UNIVERSITY OF CANTERBURY

Supporting Learning in Intelligent Tutoring Systems with Motivational Strategies

A thesis submitted in partial fulfilment of the requirements for the Degree of Doctor of Philosophy in Computer Science

by Faiza Tahir

supervised by

Professor Antonija Mitrovic

Dr. Valerie Sotardi

2021

Table of Contents

List of Figures

List of Tables

ACKNOWLEDGEMENTS

First, I would like to thank ALLAH for giving me this opportunity and supporting me throughout this journey and reaching me to the finishing line.

I am very grateful to my supervisors, Professors Antonija Mitrovic and Dr. Valerie Sotardi, for all their encouragement, advice, support and time in helping me complete this thesis. I have greatly appreciated Tanja's guidance, thoughtful insights and extraordinary ability to always find the strengths in my ideas. Valerie has been my most enthusiastic supporter and her optimistic outlook has always made me feel like I was on the right track. Together, they have been the perfect advisors to this research, and I feel very fortunate to have worked under their supervision. Importantly, I would also like to acknowledge their incredible patience on my weaknesses. Tanja and Valerie, thank you.

I could not have completed this thesis without the academic, moral and financial support of the Department of Computer Science and Software Engineering and College of Engineering. I have benefited enormously from being part of such a positive and collegial environment and I am very grateful for the encouragement and advice of my colleagues Jay, Geela, Negar, and Ja'afaru. I would like to express my thanks to the academic head during the time I have worked on this thesis, Professor Richard Green, for his understanding and support.

Finally, and most importantly, thank you to my parents and husband especially my daughter Sarah for letting me complete my thesis. They have been my inspiration; they have given me the confidence to persevere, and they have provided me with reassurance when I needed it. I cannot thank them enough for the constant love and support, and for being so patient with me!

[Faiza Tahir]

Abstract

Motivation and affect detection are prominent yet challenging areas of research in the field of Intelligent Tutoring Systems (ITSs). Devising strategies to engage learners and motivate them to practice regularly are of great interest to researchers. In the learning and education domain, where students use ITSs regularly, motivating them to engage with the system effectively may lead to higher learning outcomes. Therefore, developing an ITS which provides a complete learning experience to students by catering to their cognitive, affective, metacognitive, and motivational needs is an ambitious yet promising area of research. This dissertation is the first step towards this goal in the context of SQL-Tutor, a mature ITS for tutoring SQL.

In this research project, I have conducted a series of studies to detect and evaluate learners' affective states and employed various strategies for increasing motivation and engagement to improve learning from SQL-Tutor. Firstly, I established the reliability of iMotions to correctly identify learners' emotions and found that worked examples alleviated learners' frustration while solving problems with SQL-Tutor. Gamification is introduced as a motivational strategy to persuade learners to practice with the system. Gamification has emerged as a strong engagement and motivation strategy in learning environments for young learners. I evaluated the effects of gamified SQL-Tutor on undergraduate students and found that gamification indirectly improved learning by influencing learners' time on task. It helped students by increasing their motivation which produce similar effects as intrinsically motivated students. Additionally, prior knowledge, gamification experience, and interest in the topic moderated the effects of gamification.

Lastly, self-regulated learning support is presented as another strategy to affect learners' internal motivation and skills. The support provided in the form of interventions improved students' learning outcomes. Additionally, the learners' challenge-accepting behaviour, problem selection, goal setting, and self-reflection have improved with support without experiencing any negative emotions. This research project contributes to the latest trends of motivation and learning research in ITS.

[Faiza Tahir]

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Deputy Vice-Chancellor's Office
Postgraduate Office

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes: *Chapter 3-5 are the expanded versions of Publications in Appendices*

Please detail the nature and extent (%) of contribution by the candidate: *70%*

Certification by Co-authors:
If there is more than one co-author then a single co-author can sign on behalf of all
The undersigned certifys that:

- The above statement correctly reflects the nature and extent of the PhD candidate's contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text

Name: *Tanja Mitrovic*	Signature: *Tanja Mitrovic*	Date: *12/7/2021*
The statement from the PhD candidate is correct.		

[Faiza Tahir]

Glossary

Terminology	Description
ITS	Intelligent Tutoring System
CBM	Constraint-Based Modelling
OLM	Open Learner Model
SQL	Structured Query Language
ICTG	Intelligent Computer and Tutoring Group
SRL	Self-Regulated Learning
ACT-R	Adaptive Control of Thought-Revised
DBMS	Database Management System
GBL	Game Based Learning
LMS	Learning Management System
AOI	Area of Interest
EEG	Electroencephalogram

Chapter 1 Introduction

1.1 Motivation

In recent years, the contribution of computer tools to teaching and tutoring processes has increased substantially. In particular, learning systems have been recognised widely for teaching various hard and soft skills such as problem solving, programming, and presentation skills. Problem solving is considered an essential 21st-century skill for young and old, even for university students who either have STEM or non-STEM majors (Geisinger, 2016). Therefore, the need to develop effective instructional materials and adaptive learning systems (also called adaptive learning environments) has increased to support the problem-solving skills of every learner. These adaptive learning systems support individual learners by focusing on one's particular needs in a learning domain (Weber, 2012). One example of such systems is **I**ntelligent **T**utoring **S**ystems (ITSs), an important learning aid in today's teaching and learning environments (Anderson, Corbett, Koedinger, & Pelletier, 1995; VanLehn, 2011). ITSs mimic human tutoring (one to one, remedial tutoring) based on Artificial Intelligence techniques (Graesser et al., 2001).

Learning is a process of acquiring new knowledge, skill by practice, study, experience, or being taught (Latchman, 1997). Learning gains are the most important outcome of every teaching and tutoring process. Human tutoring is believed to be a benchmark in effective tutoring techniques because of its learning gains (effect size reported 2.0 in Bloom 1984; effect size reported 0.79 in VanLehn 2011). However, in ITSs, the most cited average learning gain was 0.1 (effect size) (Anderson et al., 1995), which has been replaced by 0.76 (effect size) in a relatively recent meta-review (VanLehn, 2011). Additionally, Ma et al. (2014) reported learning gains of ITSs when compared with large-group human instruction (effect size = 0.42), computer-based instruction (effect size = 0.57), and individual textbooks or workbooks (effect size = 0.38).

The learning gains of ITSs are getting closer to human tutoring. However, researchers have attributed the effectiveness of human tutoring to tutors' ability to provide frequent and effective feedback, accurately analyse misconceptions in student knowledge, periodically provide motivational comments and strategies, and affective state identification and regulation (Graesser, VanLehn, Rosé, Jordan, & Harter, 2001; VanLehn, 2011). To imitate human tutors, ITS researchers introduced best practices of learning science and educational psychology. However, these techniques and interventions mainly focus on developing cognitive skills (Hooshyar et al., 2020). Only a few ITSs have managed to cater to learners' motivational and affective needs (Hooshyar et al., 2019).

The research on affect-sensitive ITSs is still in its infancy (Graesser, 2020). However, those ITSs that identified and regulated learners' affective states have reported the higher motivation and engagement of learners (Arroyo et al., 2014; D'Mello et al., 2010; D'Mello et al., 2014; Munshi et al., 2018; Nye et al., 2018). However, not much evidence of affective state regulation effects on learning outcomes has been established because of technical difficulties in determining learners' affective states and complexities of emotions.

ITS researchers introduced various interventions to motivate learners, such as the Open Learner Model (OLM) for supporting metacognitive skills (Mitrovic & Martin, 2007), mastery and performance orientation (Martinez-Miron et al., 2005), self-efficacy (del Soldato & du Boulay, 1995), engagement and self-regulation (Arroyo et al., 2014), regulation of academic achievement emotions (D'Mello et al., 2011) and gamification (Long et al., 2014). However, little research has been conducted to evaluate these motivational strategies' effects on learners' learning outcomes. Additionally, ITSs providing such interventions are mostly developed for early and middle school students, and not much work has been done for undergraduate learners' motivational needs.

This brief overview of related research raises a question: why are affective, motivational, and metacognitive aids necessary when the ITSs have already achieved considerable success in improving learning by supporting cognitive skills? The simple answer in the light of the above discussion should be “for improving learning gains”. The explanation of this answer is the aim of this research project. In this project, I try to fill all the gaps mentioned above to increase the affective, metacognitive, and motivational support to learners while solving problems in ITS.

The ITS in context is SQL-Tutor (Mitrovic, 2003), a problem-solving ITS that many studies have proven as a highly effective learning environment to teach SQL (Structured Query Language). It was developed in 1997 by the Intelligent Computer and Tutoring Group (ICTG) at the University of Canterbury. It provides almost 300 problems for practising SQL queries and has been an essential teaching aid in database courses since 1998 (details in Chapter 2). Despite developing problem-solving skills successfully, learners' affective and motivational states have never been evaluated in the ITS. Likewise, limited metacognitive support was provided in the form of OLM. The reasons to select SQL-Tutor for this research are: (1) it has been used by a large number of students, and (2) SQL-Tutor has been proven to improve cognitive skill of learners, thus providing a strong foundation for evaluating meta cognitive, affective, and motivational strategies to enhance learning gains further.

1.2 Research Plan

The research reported in this thesis aims to *increase learners’ learning gains from SQL-Tutor*. I limit the scope of this dissertation to the affective and motivational aspects of learning. The affective states identification and analysis in this project is the first step towards affective support in SQL-Tutor. For the motivational aspect, I want to explore strategies which have been fairly, vigorously pursued over the past 20 years or so, and there is still lots to explore. Thus, I selected gamification, a motivational strategy, and Self-Regulated Learning (SRL). In this context, the research project has three main goals. First, I want to identify the affective states of learners while they are working with SQL-Tutor. Mainly, I want to analyse the effects of worked examples provided during problem solving on the learners' affective states. My next goal is to explore the effects of a motivational strategy: gamification. In particular, I aim to introduce badges as gamification mechanics and evaluate their effects on learners' motivation and learning outcomes. The last goal of this project is to examine the effects of another strategy i-e., SRL. I plan to introduce three interventions to support SRL in SQL-Tutor and examine the effects of each intervention on learners' motivation and learning outcomes. Based on these goals, I divide the research project into three phases: the affect detection phase, the gamification phase, and the SRL phase. The details of each phase with the relevant research questions are explained in the next section.

1.3 Scope of the Project and Research Questions

I have limited this research to students of the University of Canterbury, using SQL-Tutor in COSC265, the relational database systems course. As mentioned previously, the project has three different phases. In the initial phase, I design and implement the worked examples in SQL-Tutor, and for the determination of affective states of learners, iMotions and Tobii eye tracker is used. The research questions addressed in the study are:

RQ1: Does iMotions accurately identify learner emotions?

RQ2: Do examples help during problem solving in SQL-Tutor?

In the second phase, I want to explore the effects of gamification as a motivational strategy in SQL-Tutor. For this purpose, I extend the standard version of SQL-Tutor with badges and select goal setting, self-testing, and conflict/challenges as target learning behaviours. These selected learning behaviours are represented in SQL-Tutor in the form of goals, quizzes, and daily challenges. I want to conduct a study to evaluate the effects of gamification on learners' motivation and learning by addressing these research questions:

RQ3: What is the effect of gamification on student learning?

RQ4: Do students with different levels of prior knowledge react differently to gamification?

RQ5: What is the effect of gamification on student motivation?

In the last phase of the project, I want to investigate the effects of SRL support in the context of SQL-Tutor. I want to introduce three interventions, goal-setting support, dashboard, and self-reflection prompts, based on the Zimmerman's (2000) SRL framework. I want to conduct the third study to evaluate the effects of SRL support on learners' learning outcomes and motivation by focusing on the following research questions:

RQ 6: What are the effects of SRL support on student learning?

RQ7: What are the effects of each of the three interventions on students' learning behaviours?

RQ8: What are the effects of SRL interventions on learners' SRL skills and motivation?

Moreover, I want to examine the learners' affective states and eye gaze patterns when the SRL interventions are presented. For this purpose, I want to conduct another small lab experiment as a section of the third study to shed light on the following research questions:

RQ 9: What are the major emotions stimulated by each SRL intervention?

RQ 10: What information do students find more useful on the self-reflection prompt and dashboard?

