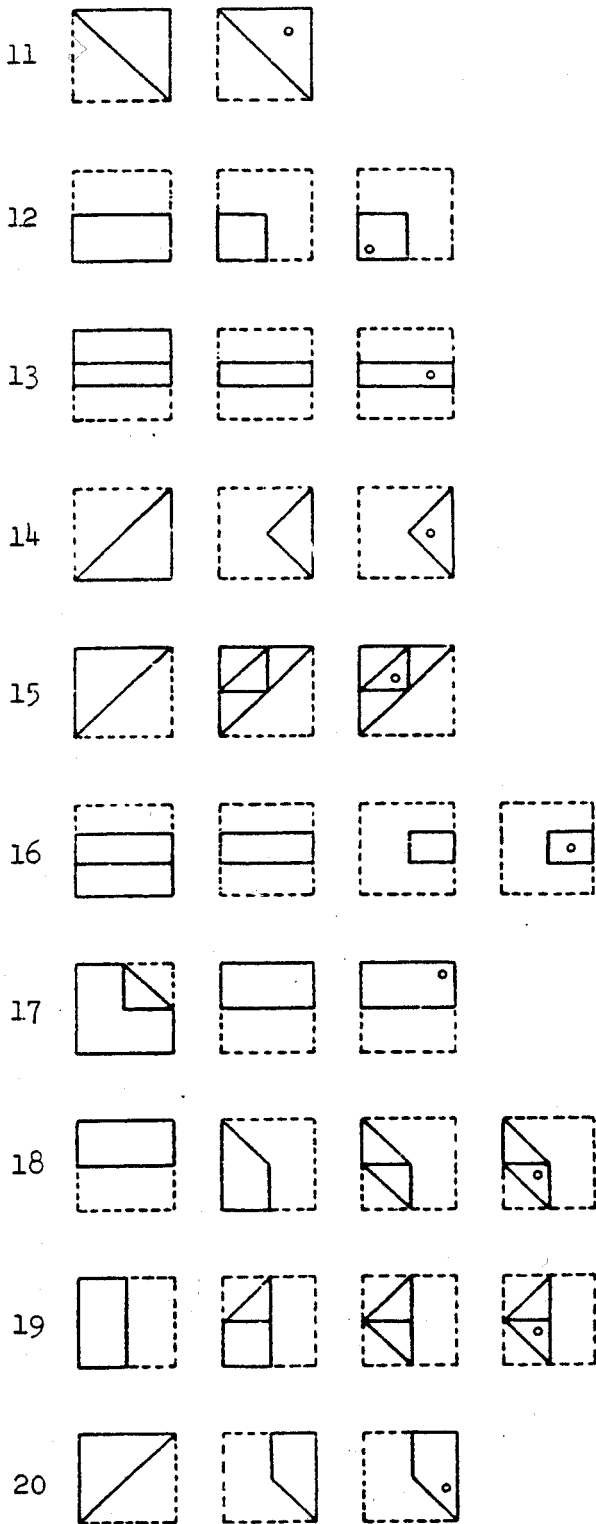


In these problems all of the folds that are made are shown in the figures at the left of the line, and the paper is not turned or moved in any way except to make the folds shown in the figures. Remember, the answer is the figure that shows the positions of the holes when the paper is completely unfolded.

Your score on this test will be the number marked correctly minus a fraction of the number marked incorrectly. Therefore, it will not be to your advantage to guess unless you are able to eliminate one or more of the answer choices as wrong.

You will have 3 minutes for each of the two parts of this test. Each part has 1 page. When you have finished Part 1, STOP. Please do not go on to Part 2 until you are asked to do so.

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.



DO NOT GO BACK TO PART 1, AND

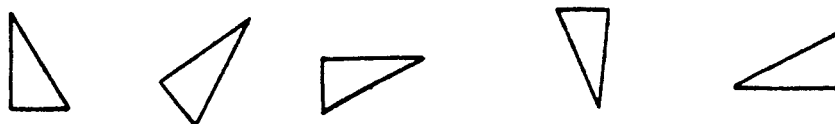
DO NOT GO ON TO ANY OTHER TEST UNTIL ASKED TO DO SO.

STOP.

D.2 Card Rotation Test

CARD ROTATIONS TEST — S-1 (Rev.)

This is a test of your ability to see differences in figures. Look at the 5 triangle-shaped cards drawn below.



All of these drawings are of the same card, which has been slid around into different positions on the page.

Now look at the 2 cards below:



These two cards are not alike. The first cannot be made to look like the second by sliding it around on the page. It would have to be flipped over or made differently.

Each problem in this test consists of one card on the left of a vertical line and eight cards on the right. You are to decide whether each of the eight cards on the right is the same as or different from the card at the left. Mark the box beside the S if it is the same as the one at the beginning of the row. Mark the box beside the D if it is different from the one at the beginning of the row.

Practice on the following rows. The first row has been correctly marked for you.



























































































| | | | | | | | | | | | | | | | |
|---|--|--|---|--|--|--|---|--|--|--|--|--|--|--|--|
| B | | | <input checked="" type="checkbox"/> S <input checked="" type="checkbox"/> D | | <input type="checkbox"/> S <input checked="" type="checkbox"/> D | | <input checked="" type="checkbox"/> S <input checked="" type="checkbox"/> D | | <input type="checkbox"/> S <input checked="" type="checkbox"/> D | | <input type="checkbox"/> S <input checked="" type="checkbox"/> D | | <input type="checkbox"/> S <input checked="" type="checkbox"/> D | | <input type="checkbox"/> S <input checked="" type="checkbox"/> D |
| | | | <input type="checkbox"/> S <input type="checkbox"/> D | | <input type="checkbox"/> S <input type="checkbox"/> D | | <input type="checkbox"/> S <input type="checkbox"/> D | | <input type="checkbox"/> S <input type="checkbox"/> D | | <input type="checkbox"/> S <input type="checkbox"/> D | | <input type="checkbox"/> S <input type="checkbox"/> D | | <input type="checkbox"/> S <input type="checkbox"/> D |
| | | | <input type="checkbox"/> S <input type="checkbox"/> D | | <input type="checkbox"/> S <input type="checkbox"/> D | | <input type="checkbox"/> S <input type="checkbox"/> D | | <input type="checkbox"/> S <input type="checkbox"/> D | | <input type="checkbox"/> S <input type="checkbox"/> D | | <input type="checkbox"/> S <input type="checkbox"/> D | | <input type="checkbox"/> S <input type="checkbox"/> D |

Your score on this test will be the number of items answered correctly minus the number answered incorrectly. Therefore, it will not be to your advantage to guess, unless you have some idea whether the card is the same or different. Work as quickly as you can without sacrificing accuracy.

You will have 3 minutes for each of the two parts of this test. Each part has 1 page. When you have finished Part 1, STOP. Please do not go on to Part 2 until you are asked to do so.