1.4 Structure of the Thesis

I organise my thesis by presenting a comprehensive literature review in Chapter 2, which provides the foundations for my research questions and hypotheses. After that, I confine each study to a separate Chapter along with the related research questions. Chapter 3 presents the first study, which analyses learners' affective states and addresses RQ1 and RQ2. Chapter 4 consists of the second study in which SQL-Tutor was equipped with gamification and addresses RQ3, RQ4, and RQ5. Chapter 5 presents the third study, which implements SRL support interventions, examines their effects on learners' learning outcomes, motivation, and affective states, and discusses RQ6-RQ10. Finally, Chapter 6 summarises my research findings and suggests future work.

1.5 Contribution

There are many contributions of this project in the field of ITS as well as learning and motivation research. Starting from the first contribution of this research, affective states detection and regulation is a relatively recent area in ITS research. However, only a few ITSs have investigated the effects of different affective states of learners because of technical complications of emotions. In SQL-Tutor, affect determination has not been studied before, and therefore this is the first step towards an affect-sensitive SQL-Tutor. Previous research found the effects of gamification in ITSs, which were developed for primary school students. Therefore, it is crucial to evaluate the gamification effects in ITS for learners of all ages and levels. In particular, this research provides gamification insights for undergraduate students who are less likely to be impressed by intangible rewards; which is the second contribution of this project.. Other achievements of this research are supporting three major phases of the SRL framework (Zimmerman, 2000) in an ITS, and analysing the separate, combined, and cyclical effects of SRL support interventions on learners' learning. Understanding the cyclical and combined effects of SRL interventions is a significant part of the SRL framework, largely ignored in SRL research. This research project provides empirical evidence for each of these contributions by conducting studies on SQL-Tutor.

Chapter 2 Background and Literature Review

2.1 Introduction

This Chapter presents a brief literature review to support and motivate this research project and shape the research questions. After this Chapter, I will be able to answer the following questions: what are the current learning gains in ITSs, and how could the proposed solution increase learning and motivation by considering affective and motivational states in SQL-Tutor. Before reviewing the work performed by other researchers in this field, I will present the purpose of the project along with the theoretical framework. This theoretical framework will help readers to understand the context and links between various aspects (affective, metacognitive, motivational) of the research.

In the main literature review, the first Section gives a brief history of ITSs and how much learning and motivation have been achieved, followed by a brief account of SQL-Tutor. This Section creates the background and context of the research project. Section 2.7 discusses how the affective states are determined and evaluated in learning environments and their effects on the learning outcomes, followed by the challenges in this research area. Section 2.8 shows the most recent research on gamification in learning environments. The Section ends with the challenges faced by the gamification research. Section 2.9 introduces SRL and discusses the latest research in this field followed by challenges in learning environments.

2.2 Purpose of the Research and Formation of Research Problem

The blended learning approach is prevalent in higher education settings these days. The reasons are ease and abundance of information. As university education is less structured than intermediate and primary education, it often gives more power to students over their course content and learning strategies. However, distractions and lack of focus are two significant downsides of this approach. The pandemic has turned things more towards online learning environments. As a result, engagement and motivation, which were once issues in learning, have now become major challenges of learning environments. Researchers have found an important cause-effect relationship between learners' motivation and their learning (Schunk & Ertmar, 2000). Many hurdles have affected this relationship, for example, students' affective states. A student should be happy, have joy, or be in the flow to engage longer in the learning process. The flow state is interchangeably used as engagement, where students are involved with the system to the extent that they do not feel boredom, tiredness, or frustration (Csikszentmihalyi & Csikszentmihalyi, 1990).

There are many aspects of motivation; for example, students may not be confident on learning a new task (self-efficacy), or they do not see the value of learning that task (outcome expectation), or sometimes they are not interested in learning the skill (interest). It is relatively straightforward for the human tutor to motivate students if they are less confident and encourage them to an extent or personalize things to boost their interest. However, learning environments with a focus on building cognitive skills lack these motivational tricks and strategies. Even identifying affective states or motivation levels is challenging in these environments, let alone their regulation. In this research, I studied the affective states of learners with two different strategies and analysed their effects in the context of SQL-Tutor. These strategies focus on boosting learners' motivation and engagement, which is believed to increase their learning. Keeping in mind the importance of engagement and motivation for learning in learning environments, I formulated the research aim: *to facilitate learning by enhancing few affective states of learners (joy, delight, engagement) and motivation through various strategies*.

2.3 Proposed Theoretical Framework

Figure 2.1 presents the complete theoretical framework of the research project. The relationship **(A)** between affect analysis and engagement is studied in the context of worked examples to examine if worked examples reduce negative affective states and increase engagement and learning with the system. Next, I studied the relationship **(B)** between motivational strategy-gamification by 1) applying gamification in SQL-Tutor, and 2) analysing the effects on engagement, motivation, and learning. In the following relationship **(C)**, I studied the effects of self-regulated learning on learners' engagement and motivation and, in turn, their learning. In the end, I examined **(D)** whether SRL support caused or increased negative affective states of learners during problem solving. In each of these relationships (A, B, C, D), I evaluated the relationship between engagement, motivation, and learning **(E)**, which is the primary goal of this research. In the next Section, I present a literature review of these relationships in the context of learning environments and ITSs.

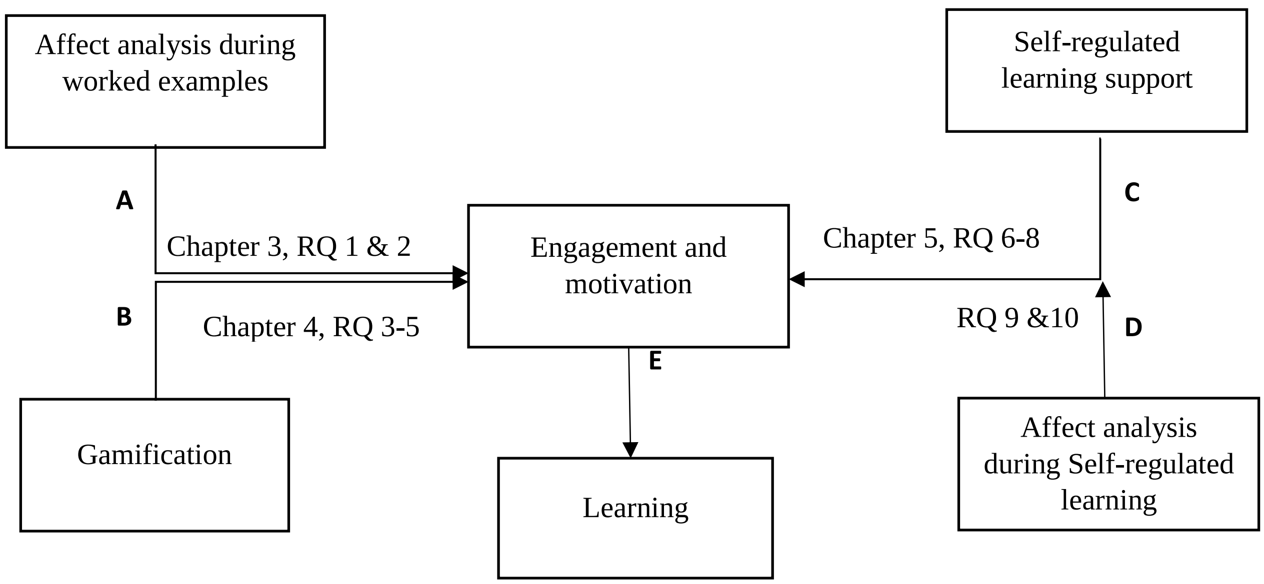


Figure 2.1. A theoretical framework of the research project along with the research questions

2.4 Learning, Engagement, and Motivation in ITSs

Learning is all about change in behaviour due to practice and experience. Engagement and motivation are drivers of learning, engagement referring to the involvement of learners in the learning process, and motivation to achieve a goal (Mitchell, 1982). In the research of learning environments, engagement and motivation are defined as similar constructs (Halverson & Graham, 2019). In many studies, engagement and motivation were analysed together, and their combined effects were reported. Therefore, in the following literature review, learning may refer to all three constructs (engagement, motivation, and learning) except if they have been reported separately.

When learning and learning outcomes are described in ITS research, they are sometimes compared with (one to one) human tutoring. Human tutoring is regarded as an effective teaching methodology because it provides tasks according to the learner's current state of knowledge, provides immediate feedback, or hints on student solutions to increase engagement, helping them understand the weakness and continuously motivating them to become better learners. The effect size of expert human teachers, when compared with classroom tutoring, is reported to be up to two standard deviations (Graesser et al., 2012).

The research on human tutors poses various challenges. First, human tutors cannot conduct extensive student modelling on all the psychological states of learners due to their limited capacity as human beings. These psychological states include cognitive, affective, and metacognitive states (Graesser et al., 2009). Many human tutors cannot provide a range of pedagogical strategies that are helpful in learning and problem solving, such as help-seeking and self-assessment. These limitations created the need for computer tutors to help in the tutoring process, which is also one of the reasons for developing intelligent tutoring systems (Graesser et al., 2012).

ITSs are computer-based learning environments and a combination of artificial intelligence, learning technologies, and educational psychology (Graesser, Conley, & Olney, 2012). The first intelligent tutor was developed in the mid 1970s. Since then, ITSs have been developed for various instructional domains such as mathematics, physics, electronics, programming languages, and information technology (Graesser et al., 2018). The main purpose of those systems is to provide individualized feedback to students and provide adaptive instruction. Empirical studies showed impressive learning gains (up to 1 standard deviation) when compared with other learning methodologies (e-g classroom instructions) (Anderson et al., 1995; VanLehn, 2011; Ma et al., 2014). Moreover, learning with these systems was affordable and found to be highly scalable and adaptable compared to human tutoring.

2.5 Architecture and Types of ITSs

The basic architecture of an ITS, as illustrated in Figure 2.2, consists of four primary modules. (a) a knowledge base that holds all the knowledge components and principles of a specific domain, (b) a pedagogical component composed of pedagogical strategies to provide practice problems and learning content to students, (c) an interface component to facilitate the interaction between the student and the ITS, and (d) the student model which tracks the student's progress. ITSs provide many other features which make them more human-like, such as feedback, hints, fine-grained adaptation (VanLehn, 2006), emotional support (Kort, Reilly, & Picard, 2001), self-regulation strategies (Azevedo, 2002), self-explanation strategies (Conati & VanLehn, 2000) and others. The following three Sections (2.5.1, 2.5.2, 2.5.3) provide a brief overview of the major types of ITSs and their learning gains.

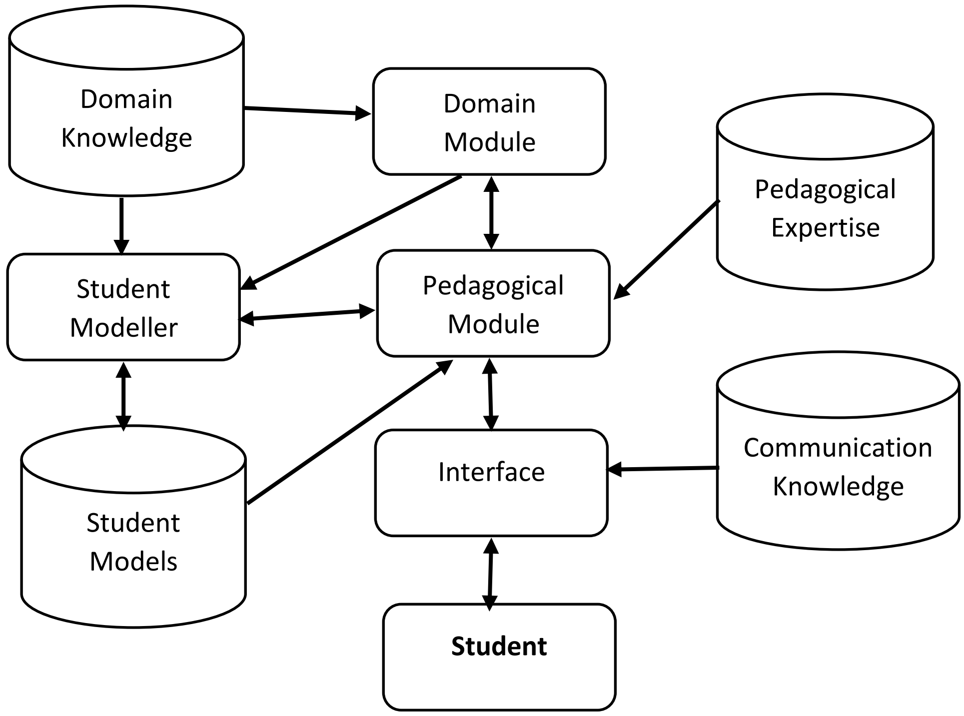


Figure 2.2. The basic architecture of ITS

2.5.1 Cognitive Tutors

Cognitive tutors are successful intelligent tutors developed at the Carnegie Mellon University based on Anderson's (1993) Adaptive Control of Thought (ACT-R) theory. These tutors consist of a psychological problem-solving model based on production and buggy rules. Production rules model the correct steps in the student's solution, whereas buggy rules define the incorrect steps that can lead to incorrect solutions. The production rules consist of the *conditions* and operations. Conditions compare the student's solution with the ideal solution and trigger the related operation. These operations could be a subgoal required at this stage of solving the problem. The mapping between student actions and production rule activation is called *model-tracing*. If the student's solution is in line with the correct solution path of the model, then it is considered a correct solution. However, if the solution is different from the correct path, then the tutor generates a feedback message explaining the error in the solution and providing hints of the correct path. Student modelling is handled by the *knowledge-tracing* approach, which estimates the student's expertise in terms of their correct and incorrect actions. Lynette (Waalkens, Aleven, & Taatgen, 2013) and The Fallacy Tutor (Diana et al., 2018) are the latest examples of these types of tutors.

Effects of cognitive tutors on students' learning outcomes were impressive; for example, when the model-tracing version was compared with the no model-tracing version, the effects size was reported to be 0.75 standard deviations and an average of 0.6 standard deviations in the case of Algebra I (Koedinger et al., 1997). The effect size is a quantitative measure of the magnitude of the experimental effect. The larger the effect size, the stronger the relationship between two variables. However, some of the evaluation studies, for example, Shneyderman (2001), reported only 0.22 standard deviations of improvement in learning outcomes of students who used cognitive tutors compared to those who studied without them. This brief overview about cognitive tutors indicates that learning gains of these tutors are considered significant in the research community; however, scalability, incompleteness of buggy rules, and resource-intensive implementation are the challenges of these tutors. The following Section describes the second approach of developing ITSs and achieving learning gains.

2.5.2 Constraint-Based Tutors

Another learning theory, learning from performance errors proposed by Ohlsson (1996), explains, "Declarative knowledge can help the student to learn from mistakes while acquiring a skill". In other words, mistakes provide a chance for learners to correct their misconceptions in the knowledge domain. This theory forms the basis of the Constraint-Based Modelling (CBM) approach, which, unlike model tracing, analyses the learners' full solution for errors. CBM believes that there can be a correct solution to a problem that does not violate domain principles.

In this approach, domain knowledge is organised in the form of constraints which consist of relevance (CR) and satisfaction (CS) conditions. The relevance condition checks whether the student's solution is relevant for that constraint, and the satisfaction condition further evaluates the solution against the correct solution. A relevant and satisfied constraint shows the correct solution or an aspect of the correct solution. In case of an incorrect solution, a constraint could still be relevant but not satisfied. The final solution is the one that satisfies all constraints. A student model consists of all the satisfied and violated constraints (Mitrovic et al., 2001) for a particular learner. CBM is suitable for both well and ill-defined domains (Mitrovic, Koedinger, & Martin, 2003).

SQL-Tutor (Mitrovic, 1998, 2003) was the first ITS developed using the CBM approach. Many versions of SQL-Tutor have been released, providing new features and upgrading the previous ones. KERMIT is the second most prominent example of these tutors that teaches conceptual database design. Its web-enabled version is called EER-Tutor, which provides multiple levels of feedback (Suraweera & Mitrovic, 2002), and self-explanation support (Weerasinghe & Mitrovic, 2003). The reported effect size of this ITS is 0.63 standard deviations when compared with a no-feedback version. NORMIT is another constraint-based tutor developed to teach data normalization (Mitrovic, 2002). The distinguishing feature of this problem-solving tutor is teaching a procedural task through self-explanations (Mitrovic et al., 2004). This brief account of CBM-based ITSs shows that the tutors are easy to develop and scalable and do not need extensive expert knowledge.

2.5.3 Other Approaches to Developing ITSs

Researchers introduced animated conversational agents in learning environments to cover the conversational part of human tutoring (Atkinson, 2002). These agents can be a mentor, tutor, avatar, peer, or players and interact with students through speech, gesture, keyboard, and a touch panel. Auto-Tutor is an example of this type of ITS (Graesser et al., 2005). It uses natural language to conduct dialogues with students. These dialogues are composed of difficult questions which require reasoning and explanation from students. It advances the dialogue by providing feedback on the student's answer, accompanied by a hint or prompt for more information. It also corrects misconceptions in the student’s answers. The learning gains in evaluation studies reported an average effect size of 0.8 standard deviations for computer literacy (Graesser et al., 2004) and physics (VanLehn et al., 2007). Another example is iSTART (McNamara et al., 2004), which facilitates self-explanation for better reading and comprehension skills. The evaluation studies reported impressive effect size varies between 0.4-1.4 standard deviations for comprehension only.

Wayang Outpost (known as MathSpring) (Arroyo et al., 2004) is an ITS supporting problem-solving skills in mathematics to K-12 students by adaptation of learning material based on the learner's cognitive, meta-cognitive, and affective states. It follows the theory of cognitive apprenticeship (Collins et al., 1988). The assessment studies revealed tutor improvement falls within 0.3 to 0.8 of the effect sizes on standardized tests versus classroom instruction (Arroyo et al., 2004).

Betty’s Brain, a teachable agent (Biswas et al., 2005), was developed for the learning-by-teaching approach (Palthepu et al., 1991).. The evaluation studies reported learning gains of 0.72 standard deviations compared with an intelligent coaching system (Leelawong & Biswas, 2008) . Another famous ITS is Crystal Island (Rowe et al., 2011), a narrative-based learning environment for microbiology students. Its main characteristics are a game-based learning environment and adaptive tutorial selection; and Dragoon (Wetzel et al., 2017), an example-tracing ITS which provides a step-by-step example for every problem to solve.

This brief overview of various developmental approaches indicates that ITSs have incorporated more human-like features, for instance, affect detection, metacognitive and motivational support, and adaptive content. These ITSs reported better learning outcomes than others. The following Section presents an account of SQL-Tutor, the context of this research project, along with the achieved learning gains.

2.6 SQL-Tutor

SQL-Tutor has been found effective in tutoring problem solving in Structured Query Language (SQL) since 1997. SQL is a language used for processing and managing data kept in relational database management systems. It is being taught as a compulsory course in computer science in the Department of Computer Science and Software Engineering at the University of Canterbury. The domain knowledge principles of SQL are taught during classroom lectures, and problem solving is practised in labs using SQL-Tutor. SQL-Tutor has been used in the labs of the University since 1998.

Figure 2.3 shows the basic architecture of SQL-Tutor. I explain a few of its features in this Chapter relevant to the research project (please see Mitrovic (2003) for a complete system description). The standalone SQL-Tutor has four primary modules: (1) The domain module, which contains all the domain knowledge, (2) an interface module, where the student interacts with the system, (3) a pedagogical module, which demonstrates the content of pedagogical actions and (4), a CBM module that evaluates the student's solution.

The web architecture is an extension of the standalone SQL-Tutor, as shown in Figure 2.4. In this version, the interface module is extended with the session manager to maintain separate sessions for students and keep student-systems interaction in log files. Along with the session manager, it contains domain knowledge structures also.

2.6.1 Domain Module

The domain knowledge in SQL-Tutor is represented in the form of constraints. Currently, SQL-Tutor contains over 700 constraints and almost 300 problems for practising problem-solving skills associated with 13 databases. The complexity of those ranges from 1 to 9, where 9 shows the most complex problem. Domain experts have determined this complexity based on various factors, such as the number of clauses of SELECT statement and the number of functions and subqueries it needs to perform in the problem (Mitrovic, 2003).

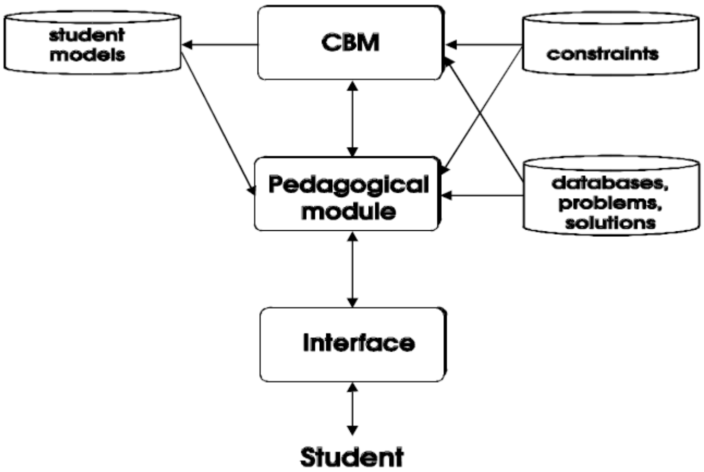


Figure 2.3. Architecture of standalone SQL-Tutor (Mitrovic, 2003)

2.6.2 Interface Module

When a student performs an action, the interface module passes it to the session manager, which puts the action into a relevant session and records the action in the log. This action is sent to the pedagogical module, which examines it and passes it to the student module, which evaluates the student's solution in terms of satisfied and violated constraints and updates the student model. The student module returns the actions to the pedagogical module, which, based on the evaluation, generates specific feedback that is delivered to the student through the session manager and repeats until the student stops problem solving.

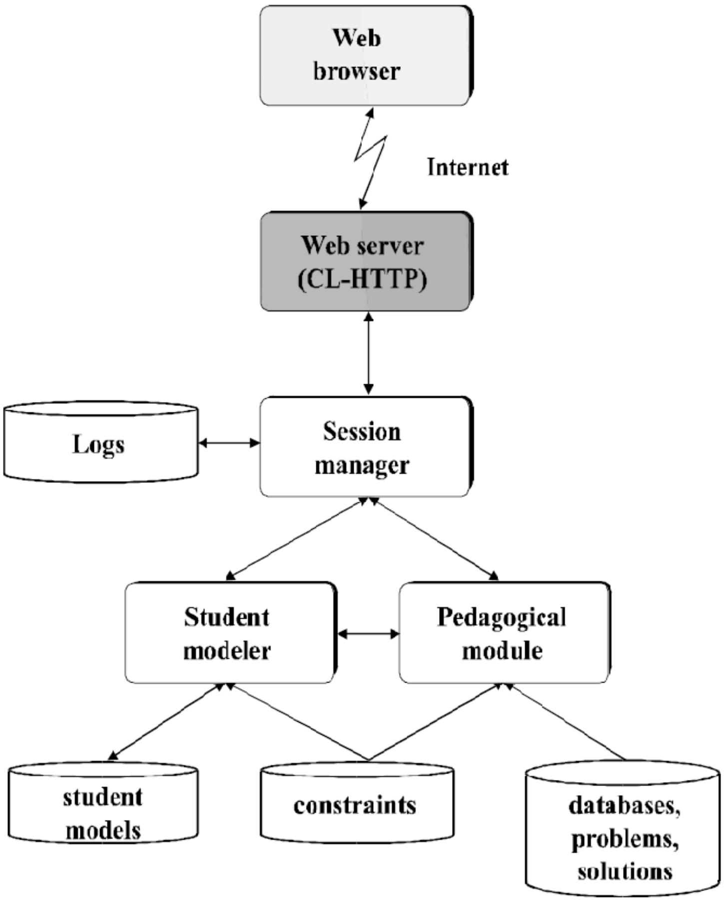


Figure 2.4. The architecture of the web version of SQL-Tutor (Mitrovic, 2003)

The problem-solving interface of SQL-Tutor is illustrated in Figure 2.5. There are a few buttons on the top of the problem-solving interface, which provide various functionalities. For example, the *student model* provides the view of the open learner model, *change database*, changes the schema, the *history* button shows the student's problem-solving history. The *run query* runs the student's submitted solution as a query and produces results in the database management system (DBMS), as shown in Figure 2.6.