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.

Part 1 (3 minutes)

| | | | | | | | | | |
|-----|---|--|--|--|--|--|--|--|--|
| 1. |  |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |
| 2. |  |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |
| 3. |  |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |
| 4. |  |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |
| 5. |  |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |
| 6. |  |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |
| 7. |  |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |
| 8. |  |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |
| 9. |  |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |
| 10. |  |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |  <input type="checkbox"/> |

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.

STOP.

APPENDIX E

ITS 2006 Short Paper

The following paper was presented at the 8th International Conference on Intelligent Tutoring Systems held in Jhongli, Taiwan, June 2006.

Milik, N., Marshall, M., and Mitrović, A. (2006). Responding to free-form student questions in ERM-tutor. In Chan, M. I., Ashley, K. D., and Tak-Wai, editors, *Proceedings of 8th International Conference on Intelligent Tutoring Systems*, pages 707–709, Taiwan. Springer

Responding to Free-form Student Questions in ERM-Tutor

Nancy Milik, Melinda Marshall, Antonija Mitrovic

Intelligent Computer Tutoring Group
Department of Computer Science and Software Engineering
University of Canterbury, Christchurch, New Zealand
{nmi14,tanja}@cosc.canterbury.ac.nz

Abstract. We present ERM-Tutor, a constraint-based tutor that teaches logical database design (i.e. mapping conceptual to logical database schemas). Students practice this procedural task in ERM-Tutor by solving each step and receiving feedback on their solutions. We also present a new feature added to the system, which enables students to ask free-form questions. A preliminary evaluation carried out on ERM-Tutor investigated how students use free-form questions, and provided promising results. We plan to perform a bigger study in early 2006.

1 Introduction

Constraint-based tutors have been successful in a variety of domains, such as conceptual database design, database queries, data normalization, UML and language learning [2,5,6]. Building on successful work, we have developed ERM-Tutor, in which students practice the algorithm for mapping conceptual database schemas (i.e. ER diagrams) into relational schemas. The ER-to-relational algorithm [4] consists of seven steps, which map the ER components in the following order: 1) regular entities, 2) weak entities, 3) 1:1 binary relationships, 4) 1:N binary relationships, 5) M:N relationships, 6) multivalued attributes, and 7) n -ary relationships. Although the algorithm is well-defined and short, students typically find it difficult to learn and apply consistently.

ERM-Tutor is a web-based system, the main components of which are the pedagogical module, problem solver, student modeler, session manager and user interface. The tutor also contains a set of problems and 121 constraints representing the domain knowledge. The problem-solving process is broken into seven tasks, corresponding to steps in the mapping algorithm, each task presented to the student on a separate page. The student has to complete the current task in order to move on to the next one. The student can request feedback at any time. The short-term student model consists of a list of satisfied and a list of violated constraints. This model is used by the pedagogical module to present feedback to the student. ERM-Tutor also maintains a long-term student model, which is used for problem selection.

2 Question Asking Module

Intelligent Tutoring Systems provide feedback on students' actions, but students do not always understand the feedback they receive. Therefore, it would be beneficial for students to be able to ask free-form questions at any time. ALPS [1,3] allows the student to ask any question, to which the system replies with a pre-recorded video clip. The results show that the rate of unprompted questions is lower than in the case of one-on-one human tutoring. Furthermore, half of the questions are not related to problem-solving, but are rather social interactions. Most of the remaining questions are performance-oriented, and not deep questions that would facilitate learning.

In this light, we added a question-asking module to ERM-Tutor. We defined 98 distinct questions, based on our experiences in teaching the mapping algorithm and our experience with other constraint-based tutors. These questions can be categorized into interface usage ("What does the button *Check Step* do?"), definitions of terms ("What is a foreign key?"), diagram notations ("How is an attribute represented in the ER-diagram?"), mapping regulations ("How is a relationship mapped?"), and deeper questions ("Why are the steps arranged in this order?"). The question database additionally includes a number of repeated questions that are phrased differently, resulting in a total of 182 questions. In contrast to ALPS, the answers to questions are textual.

The TFIDF (Term Frequency Inverse Document Frequency) vector weighting scheme [7] was chosen as the information retrieval mechanism, as is the case in ALPS. In our system, the questions are read from the database and separated into words. The weight of each question and word is calculated, and words are indexed in a hash table. When the student asks a question, the same calculations are applied to the query string: it is also broken-up into words and their weights are calculated. Each question is then allocated a query weight. Finally, the answer corresponding to the question with the highest query weight is returned to the student. To evaluate the subjective relevance of answers, students are encouraged to submit their ratings of answers; however, the system does not enforce it to avoid mode errors and distractions from the problem solving task.

3 Preliminary Evaluation

We performed a preliminary study of ERM-Tutor with students enrolled in an introductory database course at the University of Canterbury in 2005, in order to investigate the usage of free-form questions. 29 students logged into ERM-Tutor at least once, but five students used it for less than two minutes and so their logs were excluded from analyses. The average interaction time was under one hour (mean=54min, sd=63min), ranging from several minutes to 4.5 hours over several weeks. The number of sessions ranged from one to four (mean=1.67, sd=0.96). On average, students attempted 4.6 problems and completed 25% of them.

Only eight students asked questions, with a total of 24 questions submitted. The number of questions per student ranged from one to five. The questions can be categorized into task-focused (50%), definition-focused (8%) and phatic questions (42%). Task-focused questions ask directly for help solving the problem (e.g. "How could I solve this table?"). For instance, three students copied the feedback messages,

added a question mark at the end or a “How to” at the start, and submitted them as the questions. Definition-focused questions ask for definition of terms. There were only two such questions submitted: “What is foreign key?” and “What is multivalued?” Phatic questions establish a sense of social mood (e.g. “What is your name?”, “How are you?” and “How do you answer questions?”). Excluding phatic questions, 14 questions were relevant for students’ actions. Five of these questions were answered correctly, and for two of these, the students specified highest relevance. The answer could not be found for one question. The remaining questions received answers which were related to the query, but were not useful to students. This happened when the students did not formulate questions well, but instead copied a part of the feedback message, adding a question mark at the end (e.g. “Make sure the relationship is 1:1?”). We intend to enhance our question database with these questions.

4 Conclusions

The paper presented ERM-Tutor, a new constraint-based tutor that teaches the procedural task of mapping ER diagrams into relational schemas. We enhanced ERM-Tutor with a question-asking module, which allows the student to ask a free-form question, which the system processes and returns the answer with the highest relevance weight, using the TFIDF weighting scheme. Our preliminary study showed some evidence that students welcome the idea of asking free-form questions and confirmed the need for eliciting deeper questions. We are currently investigating various techniques to encourage students to use the module, such as prompting students to ask more questions and even suggesting a question to be asked based on their student model. We plan to conduct a full evaluation study of ERM-Tutor in March 2006.