The upper panel of the problem-solving interface presents the problem statement and solution space on the left side and the feedback panel on the right side. The bottom panel shows the database schema, which contains the selected database with all the relevant tables and attributes. Clicking on the table name unfolds the attributes and their characteristics. Students can build their queries by using as many clauses as required by the problem.

Once the student submits the solution, they receive feedback. There are six types of feedback available in the system, from *simple feedback* to *complete solution,* as shown in Figure 2.7. *Simple feedback* is the first and basic level of feedback that only reports either the solution is correct or not and the number of errors existing in the solution. *Error flag,* the second level of feedback, presents the part of the solution which is incorrect. *Hint*, the third level of feedback provides the messages attached with the violated constraints. The next level of feedback is *partial solution* that presents the correct solution of the clause in which the solution was incorrect. *List all errors* shows all the errors the student has made so far in that problem, and *complete solution,* the last level of feedback, presents the ideal solution of the problem. The student receives simple feedback on the first submission, whereas error flags and hints are given in the subsequent unsuccessful attempts. The remaining levels of feedback need to be selected explicitly by the learners. Students are free to submit the solution as many times as possible and even repeat the same problem.

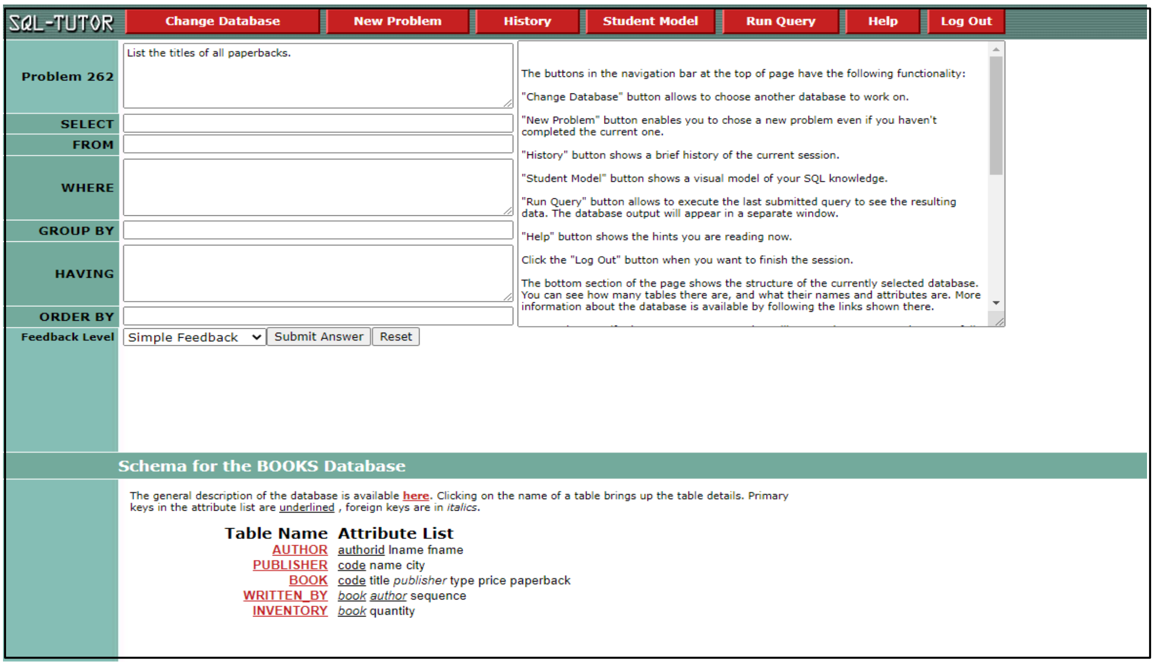


Figure 2.5. Problem-Solving interface of SQL-Tutor

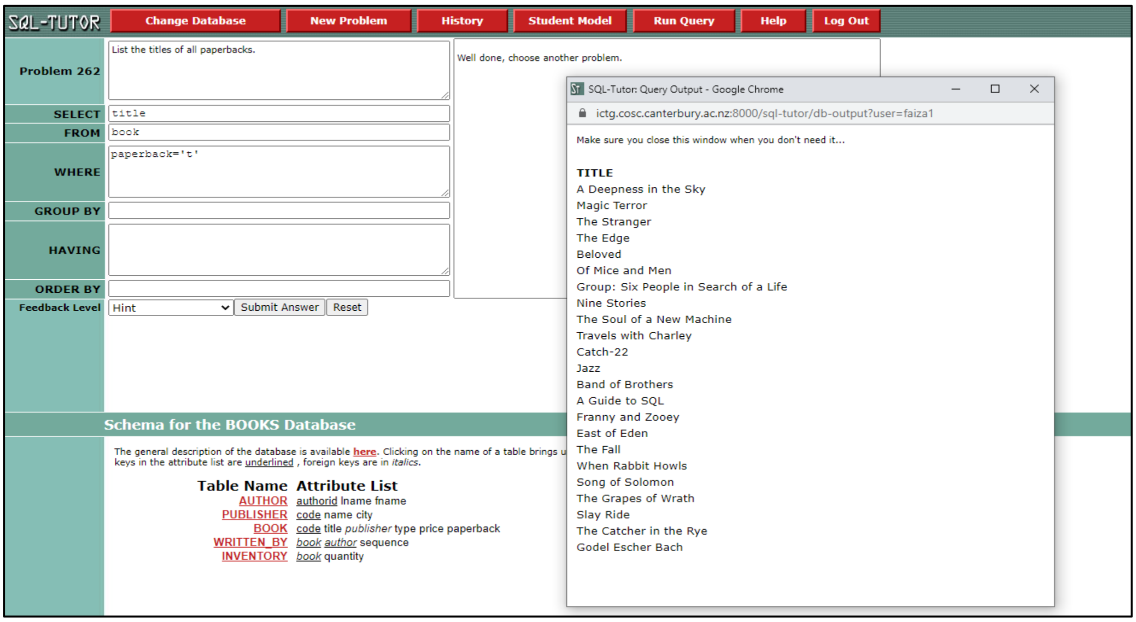


Figure 2.6. Running a query in SQL-Tutor

There are four strategies available in the system to select the next problem, as shown in Figure 2.8. The first strategy presents the following problem from the selected database. The second strategy requires the student to select a problem from a specific clause. The third strategy asks the student to select the problem according to its complexity, while the fourth strategy presents a system-suggested problem to the student according to their student model.

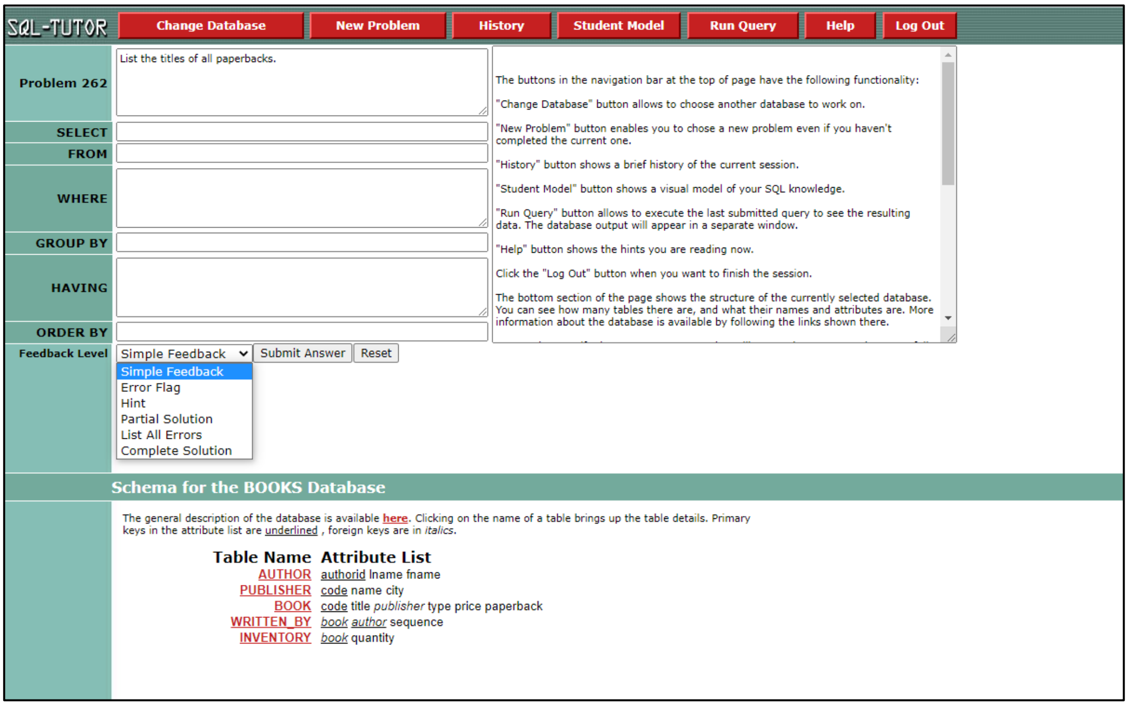


Figure 2.7. Problem-solving interface with feedback levels

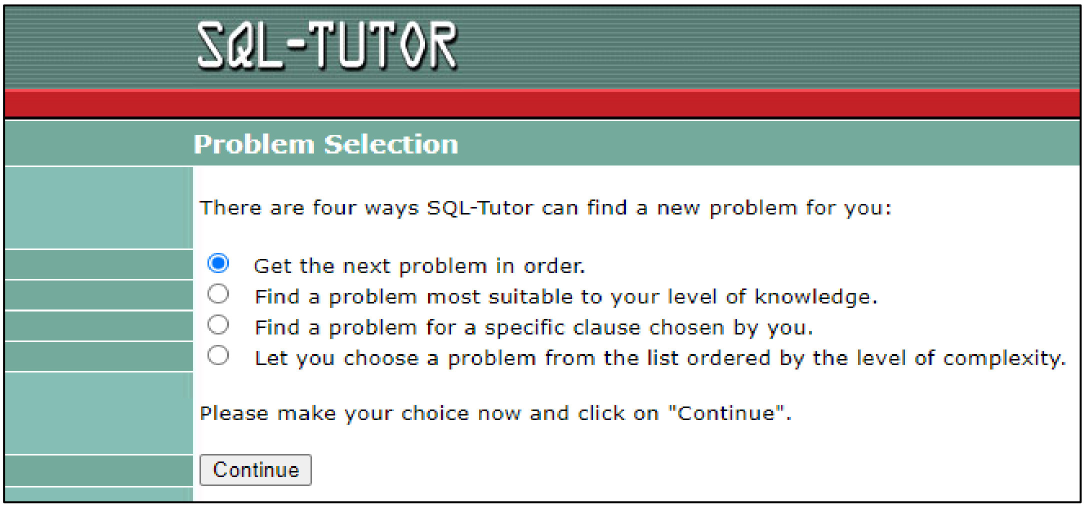


Figure 2.8. Problem selection strategies in SQL-Tutor

2.6.3 Student Modeller

The student modeller is another crucial module of SQL-Tutor, which tracks and visualizes the students' problem-solving progress, as shown in Figure 2.9. The open learner model visualizes students' knowledge of each of the six clauses of the SQL SELECT statement and presents them in skill bars. These bars contain three colours: green represents the correct percentage understanding, red shows incorrect percentage understanding, and white depicts the student has not covered that percentage of knowledge yet. For example, a student covered 13% of domain knowledge related to *where* clause, out of which 12% of the correct knowledge and 1% incorrect knowledge (Figure 2.9).