References

1. Anthony, L., Corbett, A., Wagner, A., Stevens, S., Koedinger, K. Student Question-Asking Patterns in an Intelligent Algebra Tutor. In: J. Lester, R. M. Vicar, Paraguacu, F. (eds) Proc. 7th Int. Conf. Intelligent Tutoring Systems, pp. 455-467, 2004.
2. Baghaei, N., Mitrovic, A., Irwin, W. A Constraint-Based Tutor for Learning Object-Oriented Analysis and Design using UML. In: C.K. Looi, D. Jonassen, M. Ikeda (eds), Int. Conf. Computers in Education ICCE 2005, pp. 11-18, 2005.
3. Corbett, A., Wagner, A., Chao, C., Lesgold, S., Stevens, S., Ulrich, H. Student Questions in a Classroom Evaluation of the ALPS Learning Environment. C-K. Looi, G. McCalla, B. Bredeweg, Breuker, J. (eds) Proc. 12th Int. Conf. Artificial Intelligence in Education, pp. 780-782, 2005.
4. Elmasri, R., Navathe, S. Fundamentals of Database Systems. Addison Wesley, 2004.
5. Mitrovic, A., Mayo, M., Suraweera, P., Martin, B. Constraint-based tutors: a success story. In: L. Monostori, J. Vancza and M. Ali (eds), Proc. 14th Int. Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems IEA/AIE-2001, Springer-Verlag Berlin Heidelberg LNAI 2070, pp. 931-940, 2001.
6. Mitrovic, A., Suraweera, P., Martin, B., Weerasinghe, A. (2004) DB-suite: Experiences with Three Intelligent, Web-based Database Tutors. Journal of Interactive Learning Research (JILR), vol. 15, no. 4, pp. 409-432, 2004.
7. Salton, G., Buckley, C., Term Weighting Approaches in Automatic Text Retrieval. Tech. Report #87-881 Computer Science Dept, Cornell University, 1987.

E.1 Presented Poster

Responding to Free-form Student Questions in ERM-Tutor



Intelligent Tutoring Systems provide feedback on students' actions, but students do not always understand the feedback they receive. Therefore, it would be beneficial for students to be able to ask free-form questions at any time.

ERM-Tutor teaches the ER-to-relational seven-step mapping algorithm. ERM-Tutor is a supplement to classroom teaching. The system analyzes a student's solutions, and provides the student with tailored feedback messages based on his/her knowledge.

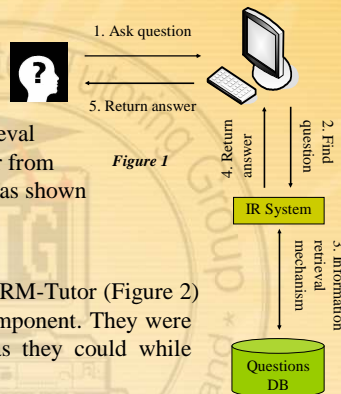
Method

We added a new question-asking module to ERM-Tutor. It stored a set of 98 predefined questions and their answers in a database.

These questions can be categorized into:

- ❖ interface usage ("What does the button *Check Step* do?")
- ❖ definitions of terms ("What is a foreign key?")
- ❖ diagram notations ("How is an attribute represented in the ER-diagram?")
- ❖ mapping regulations ("How is a relationship mapped?")
- ❖ deeper questions ("Why are the steps arranged in this order?")

When a student asks a question the system uses the TFIDF (Term Frequency Inverse Document Frequency) vector weighting scheme as the information retrieval mechanism to grab the appropriate answer from the database, and returns it to the student, as shown in Figure 1.



We invited 89 volunteer students to use ERM-Tutor (Figure 2) with the newly added question-asking component. They were encouraged to ask as many questions as they could while solving the problems.

The participants worked individually, solving problems at their own pace. We recorded their actions in log files.

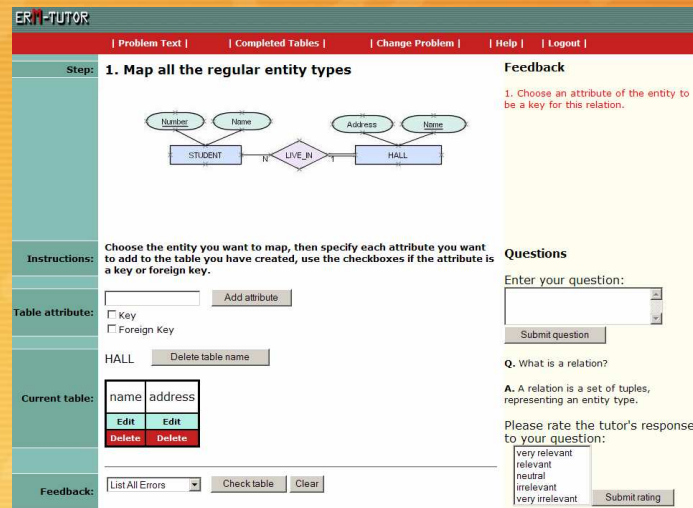


Figure 2

Results

29 students logged into ERM-Tutor at least once.

The average interaction time was under one hour (mean=54min, sd=63min), ranging from several minutes to 4.5 hours over several weeks. The number of sessions ranged from one to four (mean=1.67, sd=0.96). On average, students attempted 4.6 problems and completed 25% of them.

Only eight students asked questions, with a total of 24 questions submitted. The number of questions per student ranged from one to five. The questions can be categorized into:

- ❖ 50% task-focused (e.g. How could I solve this?)
- ❖ 8% definition-focused (e.g. What does *multivalued* mean?)
- ❖ 42% phatically-focused (e.g. What is your name?)

Conclusions

Our preliminary study showed some evidence that students welcome the idea of asking free-form questions and confirmed the need for eliciting deeper questions.



"Asking questions is a very good way to find out about something"

- Kermit the Frog

Nancy Milik,
Melinda Marshall,
Antonija Mitrovic

APPENDIX F

NZCSRSC 2007 Paper

This paper was presented at the 5th New Zealand Computer Science Research Student Conference held in Hamilton, New Zealand, April 2007.

Milik, N., Mitrović, A., and Grimley, M. (2007a). Fitting spatial ability and free-formed questions into Intelligent Tutoring Systems development. In *Proceedings of New Zealand Computer Science Research Student Conference*, Hamilton

Fitting spatial ability and free-formed questions into Intelligent Tutoring Systems development

Nancy MILIK, Antonija MITROVIC and Michael GRIMLEY

Intelligent Computer Tutoring Group
Department of Computer Science and Software Engineering
University of Canterbury, New Zealand
{nmi14,tanja}@cosc.canterbury.ac.nz

Abstract. Building effective learning tools is an art that can only be perfected by a great deal of explorations involving the tools' audience: the learners. This paper focuses on accounting for the learners' spatial ability as well as providing an additional help channel in Intelligent Tutoring Systems. We modified ERM-Tutor, a constraint-based tutor that teaches logical database design, to provide not only textual feedback messages, but also messages containing combinations of text and pictures, in accordance with the multimedia theory of learning [1]. We also added a question-asking module which enables students to ask free-form questions. Results of preliminary studies performed show a promising indication for further explorations. We plan to use these results as the basis for another evaluation study in early 2007.

1. Introduction

In today's society, people constantly face the challenge of acquiring new skills and knowledge. Rapid and widespread developments in technology have made information available and easily accessible more than ever before. Such ease of access alone however, does not necessarily result in a better learning gain for students. Although e-learning tools, such as WebCT [2], are becoming more popular in educational institutions, they do not effectively support learning. While they make it easier for teachers to present instructional material and carry out some administrative tasks, they do not provide students with individualised feedback based on their performance, which is crucial for successful learning. An effective solution that provides adaptive pedagogical assistance for each student is Intelligent Tutoring Systems (ITSs).

ITSs are interactive computerised tutors that provide an environment where students carry out problem-solving activities, and receive feedback on their actions. As the student interacts with the system, it tracks their behaviour and produces/maintains a model of the student's state. This model is used in adapting the environment towards the needs, knowledge, learning abilities and preferences of the student. This includes decisions about the timing and content of teaching actions and feedback to be presented to each individual student. Such adaptations have been shown to result in significant improvement over simplistic e-learning tools, especially

in fields that require practical proficiency [3, 4]. Nevertheless, this is still a growing discipline that is utilising findings in educational and psychological theories and new developments in Artificial Intelligence and software and hardware technology.

In this paper, we describe a master's project which focuses on enhancing ERM-Tutor, a constraint-based ITS that teaches logical database design (i.e. the algorithm for mapping conceptual to logical database schemas), by (a) adapting the feedback presentation mode towards the student's spatial ability and (b) incorporating a module where students are able to ask for additional clarifications.

The next section presents an overview of ERM-tutor, followed by an overview of spatial ability in Section 3 and the question-asking module in Section 4. We then describe the preliminary studies and the results obtained in Section 5, followed by conclusions and future work in the final section.

2. ERM-Tutor

Constraint-based tutors enhance learning in a variety of domains, such as database querying (SQL-Tutor [5]), conceptual database design (ER-Tutor [6]) and data normalisation (NORMIT [7]). ERM-Tutor [8] is another web-based tutor in which students practice the 7-step algorithm for mapping conceptual database schemas (i.e. ER diagrams) into relational schemas. Each step in the algorithm maps one ER concept by either creating a new relation or altering previously created relations by adding foreign keys and attributes.

The interface (Fig. 1) enables students to view problems, work on their solutions and receive feedback. The problem-solving area is the main part of the page, and its general layout is the same for all steps. The student creates or alters one relation at a time. Each step of the algorithm is broken into subtasks. For example, in step one, the student maps one regular entity type at a time, and the system checks the resulting relation before moving on to the next entity type. Fig. 1 illustrates a situation when the student has mapped the MEETING weak entity type, and has specified a relation (with the same name) with three attributes (*timing*, *id* and *description*). For each attribute, the student can specify whether it is a primary and/or foreign key. When the student completes the relation, he/she can request the system to check the solution. If there are any mistakes in the solution, ERM-Tutor provides feedback. In Fig. 1, the system informs the student that there are some missing attributes as well as a foreign key from the owner for the MEETING relation. If the solution is correct, the student can move on to the next entity type, or to the following step of the algorithm.

3. Spatial Ability

Spatial ability is a psychometric construct essential to activities related to spatial reasoning, such as the ability to manipulate images or spatial patterns into other arrangements [9]. Learners with high spatial abilities perform better with graphic or spatially-oriented content than those with low spatial ability.

Psychometric tests used for determining spatial ability typically consist of paper-and-pencil tasks requiring inspecting, imagining or mentally transforming shapes or objects at the *figural* scale of space. These tests do not provide a discrete value on the spatial ability scale, but rather a relative position within a sample group that determines high or low classifications. We explored short versions of two tests from the battery of cognitive tests [10]: a ten-item Paper Folding Test intended to evaluate a component of spatial ability called visualisation, and an eighty-item mental Card Rotation Test which evaluates spatial orientation. Each test has a three-minute time limit and is suitable for ages 13-18.

It is worth noting, however, that a low spatial ability score is not a deficit, and there is even evidence that it can be improved through training and practice [11]. Nevertheless, changing ITSs to accommodate low spatial ability learners, rather than providing a spatial ability training environment, could be more practical and beneficial for the system/domain's problem solving task. That is, learners with different spatial abilities should receive different types of content.

The theory of multimedia learning [1] presents a number of principles for customising instructional content towards individuals' spatial ability. Mayer defines multimedia as the presentation of material using both words and pictures, and proposes that presenting verbal explanations alone in instructional situations is less conducive to learning for some students than presenting verbal explanations in conjunction with pictures [12]. Subsequently, he defines a multimedia instructional message as communication that makes use of our dual learning channel [13] which is intended to foster learning.

The screenshot shows the ERM-TUTOR interface. At the top, there is a navigation bar with links: Problem Text, Completed Tables, Change Problem, Help, and Logout. The main content area is divided into several sections:

- Step:** 2. Map all the weak entity types
- Diagram:** An Entity-Relationship diagram showing three entities: LECTURER, STUDENT, and MEETING. LECTURER and STUDENT are connected to MEETING by a relationship named MEETS. LECTURER has an attribute 'id'. STUDENT has an attribute 'student_number'. MEETING has attributes 'description' and 'topic'.
- Instructions:** Choose the entity you want to map, then specify each attribute you want to add to the table you have created, use the checkboxes if the attribute is a key or foreign key.
- Table attribute:** A form with an 'Add attribute' button and checkboxes for 'Key' and 'Foreign Key'.
- Current table:** A table with columns 'timing', 'id', and 'description'. Below the table are 'Edit' and 'Delete' buttons for each cell.
- Relationship:** A form with 'MEETS' and a 'Delete relationship' button.
- Feedback:** A section with two numbered points:
 1. There are some missing attributes from your table. Check that you have correctly identified and spelt all the attributes for the table.
 2. For this step you need to specify all the foreign keys from the owner entities.
- Questions:** A section with a text input field and a 'Submit question' button.
- Feedback (bottom):** A section with a 'List All Errors' dropdown, 'Check table', and 'Clear' buttons.