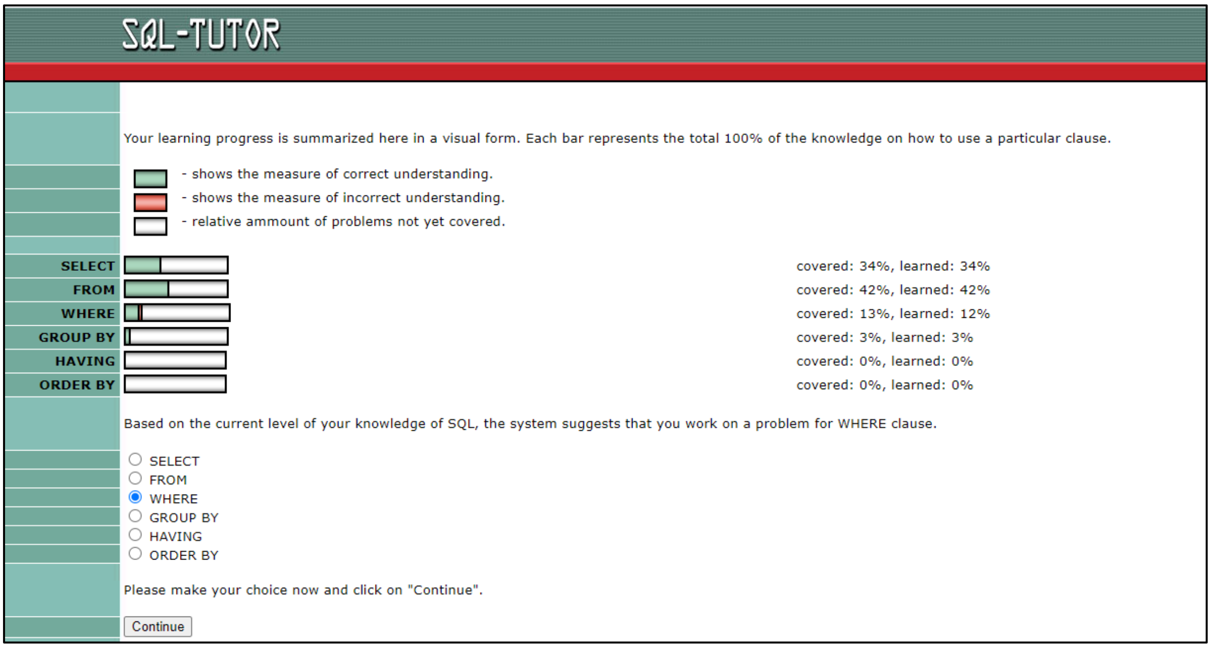


Figure 2.9. Open Learner Model of SQL-Tutor

2.6.4 Evaluation Studies of SQL-Tutor (1999-2016)

The first empirical study (Mitrovic & Ohlsson, 1999) on SQL-Tutor was conducted in 1998, which found evidence of student learning by measuring the satisfied and violated constraints and exam scores. Results revealed that students who interacted with SQL-Tutor violated significantly fewer constraints as they practised their problem-solving skills and scored significantly higher ($p < .01$) in exams than those who did not interact with SQL-Tutor. The second empirical study (Mitrovic & Martin, 2000) compared various levels of feedback and reported a higher learning rate when learners used all errors, error flags, and hints feedback instead of partial and complete solution feedback.

Mitrovic & Suraweera. (2000) introduced an animated pedagogical agent to increase the motivation and the experimental group reported significantly higher ($p < .05$) enjoyment from the system; however, no effects were found on students' learning and engagement. The probabilistic student model (Mayo & Mitrovic, 2000) was introduced in SQL-Tutor and found that students who selected problems based on the student models had fewer attempts on simpler problems and more complex problems than the students who selected problems without the student model. In another study (Mitrovic, 2001), self-assessment skills of students were supported in SQL-Tutor, and significant learning gains ($p < .001$) were found from pre-to post-tests. In a subsequent related study Mitrovic & Martin. (2002) reported that students with low prior knowledge significantly increased their learning from pre to post-test when they self-assessed themselves. Students with higher prior knowledge abandoned fewer problems when they selected problems from the open learner model.

The positive feedback implemented and evaluated in SQL-Tutor (Mitrovic et al., 2013) showed that students who received positive and negative feedback solved the same number of problems in significantly less time ($p < .05$) than those who received only negative feedback. Most recently, worked examples introduced in the ITS and evaluation study (Najar et al., 2014) revealed that novices learned significantly more on conceptual knowledge when provided with both worked examples and problem solving alternately. Additionally, Chen et al. (2016) demonstrated erroneous examples effects, reporting higher problem-solving skills for those who studied erroneous examples than worked examples.

This brief overview of various studies conducted on SQL-Tutor reveals its effectiveness in increasing problem solving and learning outcomes. However, affective states have not been analysed in the ITS, which is critical for improving learning. Moreover, SQL-Tutor has demonstrated the effects of motivational strategies such as animated pedagogical agent and positive feedback. However, none of these strategies influenced learning and engagement with the system. Few studies evaluated the effect of the open learner model and metacognitive activities such as self-explanation and self-assessment. However, the open learner model is not a complete representation of learners' progress in the ITS, and therefore, a more comprehensive and robust learner model is required. This research tries to fill these gaps and improve learner motivation engagement and learning outcomes from SQL-Tutor. The following Section presents the latest trends in research in each affective and motivational support.

2.7 Affective Support in Learning Environments

Emotions are defined as:

> "A complex set of interactions among subjective and objective factors, mediated by neural/hormonal systems, which can (a) give rise to affective experiences such as feelings of arousal, pleasure/displeasure; (b) generate cognitive processes such as emotionally relevant perceptual effects, appraisals, labelling processes; (c) activate widespread physiological adjustments to the arousing conditions; and (d) lead to behaviour that is often, but not always, expressive, goal-directed, and adaptive"(Kleinginna & Kleinginna, 1981).

In simple words, emotions are biologically behavioural responses, feelings, thoughts, and intensity of pleasure and displeasure. *Affect* is defined as a "neurophysiological state consciously accessible as a simple primitive non-reflective feeling most evident in mood and emotion but always available to consciousness" (Barrett & Bliss-Moreau, 2009). In psychology, affect sometimes refers to underlying experiences from emotions, moods, and feelings. The difference between affect and emotion is that emotion can be a state that occurs for a short period.

In a series of studies, Pekrun et al. (2002) found that hope, pride, enjoyment, anger, anxiety, shame, boredom, and hopelessness are mostly reported in academic settings that are different from the basic emotions suggested by Ekman (1999). However, affect is an experience of those emotional states which lasts longer than emotion (Ekkekakis, 2012). The control-value theory (Pekrun, 2006), associates emotions with one's control and value outcomes. Higher levels of perceived control and increased task value lead to positive emotions such as enjoyment or pride, whereas lower levels of control lead to negative emotions such as hopelessness or anxiety.

Affective states play an essential role in motivating students and affecting their learning outcomes (Pekrun & Perry, 2014). Research shows that boredom predicts low future performance in low and high-performing students, and only high-performing students showed delight during the assessments (Munshi et al., 2018). Surprise often negatively influences learners' academic efforts (Nye et al., 2018), and confusion sometimes positively influences learning (D'Mello et al., 2014).

After the evidence of affective states influencing student's learning outcomes and efforts, the next step is to identify and regulate the affective states in learning environments and ITSs. The most comprehensive example of affect-aware ITSs is MathSpring (Arroyo et al., 2014). MathSpring identified the affective states of learners using sensors and introduced animated companions to provide support when learners experienced boredom and frustration. The results demonstrate that a female learning companion increased students' interest in the topic and improved their self-belief in learning, particularly the underperforming group. AutoTutor (D'Mello et al., 2011) is another example of identifying affective states and supporting learners with the adaptive pedagogical content. Results of the evaluation study reported higher learning gains by students who had insufficient domain knowledge. However, students liked the affect-aware tutor and attributed it as being similar to the human tutor.

Researchers use a few methodologies for affective state identification in learning environments. For example, self-reports and think-aloud protocols are the most common and straightforward methods, and they are suitable when the research into affective state is in the early stages (Graesser, 2020). Expert judgment, for example, BROMP (Ocumpaugh, 2015), is another methodology for affect detection, which relies on human judgment to record affective states. Other ways of identification are physiological measures such as using various sensors and cameras for affective state detection, for example, sitting posture, facial expressions, eye tracking, and body pressure (Sik & Sungho, 2018; Nalepa, Kutt, & Bobek, 2018). However, this method of affective states identification is unreliable due to noise in data collection and frequency differences between sensors. Another method is to examine the students' behavioural data by using data mining techniques that need thousands of affective states to train the input features to produce desired outputs (Picard et al., 2004). None of these methods is considered standard because of the variable and highly subjective nature of emotions. Researchers are now focusing on multichannel or multimodal affect identification methods. These methods combine two or more affect identification methodologies discussed above to increase the predictability of those states.

In a nutshell, affective state identification and regulation are complex and resource-intensive research requiring extensive expert knowledge. However, this identification and regulation pay off as the learning outcomes and efforts were greatly influenced by learners' affective states. The following Section will discuss the major challenges faced by affective state researchers while identifying and regulating them in learning environments and ITSs.

2.7.1 Challenges in Affect-Aware Learning Environments

Research on affect in learning environments has been conducted for more than 20 years but still has a long way to go, and researchers face many issues and challenges to detect and regulate those states (Graesser, 2020). The first and obvious challenge is the small sample size of evaluation studies which hindered the real effects of affects and emotions. The second challenge is the novelty effect of the system and intervention. Research revealed that the first 15 minutes of learner interaction with the system are crucial and shape their liking or disliking about the system or the intervention, also known as novelty effects (D'Mello & Graesser., 2012). However, after 10 hours, those novelty effects started fading or no longer remain intense. Most of these studies span 30 min to 2 hours and might not give enough exposure to reduce the novelty effect. It has also been noticed that the novelty effect causes the cognitive load, which might become a distraction during learning (Mayer, 2011).

Another challenge is that different emotions can be experienced simultaneously, which means analysing the combinational effect of affects. This area of research is in its early stages, and not much has been investigated so far. However, the transition between emotions has been investigated during interaction with the learning environments (D'Mello & Graesser, 2012) and could be a way forward. Another challenge is to regulate and respond to these affective states in learning environments. Most systems provide various cognitive strategies to support learners when found in a negative affective state. However, not much empirical evidence is available to prove the appropriateness of these strategies (Johnson & Lester, 2016; Rowe et al., 2011).

2.8 Gamification as a Motivational Strategy

Game-based learning (GBL) has proved its effectiveness in improving self-monitoring, problem recognition, problem solving, decision making, short/long-term memory retention, and social skills (Corti, 2006; Ellis et al., 2006; Mitchell & Savill-Smith, 2004; Prensky, 2003; Rieber, 1996). However, the development of games is time and resource intensive and subject to various technical and social concerns (Sanford & Madill, 2006; Susi et al., 2007). The idea of gamification focuses on learning, not on the play; this means separating gamification from playfulness.

Gamification is defined as "the use of game design elements in non-game contexts" (Deterding et al., 2011). In other words, gamification is the mechanism that provides the game-like experience in settings where game-based development is not viable. It is considered less expensive than standalone games (Dicheva et al., 2015; Landers et al., 2017). Gamification aims to increase motivation by combining the efficiency of utilitarian systems and the enjoyment of hedonic systems (Koivisto & Hamari, 2019).

The ease of applying gamification and its benefits are reasons for its popularity. Three meta-analysis studies (Alhammad & Moreno, 2018; Hamari, Koivisto, & Sarsa, 2014b; Koivisto & Hamari, 2019) reported education as the major area of gamification influence with the largest effects found on student motivation and engagement. These studies identified two major trends in gamification research: (1) focus on behavioural changes targeting engagement, enjoyment and motivation, and (2) adaptation of gamification based on user characteristics such as playing attitude, personality, traits and gender (Klock et al., 2020). The authors of these studies identified several methodological problems with reported studies, including small sample size, short durations with no control conditions, implying several gamification mechanics, not reporting negative or neutral effects, and reliance on self-reporting instruments.

Nicholson (2015), in the RECIPE of meaningful gamification, divides the concept into two categories: reward-based gamification applied when users have short-term goals and the system needs to engage them to foster performance, and meaningful gamification which deals with real long-term behavioural changes. The proposed framework (RECIPE) elaborates the features of meaningful gamification: it should provide a narrative as context (exposition), allow players (students) to accept the defeat while learning (play), encourage students to seek more knowledge (information), give options and autonomy (choice), encourage students to discover (engagement), and reflect and relate on experiences (reflection). The author suggests that reward-based gamification should be applied first when introducing gamification in an environment, and then gradually transform into meaningful gamification that leaves learners with a real behavioural change to interact with the environment purposefully.

2.8.1 Gamification and Various Motivational Aspects

Most of the gamified systems explored the effects on student engagement and motivation as mentioned in studies above and in (Hamari et al., 2014). However, in those studies, motivation was measured either by the number of awards a student has achieved, the effort to achieve those awards (number of problems attempted, number of edits, etc.), or learners' opinions about future use of the system. The interplay of various motivational aspects was neglected in the research. For example, self-efficacy, mentioned in the social cognition theory, is a powerful tool to influence students' motivation, achievement, and self-regulated learning (Schunk & DiBenedetto, 2020). Bandura (1981) reported that self-efficacious individuals tend to work harder and persist longer in challenging tasks. At the start of a task, students' self-efficacy is based on their prior experience. As they are working, the attitude towards the goal, information processing, and feedback from the teacher on the effort and rewards gave them signals on how they were learning, which helped them assess their efficacy (Schunk, 1991).

Rewards are considered a mechanism to increase self-efficacy if they are linked with students' achievement and learning (Bandura, 1986) and deliver the highest efficacy and learning when combined with goals (Schunk, 1984). Another motivation aspect that can be influenced by gamification is perceived competence. The cognition evaluation theory (Deci & Ryan, 2010) suggested that when rewards are combined with goals, they stimulate intrinsic motivation and perceived competence. The theory mentioned that increase or decrease in intrinsic motivation also increases and decreases one's perception and feelings of competence. Topic interest is a relatively less explored motivational aspect in literature. It is known as the interest develops when individuals are introduced to a topic and influences students' affective responses related to their persistence and learning (Ainley et al., 2002). The four-phased model of interest development mentioned that interest in a topic can be increased with provided facilitation and students' self-efficacy (Hidi & Renninger, 2006).

From this brief literature review, I infer that those various aspects of motivation are strongly related to learning, and rewards might help to strengthen the relationship between motivation and learning. These motivational aspects, such as self-efficacy, topic interest, and perceived competency, are linked and complement each other (Mayer, 1998). There is a need to explore the influence of these aspects in gamification to determine which ones are impacting or what unique contribution these aspects have on student motivation.

2.8.2 Theory of Gamified Learning

The theory of gamified learning presented by Landers (2014) specifies the causal relationship between gamification and learning. This theory elaborated that gamification does not directly impact learning outcomes; however, for gamification to be applied successfully, the learning behaviour must be influenced and affected. In short, gamification must be mediated or moderated by learning behaviours, which act as a causal force between gamification and learning outcomes. This modified learning behaviour, in turn, yields the learning outcomes.

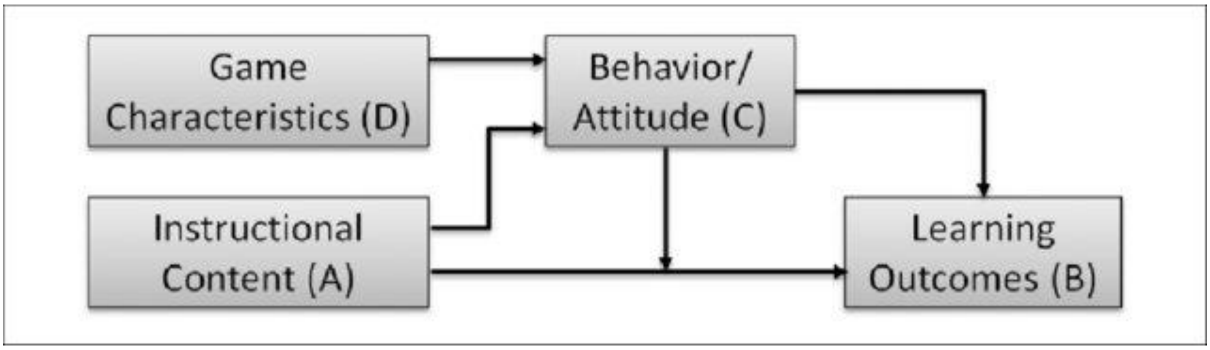


Figure 2.10. The framework of gamification by Landers (2014)

Landers proposed two processes of how gamification influences learning, as shown in Figure 2.10. The first, more direct process, is called mediator (D→C→B), in which game elements directly affect a learner's behaviour which in turn increases/decreases learning outcomes. Examples of this behaviour are time-on-task, practising, and revising. If gamification is capable of successfully increasing time-on-task, practising on a specific task, or motivating the learner to revise the material/problems to increase their understanding, these behaviors positively impact the learner's performance.

The other, less direct process is called moderator (C on A→B), where game intervention affects learner behaviour (psychologically), which subsequently affects the existing established (positive) relationship between instructional content and learning outcomes. Playfulness can be an example of this type of relationship. If gamification helps learners focus, understand, and interact more with instructional content in a playful way, it successfully strengthens the relationship between instructional content and learning outcomes.

In a study using a leader board as a game mechanic and time-on-task as the mediating behaviour, Landers and Landers (2014) found a 27% improvement in learning in the experimental group compared to the control group. Helmefalk (2019) proposed another gamification framework (M-PM-O) having a similar path between game mechanics and outcomes mediated by psychological processes (flow, enjoyment, engagement, motivation). The author suggested that different moderators such as demography, time, space, or platform may affect the mediating relationship. Moreover, this study emphasized evaluating the effect of a single game mechanic on a particular learning outcome under the influence of a particular mediator or moderator.

2.8.3 Gamification in Learning Environments

Gamification has been applied to various learning environments with mixed results. A study with Code Academy and Khan Academy found that gamification did not always motivate students to start using the system but helped them engage with the system for a longer time once they started using it (van Roy et al., 2018). Another study with Stack Overflow reports that badges motivate users to edit more questions but do not help them to ask more questions (Marder, 2015). Another similar study (Suh et al., 2018) reports the mediation effects of need satisfaction between gamification and enjoyment in the Q&A website. In this study, authors found that rewards implemented as points, levels, badges, and leader boards had a significant effect on psychological behaviours (competence, autonomy, and relatedness), which increased enjoyment and engagement with the system. A potential explanation for mixed results could be the voluntary nature of these systems. Students who were already motivated to use such systems did not require external stimuli. However, in some cases where the use of a system is compulsory or a part of the course, such as learning management systems (LMS), it mostly yielded positive results. For example, O'Donovan et al. (2013) gamified an undergraduate course on developing computer games by adding experience points, badges, leader boards, storylines, and themes. They reported significant improvement in student engagement and motivation by influencing attendance and self-testing behaviour. The leader board was found to be the best motivational element. Denny and colleagues (2018) investigated the effects of badges on learning outcomes by mediating self-testing behaviour in a peer learning system. They found a 4.5% improvement in the exam score of the gamified group and regarded gamification as a valuable activity to increase student engagement. Another recent short-duration study conducted with undergraduate and postgraduate students (Legaki et al., 2020) analysed the effects of challenge-based gamification on learning performance. The gamified system targeted the playing behaviour of students, in comparison to the reading-only group and the no-intervention groups. The authors reported a 34.75% improvement in students' learning performance in the gamified group. Besides the gamified system being part of the course, another major contributing factor of these successes is targeting students' learning behaviours which caused performance and learning.

However, not all gamification studies in learning revealed positive results, with some even revealing negative effects of gamification on motivation and learning. Haaranen and colleagues. (2014) investigated the effects of badges in a data structures and algorithm course.

The badges were awarded for time management, early submission, and completing exercises. The results showed no significant effects on learning outcomes, and the students were mostly indifferent about badges. The authors reported that students stopped working once they achieved enough scores for passing the course. This finding opposed the effects of motivation through gamification. Other adverse effects include loss of performance, undesired behaviour, indifference, and declining motivation (Hanus & Fox, 2015; Toda, Valle, & Isotani, 2017).

Although numerous papers are investigating the effect of gamification in online educational systems, there is very little research focusing on gamification of ITSs. Long and Aleven. (2014) investigated the effects of two gamification features in Lynette (the equation solving ITS), re-practising previously completed problems, and rewards for completed problems. The authors reported that gamifying Lynette did not result in increased learning or enjoyment. However, the highest learning gains were found for those students who re-practised previously completed problems but received no rewards on their performance. In the subsequent study (Long & Aleven, 2016), Lynette rewarded students in the form of stars and badges when they selected problems and showed perseverance in practising new problems. As a result, the gamified group showed higher learning outcomes than non-gamified groups with improved problem-selection strategies. In another study, Abramovich and colleagues (2013) studied gamification in a CS2N intelligent tutoring system. Badges were awarded for skills mastery or continued use of the system. The results revealed that, although in the badges group, students' interest in the topic has increased and their counter-productive behaviour has decreased, badges did not improve learning. . Additionally, the study highlighted the interplay of motivation for students with different background knowledge and attributed badge design to poor motivation in students.

This brief overview shows that gamification has both positive and negative effects on students' learning and performance depending on the context and game mechanics. It is observed that selecting appropriate learning behaviours for gamification played a significant role in gamification success.

2.8.4 Challenges in Gamification

Researchers experienced many challenges while implementing gamification in learning environments. For instance, gamification projects inconsistently considered students' learning behaviour and gamification frameworks. Most of the studies that reported positive results neither followed a gamified theory or a specific framework nor reported design guidelines. Nearly all the studies applied multiple game mechanics to influence more than one learning behaviour. Therefore, understanding which game mechanic is suitable for a particular learning behaviour remained unclear (Helmefalk, 2019). Lack of empirical studies and especially controlled experiments is another reason to remain inconclusive about the effects of gamification as mentioned in the meta studies (Hamari et al., 2014, Koivisto et al., 2019).

2.9 Self-Regulated Learning (SRL)

Self-regulated learning can be specified as "self-generated thoughts, feelings and behaviours that are oriented to attain goals" (Zimmerman, 2000). This concept goes beyond the knowledge of a skill and involves the behaviour to apply that skill properly. In the higher education environment, students are less reliant on teachers and more on the strategies to self-regulate their learning which positively impact their performance (Broadbent, 2017). According to the study, self-regulated components (self-efficacy, persistence, effort, and goal level) accounted for 17% of the variance in learning (Sitzmann & Ely, 2011). With these results, researchers should not assume that learners know how to self-regulate in learning environments because failing in self-regulated learning is also one of the indicators of the high dropout rate in higher education settings (Lee & Choi, 2011).

SRL is not just a personality attribute that a student can acquire or lack, but it is a complete set of processes and activities which a learner can apply to complete a learning task, for example, setting goals, applying effective strategies to attain those goals, monitoring performance, and managing time and environment effectively. This brings us to a critical question: how do these selective learning processes combine to form a self-regulated learner? Zimmerman (2000) proposed a cyclical model to connect and combine various SRL processes. In the following Section, I briefly discuss the model and its processes.

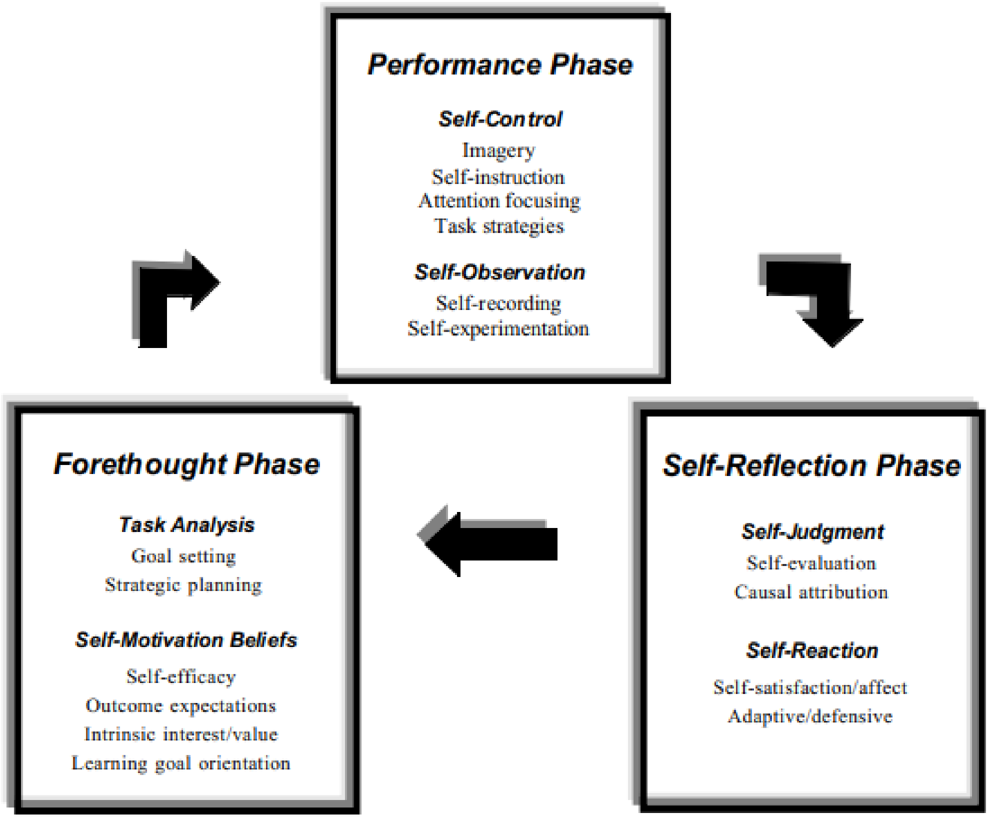


Figure 2.11. Self-regulated Model by Zimmerman (2001)

2.9.1 SRL Model

From the social-cognitive perspective, self-regulatory processes are categorized into three phases, as shown in Figure 2.11:

1. The forethought phase, which consists of processes that take place before a student starts learning.
2. The performance phase which consists of those processes considered during learning.
3. The self-reflection phase which refers to those processes which occur after learning

The following is a brief discussion on each of these phases.