Fig. 1. Screenshot of ERM-Tutor

Fig. 2 shows a representation of the dual channel theory. One channel is dedicated to processing words, whether printed or spoken, and the other is for processing pictorial forms. Based on this assumption, along with the assumptions that each channel has a limited capacity and require active processing, Mayer defines the Cognitive Theory of Multimedia Learning [1]. The theory states that learning occurs when learners attend to relevant incoming information (sensory memory), select and organise important information and integrate it with their prior knowledge (working memory) into mental representations (long-term memory). Mayer argues that making use of both visual and auditory channels when presenting learning instructions aids in deep, or meaningful, learning, indicated by good retention and transfer performance. His rationale is that when presenting a message combining an image and text, the information is effectively being perceived and processed twice (once through each channel). Moreover, the words and pictures complement each other, aiding the learner to mentally encode and integrate the information.

Mayer defines a number of principles for designers of instructional environments to follow in order to make the maximum use of the learners' dual channels. The principle that is of most interest to us however, is the individual differences principle, which states that "[multimedia] design effects are stronger for low-knowledge learners than for high-knowledge learners and for high spatial learners rather than from low spatial learners" (p. 161) [1]. This is because high-knowledge learners are able to use their prior knowledge to compensate for the cognitive processing needed to integrate the information received by the dual-channel. On the other hand, low-spatial learners must devote so much cognitive capacity to mentally integrate the information. Therefore, it is the combination of the learners' spatial ability and level of knowledge that influences their meaningful/deep learning.

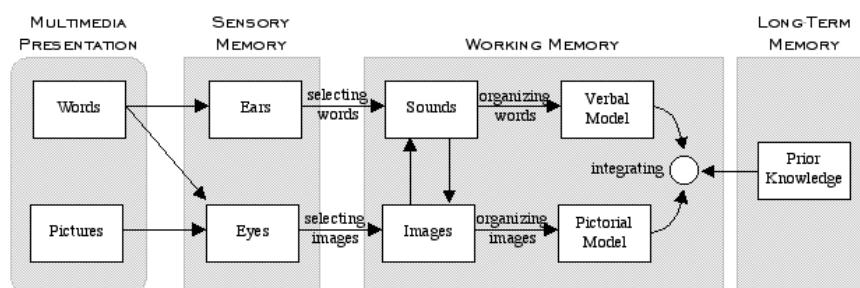


Fig. 2. Information processing via dual learning channels (Figure 3.2 from [1])

Influenced by Mayer's work, we created a new version of the system. The original ERM-Tutor only provides text-based feedback. Following the multimedia learning theory, we decided to incorporate a pictorial aspect in the messages; for each feedback message, we created a graphically annotated version. To make the original and the newly created messages comparable, we kept the text identical in both versions. The only difference is the addition of a pictorial representation in the new version. Fig. 3 shows the multimedia (text and picture) version of the second feedback message given in Fig. 1. A total of 112 images were created, each corresponding to a

single feedback message. In addition, ERM-Tutor was modified to cater for both versions of feedback and prepared for an evaluation study described in Section 5.

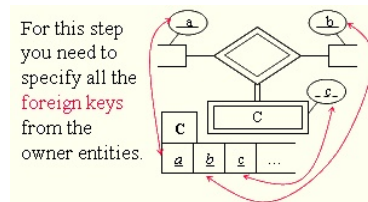


Fig. 3. An example feedback message in multimedia representation

4. Question Asking Module

ITSs provide feedback on students' actions, but students do not always fully understand the feedback they receive. Therefore, it would be beneficial for students to be able to ask questions at any time. Although researchers in cognitive science and education have reported learning benefits for environments that encourage students to generate questions [14], question-asking is still a young concept in the ITS field (e.g. [15]). Question generation is believed to be a primary attribute of active learning which reveals how deeply the learner has mastered the material and even shifts the student's goals from performance toward learning orientation [16].

In this light, we added a question-asking module to ERM-Tutor. We defined 98 distinct questions, based on our experiences in teaching the mapping algorithm and our experience with other constraint-based tutors. These questions can be categorised into interface usage ("What does the button Check Step do?"), definitions of terms ("What is a foreign key?"), diagram notations ("How is an attribute represented in the ER-diagram?"), mapping regulations ("How is a relationship mapped?"), and deeper questions ("Why are the steps arranged in this order?"). Each question is stored along with its textual answer. The question database additionally includes a number of repeated questions that are phrased differently, resulting in a total of 182 questions.

The TFIDF (Term Frequency Inverse Document Frequency) vector weighting scheme [17] was chosen as the information retrieval mechanism. In our system, the questions are read from the database and separated into words. The weight of each question and word is calculated, and words are indexed in a hash table. When the student asks a question, the same calculations are applied to the query string: it is also broken-up into words and their weights are calculated. Each question is then allocated a query weight. Finally, the answer corresponding to the question with the highest query weight is returned to the student. To evaluate the subjective relevance of the answers returned, students are encouraged to submit their ratings of the answers; however, the system does not enforce it to avoid mode errors and distractions from the problem solving task.

5. Preliminary Studies

We performed two preliminary studies with students enrolled in an introductory database course at Canterbury University in November 2005 and March 2006. The aim of the first study was to investigate the usage of free-form questions. 29 students logged into ERM-Tutor at least once, but five students used it for less than two minutes and so their logs were excluded from analyses. The average interaction time was under one hour (54min, $sd=63min$), ranging from several minutes to 4.5 hours over several weeks. The number of sessions ranged from one to four (mean=1.67, $sd=0.96$). On average, students attempted 4.6 problems and completed 25% of them.

Only eight students asked questions, with a total of 24 questions submitted. The number of questions per student ranged from one to five. The questions can be categorised into task-focused (50%), definition-focused (8%) and phatic questions (42%). Task-focused questions ask directly for help solving the problem (e.g. "How could I solve this table?"). For instance, three students copied the feedback messages, added a question mark at the end or a "How to" at the start, and submitted them as the questions. Definition-focused questions ask for definition of terms. There were only two such questions submitted: "What is foreign key?" and "What is multivalued?" Phatic questions establish a sense of social mood (e.g. "What is your name?", "How are you?" and "How do you answer questions?"). Excluding phatic questions, 14 questions were relevant for students' actions. Five of these questions were answered correctly, and for two of these, the students specified highest relevance. The answer could not be found for one question. The remaining questions received answers which were related to the query, but not useful to students. This happened when the students did not formulate questions well, but instead copied a part of the feedback message then added a question mark at the end (e.g. "Make sure the relationship is 1:1?").

Our hypothesis for the second study was that students with a high spatial ability level will benefit more from multimedia feedback than students with a low spatial ability, given the same background knowledge. As each student's spatial ability level (either high or low, as opposed to the actual value) is determined relatively to the sample group, it was decided to compute it in a post-hoc manner. The students were randomly allocated to one version of the system, providing either textual or multimedia feedback. The assumption was that each group would ultimately include students with high and low spatial abilities. Therefore, the experiment allows for a 2x2 comparison: textual messages for high (TH) and low spatial ability students (TL), and multimedia messages for high (MH) and low spatial ability students (ML).

The study was conducted in two sessions of scheduled labs on ER mapping, straight after students had attended lectures on the topic. Each participant attended one of the sessions, and worked with ERM-Tutor individually, solving problems at his/her own pace. At the start of a session, the students were given an information sheet describing the study, a consent form, and a pre-test on paper (with the maximal score of 4). The average score on the pre-test for all students was 2.23 ($sd=1.15$). To make the results of the pre-test and post-test comparable, two tests were used; students in the first session used version A as the pre-test and version B as the post-test and students in the second session used the reverse.

When a student logged onto the system, he/she was presented with a set of instructions explaining the two spatial ability tests, with sample problems.

Additionally, for each test, they were asked to rate their own ability on a scale of 1 to 5 before sitting the tests. They had three minutes to solve the problems in each test. Once the spatial tests were completed, or their time was up, the students were randomly assigned to one of the two versions of the system. At the end of the session, students were asked to fill in a post-test and a questionnaire about the system. Finally, the students were encouraged to use the system at any time until the end of the course.

Out of 74 students enrolled in the course, 55 students participated and completed both spatial tests. Before completing each test, the students were asked to rate their own ability of the spatial skill. For the paper fold test, the average rating of 6.62 was close to the actual test score of 6.89 out of a possible 10. The students' personal rating for the card rotation test had a mean of 7.6. As explained, the total possible score for the card rotation test is 80. We computed the total test score by dividing the score by 8, thus giving a range of 1–10. The students scored a mean of 6.43. To compute the spatial ability of each student, we added both test scores giving a possible range of 1–20. Using a median split, a total of 28 students scored above the median and were classified as high spatial, and the other 27 students were classified as low spatial.

The pre-test was collected at the start of the session, while the post-test was administered after two hours of interaction. Only 13 students completed both tests, scoring a mean of 1.92 (sd=1.04) on the pre-test, and 3 (sd=1.15) on the post-test, resulting in significant improvement of their performance ($t=3.09$, $p < 0.001$). The scores for the four groups are given in Table 1. These preliminary results (although with small numbers) seem to refute Mayer's prediction that high spatial learners will benefit most from multimedia messages. However, it does seem that the subsets of participants from the TH and MH groups who completed both tests started with higher pre-existing knowledge, and therefore Mayer's individual differences principle may be more pertinent in that low knowledge individuals will have a higher gain. Of course, with such low numbers of submitted tests, we might expect a lot of error and therefore further investigation is warranted.

The system recorded all student actions in logs. Due to a technical problem however, the logs from the first session could not be used. A total of 17 students used the system for more than 10 minutes. On average, students attempted 3.4 problems and completed 33% of them. The numbers of valid logs in each condition are too small, and we are therefore unable to closely analyse the effect of the students' spatial ability on their performance. Analyses of the questionnaires showed that students who received multimedia feedback rated the overall quality of the feedback messages 25% higher (mean of 4 out of a possible 5) than those who received textual feedback (mean of 3 out of a possible 5). There were also some encouraging comments from the questionnaires submitted in the second session. Students liked the system, and appreciated the problem solving environment provided to solve the problems.

Table 1. Pre/post test results for the students who sat both tests

| Feedback | Low Spatial | | | High Spatial | | |
|------------|-------------|-----------|-----------|--------------|-----------|------------|
| | No. | Pre-test | Post-test | No. | Pre-test | Post-test |
| Textual | TL: 4 | 1.5 (1) | 3.5 (0.6) | TH: 3 | 2 (1) | 2.3 (0.6) |
| Multimedia | ML: 2 | 1.5 (0.7) | 3.5 (0.7) | MH: 4 | 2.5 (1.3) | 2.75 (1.9) |

6. Conclusion

Rapid and widespread development of computerised learning tools have proven the need for further exploration of the learners' personal characteristics in order to maximise the use of the current technology. In particular, this paper has looked at the potential of accounting for spatial ability and providing an additional help channel in ERM-Tutor; a constraint-based tutor that teaches the procedural task of mapping ER diagrams into relational schemas. We presented results from two preliminary studies.

The first study, which investigated the question-asking module, showed some evidence that students welcome the idea of asking free-form questions and confirmed the need for eliciting deeper questions. The results from the second study, which evaluated the effectiveness of the type of content representation (text only vs. multimedia) to the learner's spatial ability level, show an overall improvement in the students' domain knowledge level after using ERM-Tutor for the duration of the study (2 hours). Although the amount of data collected was small, the results show a promising indication for further explorations. We plan to use these studies as the basis for another evaluation study testing the same hypotheses in early 2007.

References

1. Mayer, R., *Multi-media Learning*. 2001: University of California.
2. Lu, J., C.S. Yu, and C. Liu, *Learning style, learning patterns, and learning performance in a WebCT-based MIS course*. Information and Management, 2003. 40(6): p. 497-507.
3. Koedinger, K.R., et al., *Intelligent tutoring goes to school in the big city*. International Journal of Artificial Intelligence in Education, 1997. 8(1): p. 30-43.
4. Mitrovic, A., et al., *Constraint-Based Tutors: A Success Story*. IEA/AIE, 2001: p. 931-940.
5. Mitrovic, A., *A Knowledge-Based Teaching System for SQL*. MEDIA, 1998: p. 1027-1032.
6. Suraweera, P., *An Intelligent Tutoring System for Entity Relationship Modelling*. International Journal of Artificial Intelligence in Education, 2004. 14(3): p. 375-417.
7. Mitrovic, A., *NORMIT, a Web-enabled tutor for database normalization*. Proc. ICCE, 2002: p. 1276-1280.
8. Milik, N., M. Marshall, and A. Mitrovic. *Responding to Free-form Student Questions in ERM-Tutor*. in *Proc. Intelligent Tutoring Systems*. 2006. Taiwan: Springer.
9. Carroll, J., *Human cognitive abilities*. 1993, England: Cambridge University Press.
10. Ekstrom, R., J. French, and H. Harman, *Manual for kit of factor referenced cognitive tests*. 1997: Princeton.
11. Vicente, K.J. and R.C. Williges, *Accommodating individual differences in searching a hierarchical file system*. Journal of Man-Machine Studies, 1988. 29(6): p. 647-668.
12. Mayer, R., *Multimedia learning: Are we asking the right questions?* Educational Psychologist, 1997. 32: p. 1-19.
13. Paivio, A., *Mental Representations: A Dual Coding Approach*. 1986: Oxford University.
14. Beck, I.L., et al., *Questioning the author: An approach for enhancing student engagement with text*. Delaware: International Reading Association, 1997.
15. Anthony, L., et al. *Student Question-Asking Patterns in an Intelligent Algebra Tutor*. in *Proc. Intelligent Tutoring Systems*. 2004. Brazil: Springer.
16. Graesser, A.C. and N.K. Person, *Question asking during tutoring*. American Educational Research Journal, 1994. 31: p. 104-137.
17. Salton, G. and C. Buckley, *Term-weighting approaches in automatic text retrieval*. Information Processing and Management: an International Journal, 1988. 24(5): p. 513-523.