Forethought Phase. This phase has been defined by two processes: task analysis and self-motivation. Students set specific goals (goal setting) during task analysis and devise a plan to attain those goals (strategic planning). There is extensive research (Latham & Yukl, 1975; Locke & Latham, 2002; Locke & Latham, 2019) on goal setting, which provides evidence of learning for those students who set specific domain-oriented goals and achieve them using different strategies; for example, solving numerous problems to learn, evaluating algebraic formulas.

Self-motivation refers to various beliefs of students about their learning capabilities (self-efficacy) and the expectations from learning outcomes (outcome expectations). Another perspective of self-motivation is learners' intrinsic/extrinsic interest in a particular topic, which increases the value of a task related to that topic. Learners' goal orientation is somewhat connected with their intrinsic interest in the topic. The difference is that learners' goal orientation focuses on learning a topic in which they are interested. For example, the student who finds algebra fascinating attains mastery of it and is motivated to adopt the self-regulated way of learning.

Performance Phase. The performance phase consists of two major processes: self-control and self-observation. Self-control deals with the implementation of goals and strategies selected in the forethought phase. The crucial mechanisms of this process are imagery, self-instruction, attention focusing, and task strategies. A student who is interested in learning expression evaluation in mathematics can make different flow charts for understanding (imagery), making a connection with various algebraic formulas (self-instruct), can locate themselves away from distractions (attention focus), and group expressions based on their complexity or evaluating formulas (task strategies).

The self-observation process consists of self-recording and self-experimentation sub-processes. The purpose of these processes is to observe oneself while studying and find causes of events. Students can record themselves during the study to check how much time they have spent on the learning and how much time spent on distractions. In self-experimentation, the student can change the study place or start collaborative learning to see if learning time can be utilized more productively. Self-observation is usually known as self-monitoring. In this process, the learner monitors their performance in attaining the goals decided in the forethought phase and tries to comprehend the reasons for failures in performance.

Self-reflection Phase. Self-reflection also consists of two categories: self-judgment and self-reaction. Self-judgment can be considered a form of self-evaluation that defines one’s self-observed performance with either one’s prior performance, or the performance of peers, or with a performance standard. Another form of self-judgment is causal attribution, which specifies one’s belief about the source of failures and successes. If a student attributes the roots of low scores in mathematics tests to their poor ability, then the motivation for learning mathematics remains low. However, suppose the cause of a poor score is attributed to a controllable process, such as using a different task strategy or proper time management. In that case, it can increase motivation for learning.

Self-reaction involves the feeling of self-satisfaction and positive affect. If the student is satisfied with their performance and delighted while working, then they show more motivation towards work. However, negative affective states or dissatisfaction with the work leads to lower motivation. Defensive or adaptive approaches are another form of self-reaction. A learner facing the prospect of failure on a mathematical test either remains absent on test days or drops the course in a defensive approach. On the other hand, in the adaptive approach, the learner changes the learning strategy or revises the strategic planning to improve learning.

The process of self-regulation is repetitive and cyclical because it reflects on previous efforts and provides reasons for the next task. If a student is experiencing self-dissatisfaction, they be less self-efficacious in subsequent learning efforts. Moreover, self-regulation phases highly correlate with each other, such as the student setting a specific goal; they will likely achieve this by continuously putting in effort while monitoring performance and adapting learning strategies.

Research shows that self-regulated processes are not one's ability, which is considered fixed, but these processes can be taught, learned, and attained (Schunk & Zimmerman, 1998). Teachers sometimes encourage students to set goals for themselves or give various learning strategies in the classroom settings. Few teachers provide various learning strategies, but students are seldom asked to reflect upon the adequacy of efforts and strategies. The teacher does not pay much attention to students' motivational beliefs about learning and hardly observes the causal attribution of failures. Even though most people believe in self-regulated learning as fixed and uncontrollable processes, these processes can be learned; in fact, learners actively seek help to improve their learning (Zimmerman, 2000).

2.9.2 Self-regulated Learning Support in Learning Environments

Two significant approaches are used to implement SRL in learning environments (Panadero et al., 2016). First, technologies (mobile apps) are developed to directly instruct how to develop and enhance students' SRL skills. These instructions are given before or parallel with course instructions. An example of this type of intervention is online SRL training sessions and learning diaries focusing on skill development in all the phases of the SRL model (Dörrenbächer & Perels, 2016). Training sessions are conducted several weeks before the learning session, and diaries are filled in during learning sessions. The downsides of this approach were time management, as the learners must perform extra tasks to complete diaries and write them up every day. The combined effects of both interventions were found to be significant.

In the second approach, SRL interventions are combined within the learning environment and provide support while completing their tasks. This approach scaffolds through prompts for students when they do not use SRL interventions and provides feedback for effectively using them. An example of this type of intervention is nStudy (Winne & Hadwin, 2013). nStudy provides a combination of tools to help students learn about a specific topic by using various SRL resources (summaries, comments, and underlined texts) and collects their trace data regarding learners' learning experience. Additionally, it provides peers' and teachers' feedback on student learning which helps adapt future learning behaviours. The drawback of this approach is that it is mostly suitable for well-defined domains of study where the problem solving follows a similar step-by-step fashion. For ill-defined domains where problem solving can be done in various ways, capturing traces might be complicated, which eventually makes tracing learner' SRL and metacognition behaviours more difficult.

The second example of this type of ingrained SRL intervention is found in intelligent tutoring systems. These systems combine the pedagogical modules with the student learner module, which estimates student learning in correct knowledge, misconceptions, and topics that are not covered yet (Mitrovic & Martin, 2002). The learner module also helps understand the learners' metacognitive and motivational requirements (Arroyo et al., 2014). A prominent example of self-regulated support in learning environments is the open learner model (OLM). The purpose of incorporating OLMs is to increase student awareness regarding their learning and misconceptions. A systematic meta-analysis about OLMs (Hooshyar et al., 2020) revealed that OLM has links with two SRL phases, performance and self-reflection, but does not support the forethought phase. This review reports that OLMs mainly provide cognitive support, a few provide metacognitive support, and almost none provide affective state support.

The second more comprehensive example of this approach is MetaTutor (Duffy & Azevedo, 2015) which scaffolds students through four agents: Pam for planning, Sam for helping in making strategies, Mary for monitoring, and Gavin to inform about system features. This system provides a set of resources to support separate SRL phases. In the forethought phase, goal-setting support is provided by Sam. In the monitoring and self-reflection phases, support is provided by the feeling of knowing (FOK) and judgment of learning (JOL) questions. Sam keeps prompting learners to use different SRL activities such as note-taking and making summaries during the study. Results of the system evaluation revealed that receiving prompts for using SRL activities did not significantly increase students' learning. MetaTutor trained learners for SRL activities before they started using the system. During learning, the agents prompted learners to use SRL resources and provided adaptive feedback on effective and ineffective learning strategies. The problem with the system is that it does not know how much scaffolding a student requires and when it is needed to remain motivated to use SRL resources (Duffy & Azevedo, 2015).

MetaTutor is the only ITS to support all SRL phases; however, most ITSs support a single SRL phase. In a study (Aleven et al., 2010), one SRL activity, help-seeking strategy, was supported and evaluated in the geometry tutor. The tutor provided help based on students' help-seeking behaviour. The evaluation study found improvement in students' help-seeking behaviour, but no significant improvements were found in learning. A subsequent longitudinal study found improved help-seeking skill transfer in the fourth month of the study session. Another study (Arroyo et al., 2007) demonstrated the single SRL process: self-monitoring through a simple intervention for disengaged students of AutoTutor. The intervention consists of a graph showing the student's learning progress and a message that provides a tip for solving the following problem. Students in the intervention group showed higher post-test and exam scores and mentioned the intervention as helpful in an after-study survey.

Another study (Leelawong & Biswas, 2008) with Betty's Brain revealed improved learning for students who received explicit support for five SRL processes and activities such as self-assessment, information seeking, social interaction, and monitoring.

ITSs rely heavily on content; therefore, designing, and ingraining SRL supporting activities needs a proper learning design that is time-consuming. Meta-Tutor is one of the pioneer ITSs which incorporated SRL activities in their learning environment; however, the cost of implementing such systems is much higher than other ITSs.

2.9.3 Challenges in SRL Research

Research into self-regulated learning support in learning environments has several challenges. The first challenge is to figure out those behaviours which depict SRL skills or processes. As these behaviours are collected from self-reports, and from the system's trace data, it is very important first to understand the nature of these behaviours and then carefully map them on various SRL processes. Furthermore, analysing those specific behaviours is yet another challenge researchers deal with while interpreting the results.

The second challenge is that only a few studies explored the effect of SRL interventions on the learning and performance of students. Broadbent and Poon. (2015) mentioned that between 2005 and 2015, only ten studies measured the effects of SRL on learning outcomes (GPA, grades). The third challenge is that learning environments do not cover all SRL phases. The recursive and combined impacts of all phases are hidden in these environments, which support certain SRL behaviours. Incorporating all phases of SRL takes a lot of time and effort because the researcher has to figure out which tools and procedures work best for each phase. Another difficult decision is whether interventions should be domain specific or independent, having advantages and disadvantages on both.

2.10 Discussion

To summarise this Chapter, I started with a brief overview of different ITSs and their comparison with human tutoring, which showed that ITSs achieve comparable learning gains over human tutors. Additionally, ITSs have the edge over human tutoring in student modelling, adaptive task provision, and providing multiple pedagogical strategies. However, identifying learners' affective states and increasing their motivation and engagement with learning content is still very important yet challenging for ITSs. As mentioned above, affect detection is a challenge in learning environments because of the novelty effects, the unreliability of sensor data, and moment-by-moment changing emotions. Few ITSs, for example, AutoTutor and MathSpring, have incorporated interventions to analyse and regulate affective states. However, limited learning effects provide evidence that affect and emotion research in ITS needs much more attention.

CBM tutors have not done much in analysing affective states, so it is important to take the first step in this direction (addressed in RQ1 & 2). Gamification, the proposed motivational method, was theorised in 2010 and has since gained prominence in learning environment research. Despite all of gamification's appealing characteristics, its impact on student learning remains a major research subject. The application of gamification in ITSs is scarce, and only two ITSs, Lynette and CS2N, have analysed its effects. However, these systems were developed for primary school students; therefore, it is important to see the learning, motivation, and engagement effects of gamification in ITS for undergraduate students (addressed in RQ 3-5). The other strategy reviewed in the literature is SRL which is considered as one of the most crucial learning strategies in today’s online learning space. However, its effects on learning outcomes are not investigated much. Another challenge in SRL research is to identify various activities and interventions that may improve students’ SRL skills and motivate them to use those interventions during their learning sessions. Many learning environments applied either one or two SRL processes, making it impossible to analyse and interpret the cyclical effects on learning and motivation (addressed in RQ6-10). Therefore, in this project, I try to cover these gaps by conducting three classroom studies, examining the effects of affect, gamification, and SRL on student motivation, engagement, and finally on their learning outcomes.

The next Chapter explains how I identify learners' affective states and address research questions 1 and 2.

Chapter 3 Affect Detection Study

3.1 Purpose of the Study and Research Questions

This Chapter provides an account of affect detection study, which evaluates the efficacy of iMotions as an affect detection tool and the use of worked examples in SQL-Tutor. Learners' affective states were determined by analyzing their facial expressions using the iMotions software package (iMotions.com) and the Tobii eye tracker (tobii.com). Worked examples are introduced in SQL-Tutor for aiding the learners during problem-solving. This study helps address the following research questions.