APPENDIX G

AIED 2007 Short Paper

The following paper was presented at the 13th International Conference on Artificial Intelligence in Education held in Los Angeles, United States, July 2007.

Milik, N., Mitrović, A., and Grimley, M. (2007b). Fitting Spatial Ability into Intelligent Tutoring Systems Development. In *Proceedings of 13th International Conference on Artificial Intelligence in Education*, pages 617–619, Los Angeles. IOS Press

Fitting Spatial Ability into Intelligent Tutoring Systems Development

Nancy MILIK¹, Antonija MITROVIC¹ and Michael GRIMLEY²

¹*Department of Computer Science and Software Engineering*

²*Department of Education*

University of Canterbury, New Zealand

Abstract. Building effective learning environments is an art that can only be perfected by a great deal of explorations involving the environments' audience: the learners. This paper focuses on taking into account the learners' spatial ability into the development of Intelligent Tutoring Systems. We modified ERM-Tutor, a constraint-based tutor that teaches logical database design, to provide not only textual feedback messages, but also messages containing combinations of text and pictures, in accordance with the multimedia theory of learning [1]. Results of a preliminary study performed show a promising indication for further explorations. We plan to use these results as the basis for another evaluation study in early 2007.

Introduction

Intelligent Tutoring Systems (ITSs) are effective learning tools due to the adaptive pedagogical assistance they provide. Students differ in their capabilities for learning and processing information. This paper describes a project which focuses on spatial ability, a psychometric construct essential to activities related to spatial reasoning, such as the ability to manipulate images or spatial patterns into other arrangements [2]. Learners with high spatial abilities perform better with graphic or spatially-oriented content than those with low spatial ability. It is worth noting, however, that a low spatial ability score is not a deficit; there is evidence that it can be improved through training and practice [3]. Nevertheless, enhancing ITSs to accommodate low spatial ability learners could be beneficial for their problem solving skills. As a consequence, learners with different spatial abilities should receive different types of content.

The theory of multimedia learning states that "*[multimedia] design effects are stronger for low-knowledge learners than for high-knowledge learners and for high spatial learners rather than from low spatial learners*" (p. 161) [1]. Low-spatial learners must devote much of their cognitive capacity to process multimedia information. High-knowledge and low-spatial learners are able to use their prior knowledge to compensate for the cognitive load needed to integrate the information received by the dual-channel. Therefore, it is the combination of the learners' spatial ability and level of knowledge that influences their meaningful/deep learning.

We present an approach to support the learners' spatial ability in ERM-Tutor [4], a constraint-based ITS that teaches logical database design (i.e. the algorithm for mapping conceptual to logical database schemas). The next section presents the modifications made in this project. We then describe the preliminary study and the results obtained, followed by conclusions and future work in the final section.

1. Spatial Ability Support in ERM-Tutor

Influenced by Mayer's work, we created a new version of the system. The original ERM-Tutor provides only text-based feedback. Following the multimedia learning theory, we decided to incorporate a pictorial aspect in the messages; for each feedback message, we created a graphically annotated (multimedia) version.

Each feedback message in ERM-Tutor is associated with a constraint. In other words, each constraint has a feedback message which is returned when the constraint is violated. Consequently, each message provides a hint on how to satisfy its particular constraint. To make the original and the newly created messages comparable, we kept the text identical in both versions. The only difference is the addition of a pictorial representation in the new version.

Figure 1 shows an example message in multimedia representation. A total of 112 images were created, each corresponding to a single feedback message. In addition, ERM-Tutor was modified to cater for both versions of feedback and prepared for an evaluation study described in the following section.

We also explored two cognitive tests for testing spatial ability [5]: a ten-item Paper Folding Test intended to evaluate a component of spatial ability called visualization, and an eighty-item mental Card Rotation Test which evaluates spatial orientation. Each test has a three-minute time limit and is suitable for ages 13-18.

2. Experiment

We performed a preliminary study with students enrolled in an introductory database course at Canterbury University in March 2006. Our hypothesis is that students with a high spatial ability level will benefit more from multimedia feedback than students with a low spatial ability, given the same background knowledge. As each student's spatial ability level (either high or low, as opposed to the actual value) is determined relatively to the sample group, it was decided to compute it post-hoc. The students were randomly allocated to one version of the system, providing either textual or multimedia feedback. The assumption was that each group would ultimately include students with high and low spatial abilities. Therefore, the experiment allows for a 2x2 comparison: textual messages for high (TH) and low spatial ability students (TL), and multimedia messages for high (MH) and low spatial ability students (ML).

The study was conducted in two two-hour sessions of scheduled labs on ER mapping, straight after students had attended lectures on the topic. Each participant attended one of the sessions, and worked with ERM-Tutor individually, solving problems at his/her own pace. The pre-test was collected at the start of the session, while the post-test was administered after two hours of interaction.

55 students participated and completed both spatial tests. The test score for the paper fold test was 6.89 out of a possible 10. The total possible score for the card rotation test is 80. We computed the total test score by dividing the score by 8, thus giving a range of 1-10. The students scored a mean of 6.43. To compute the spatial

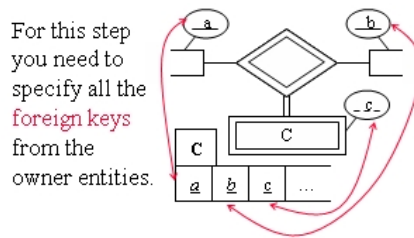


Figure 1. An example feedback message in multimedia representation

ability of each student, we added both test scores giving a possible range of 1–20. Using a median split, a total of 28 students scored above the median and were classified as high spatial, and the other 27 students were classified as low spatial.

The system recorded all student actions in logs. Due to a technical problem however, the logs from the first session could not be used. A total of 17 students used the system for more than 10 minutes. On average, students attempted 3.4 problems and completed 33% of them.

Only 13 students completed both tests, scoring a mean of 1.92 (sd = 1.04) on the pre-test, and 3 (sd = 1.15) on the post-test, resulting in significant improvement of their performance ($t=3.09$, $p < 0.001$). The scores for the four groups are given in Table 1. These preliminary results (although with small numbers) seem to refute Mayer’s prediction that high spatial learners will benefit most from multimedia messages. However, it seems that the participants from the TH and MH groups who completed both tests started with higher pre-existing knowledge, and therefore the theory may be more pertinent in that low knowledge individuals will have a higher gain.

The numbers of valid logs in each condition are too small, and we are therefore unable to closely analyze the effect of the students’ spatial ability on their performance. Analyses of the questionnaires showed that students who received multimedia feedback rated the overall quality of the feedback messages 25% higher (mean of 4 out of a possible 5) than those who received textual feedback (mean of 3 out of a possible 5).

Table 1. Pre/post test results for the students who sat both tests

| Feedback | Low spatial | | | High spatial | | |
|------------|-------------|-----------|-----------|--------------|-----------|------------|
| | No. | Pre-test | Post-test | No. | Pre-test | Post-test |
| Textual | TL: 4 | 1.5 (1) | 3.5 (0.6) | TH: 3 | 2 (1) | 2.3 (0.6) |
| Multimedia | ML: 2 | 1.5 (0.7) | 3.5 (0.7) | MH: 4 | 2.5 (1.3) | 2.75 (1.9) |

3. Conclusions

This paper has looked at the potential of accounting for spatial ability in ERM-Tutor. Results from a preliminary study show an overall improvement in the students’ domain knowledge level after two hours of interaction. We could not however report any findings on the correlation between spatial ability, content representation and the learning experience due to a technical problem. Although the amount of data collected was small, the results show a promising indication for further explorations. We plan to use this study as the basis for another evaluation study testing the same hypothesis.

References

1. Mayer, R., *Multi-media Learning*. 2001: University of California.
2. Carroll, J., *Human cognitive abilities*. 1993, England: Cambridge University Press.
3. Baenninger, M. and N. Newcombe, *The role of experience in spatial test performance: A meta-analysis*. *Sex Roles*, 1989. 20(5): p. 327-344.
4. Milik, N., M. Marshall, and A. Mitrovic. *Responding to Free-form Student Questions in ERM-Tutor*. in *Proceedings of 8th International Conference on Intelligent Tutoring Systems*. 2006. Taiwan: Springer.
5. Ekstrom, R., J. French, and H. Harman, *Manual for kit of factor referenced cognitive tests*. 1997: Princeton.

G.1 Presented Poster

Fitting Spatial Ability into ITS Development

Domain:

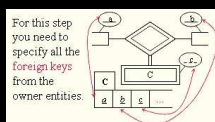
ERM-Tutor is a constraint-based ITS that teaches logical database design (ER-to-relational 7-step mapping algorithm) in a multimedia environment.

Facts:

- Multimedia learners need to be able to form, hold, and use mental images.
- Some of the skills required to engage fully in multimedia learning resemble the definition of spatial ability.
- Spatial ability is the ability to engage in spatial reasoning/cognition.
 - E.g. manipulate images or spatial patterns into other arrangements.
- Learners with high spatial abilities perform better with graphic or spatially-oriented content than those with low spatial ability.

Decision:

- Incorporate a pictorial aspect in the messages.
 - For each feedback message, we created a graphically annotated (multimedia) version.



Problem:

What does the feedback mean?



A Solution:

Customised feedback messages

Experiment:

- Aim: investigate differences in students' performance.
- Invited 74 volunteer students.
 - Presented with pre- & post- tests, spatial tests, questionnaire.
 - Randomly assigned to feedback mode.
 - Worked individually, solving problems at their own pace.

Results:

- 55 students completed the spatial tests.
 - Score: mean 12.03, s.d. 3.06, median 11.90.
- Only 13 students completed both pre- & post- tests.
 - Significant improvement for all students ($t=3.09$, $p < 0.001$).
 - Numbers are small for comparing groups.
- Positive subjective results about ERM-Tutor.

Conclusions:

- Our study shows:
 - An overall improvement in the students' domain knowledge.
 - A promising indication for further explorations.
- We plan to use this study as the basis for another evaluation study testing the same hypothesis.

Nancy Milik, Antonija Mitrović and Michael Grimley

APPENDIX H

College Poster Competition

The following poster was entered in the College of Engineering 2006 Research Poster Competition.

Ask and You Shall Be Answered

while being tutored

“Asking questions is a very good way to find out about something”

– Kermit the Frog

Background

Often when we are learning new skills, whether it's algebra, how to play a musical instrument or how to drive, we need to be tutored.

The tutor's job is to give us individual instruction and personalised feedback about our performance; give us praise, tell us what needs to be improved and hints on how to improve it.

Human one-to-one is the most effective form of tutoring. However, it is nearly impossible for the majority of students to be privately tutored.

This is why Intelligent Tutoring Systems (ITSs) are being developed. They are computer-based educational programs that provide problems for students to solve while mimicking the human tutoring behaviour.

We learn better when we ask questions that clarify our understanding. Being able to ask questions is vital because questions:

- ❖ point to holes in our memory structures
- ❖ provide the starting point for integrating new information
- ❖ tie old information together in new ways
- ❖ correct faulty generalisations

Do we really learn better if we ask questions?

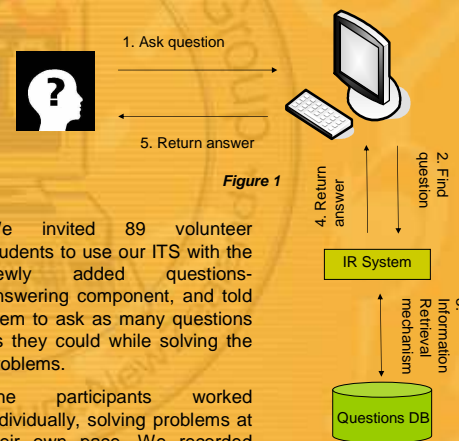
To answer this we added a new feature to one of our ITSs. The component enables students to ask questions in an ITS. We then put our new component to the test.

Hypothesis: Students who ask questions and receive answers while using an ITS will understand and perform better.

Method

The new questions-answering component stored a set of 98 predefined questions and their answers in a database.

When a student asks a question the system uses an *Information Retrieval* (IR) method to grab the appropriate answer from the database, and returns it to the student, as shown in Figure 1.



We invited 89 volunteer students to use our ITS with the newly added questions-answering component, and told them to ask as many questions as they could while solving the problems.

The participants worked individually, solving problems at their own pace. We recorded their actions in log files.

Results

The participants used the system for an average of just under an hour. The questions they asked the system can be categorised as follows:

- ❖ 50% Task-focused (e.g. How could I solve this?)
- ❖ 8% Definition-focused (e.g. What does x mean?)
- ❖ 42% Socially-focused (e.g. What is your name?)

Conclusions

ITSs are extending the tertiary education system especially in the area of e-learning. They have a great potential to support learning in all instructional areas and levels.

Our initial study showed some evidence that students welcome the idea of asking questions while using an ITS.

We are currently investigating various techniques to encourage students to use the questions-answering component, such as prompting students to ask more questions and even suggesting a question to be asked based on their actions.

Presented by: Nancy Milik
Supervisor: Assoc Prof. Tanja Mitrović