Research Question 1: Does iMotions accurately identify learner emotions? There is very little evidence found in the literature which proves the reliability and validity of iMotions, notably its AFFDEX algorithm. However, in the new versions of iMotions (7 and above), the support for FACET has been discontinued in 2020. To find the suitability of the software for the research, it must be (1) reliable in identifying emotions and (2) comparable with previously established techniques (Kulke et al., 2020). In a study, Taggart et al. 2016 examined the reliability of iMotions on human respondents; however, only the FACET module was investigated without accuracy estimation.

Another study (Stockli et al., 2018) compared the FACET and AFFDEX algorithms on prototypical pictures and human respondents. This study used three libraries of photos, including IAPS (which I am using in this study) for affective state detection. The results revealed that the accuracy of iMotions on prototypical pictures is 73% using AFFDEX and 97% using FACET. When the prototypical pictures were replaced with human respondents, the accuracy dropped to 55% using AFFDEX and 57% using FACET. The only study which compared the iMotion's AFFDEX module with another facial expression technique Electromyography (EMG) is conducted by Kulke et al. 2020. The analysis based on three emotions, happiness, angry and neutral, revealed that both approaches produced similar results when respondents imitate the facial expressions of emotions. However, AFFDEX wrongly predicted neutral to angry emotion. This brief overview indicates that iMotion's AFFDEX algorithm's accuracy has not been established for human respondents in the natural environment. Therefore, it is necessary to establish the accuracy of the iMotions AFFDEX first to gain confidence in our results. The first research question handles this situation, and I expect that iMotion's AFFDEX algorithm accurately detects the emotions of human participants under realistic conditions (**H1**). (Affect and emotion will be used interchangeably in this work)

Research Question 2: Do worked examples help during problem-solving in SQL-Tutor? Existing research has not provided much evidence of learners' negative/positive affective states while interacting with different interventions in ITSs. A recent study (Borracci et al., 2020) explicitly asked students about their negative affective states while solving problems from an ITS. The authors found no significant differences in frustration, boredom, and anxiety when learners have been provided with an isomorphic worked example or just a worked example. These results are based on learners' self-reports and do not reveal the real-time affective states of learners while using worked examples or problem solving. To fill this gap, I intend to use iMotions and Tobii eye tracker to capture the real-time face expressions of learners when they solve problems. With this context, I hypothesize that worked examples will help learners improve learning and reduce the negative affective states during problem solving **(H2).**

The above research questions are addressed individually in two phases. Firstly, to establish the reliability of iMotions, the photos from the International Affective Picture System (IAPS) library are presented to participants. Their emotions for each photo were captured and analysed by iMotions and compared with the established emotion categories by Mikels et al. (2005). The comparison determined whether iMotions can correctly predict the emotion or not.

In the second phase, a version of SQL-Tutor is extended with worked examples. The use of worked examples was entirely voluntary. If participants used these examples, they were asked to complete a short questionnaire regarding the usefulness of the example. iMotions and Tobii eye tracker recorded the facial expressions and eye gaze of participants during problem-solving and convert them into emotion labels These worked examples eventually help determining whether working examples helped cognitively and iMotions determine their effectiveness in affective states regulation during problem solving with SQL-Tutor.

3.2 Experimental Design

In the first phase, basic demographic information about each participant and their level of familiarity with IAPS and SQL-Tutor was obtained using a demographic and emotional intensity questionnaire (see Appendix B). This questionnaire enabled participants to determine their emotions on a 9-point Likert scale (1 = very strongly disagree, 9 = very strongly agree). Students marked the intensity of their emotion on the scale. For instance, (Q1) *I am easily excited*, and (Q2) *I am easily angered.*

After the questionnaire, each participant sat in front of the Tobii screen, and the standard Tobii calibration was completed. During calibration, the participant was requested to track the movement of a ball on the screen. This test took 6 seconds, and the experiment started when the result was excellent. Otherwise, the participant's position was adjusted in front of the camera, and recalibration took place. The experimenter sat to the other side with a monitoring screen and watched the participant's entire face and eye gaze captured by both iMotion's camera and Tobii. This monitoring ensured that the participant's full face was captured during the session.

Forty-eight photos of different emotions were selected from the IAPS library. IAPS is a collection of 1,182 emotion-intensive-coloured photographs, which provide a standard to researchers to study emotions and attention (Lang, 2008). Those photos are grouped into eight discrete emotions: anger, disgust, fear, sadness, amusement, awe, excitement, and contentment (Mikels et al., 2005). Those photos are categorised as negative and positive based on their mean arousal and mean pleasure values. For example, photos regarded as positive have mean pleasure values between 5.00-8.34 and mean arousal values between 2.90-7.35. By contrast, the negative photos have the mean pleasure values in the range of 1.45-4.59, and the mean arousal values in the range of 2.63-7.35.

The photos were chosen from the IAPS library based on the above categorization of positive (24 photos) and negative (24 photos) emotions, with six photos representing content corresponding to each emotion. These photos were selected based on the highest mean arousal values. For example, if a positive photo A has a mean arousal value of 6.5 and positive photo B has a mean arousal value of 5.9, then photo A was selected. However, disgust-related photos were selected from the range of 4.0-5.0 mean arousal due to their disturbing content. These eight emotions were mapped to seven emotion categories provided iMotions as described in Table 3.1.

Table 3.1. *Comparison of Emotions from iMotions, and Mikels Classification*

iMotions categories	**Mikels Classification**
Anger	Anger
Disgust	Disgust
Fear	Fear
Contempt	Anger/Fear/Sadness
Sadness	Sadness
Joy	Amusement
Surprise	Awe
Joy	Contentment
Joy, Surprise	Excitement

Photos were presented according to the instructions of the Self-assessment Manikin (SAM) scale (Mikels et al., 2005) in which the participants need to report on a 9-point Likert scale (see Figure 3.1). The 48 photos were split into three blocks (with 15 seconds intermittent break), each with 16 images, two related to each emotion category. Three photos were presented for training purposes before starting the first block, and a 15 second rest time was provided before starting each block to normalize participants' facial expressions. Each photo was presented to the participant for 6 seconds, and then a self-reporting scale (Figure3.1) was shown for 15 seconds to mark their immediate emotion.

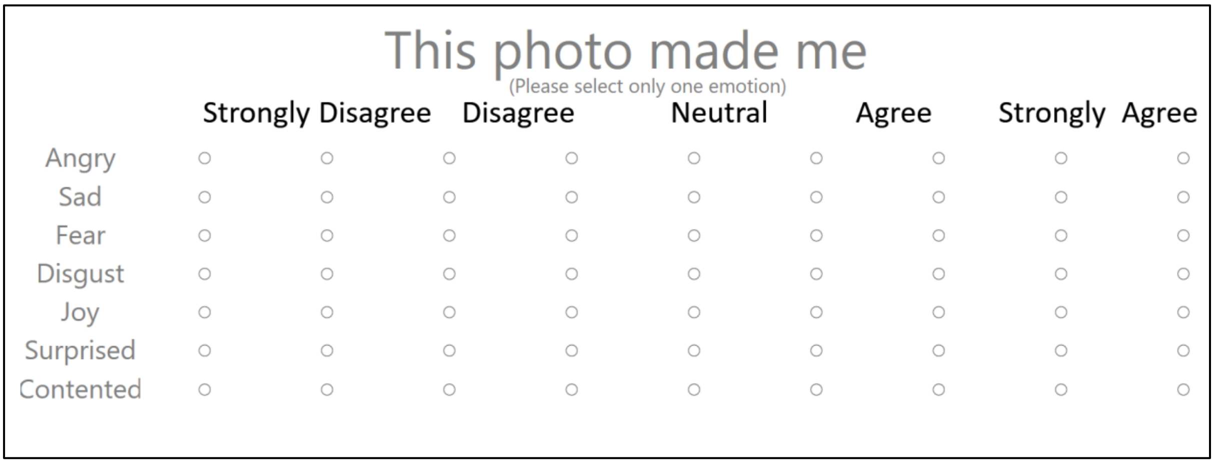


Figure 3.1. Self-reporting scale

In the second phase, the version of SQL-Tutor consisted of ten problems defined for the *CD-Collection* database, as illustrated in Figure 3.2. For the study, I added a new *Example* button, which allows the participant to view a worked example for the current problem.

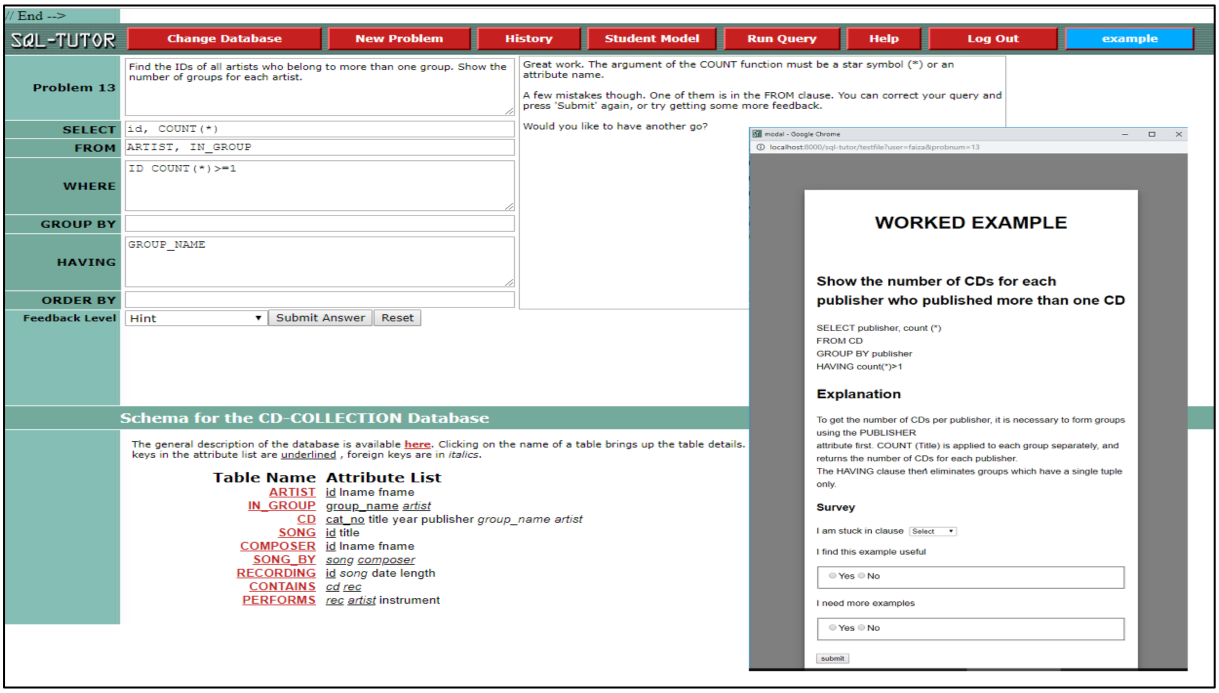


Figure 3.2. Screenshot of the worked example mode of SQL-Tutor.

For each problem, a worked example isomorphic to the problem was provided. Figure 3.2 shows a worked example, which includes the problem statement, the solution accompanied with an explanation. After the explanation, a short survey was presented. Firstly, participants were expected to specify which clause of the select statement they had difficulty with. The participants were also asked whether the example was useful and whether additional examples were needed. The questions were mandatory, and participants had to answer them to continue with problem solving. While the participants were working with SQL-Tutor, their eye gaze and facial expressions were recorded in iMotions.

3.3 Procedure

Each participant had an individual session that combines both phases of the study. At the beginning of the session, the participants provided informed consent and filled the demographic and emotion intensity questionnaire. After the calibration test, the first phase of the study started. The participant was first presented with three practice photos and their survey to familiarize them with the process. After the practice photos, the participant was allowed to rest for 15 seconds, and then the main experiment started. The participant viewed the first block of 16 photos each for 6 seconds (Figure 3.3). After each photo, the participant had 15 seconds to mark on the self-reporting scale. The participant could move to the next photo after completing the self-reporting scale. Once the first block was completed, the participant received 15 seconds of rest time, followed by the second block. This phase took 15-20 minutes to finish.

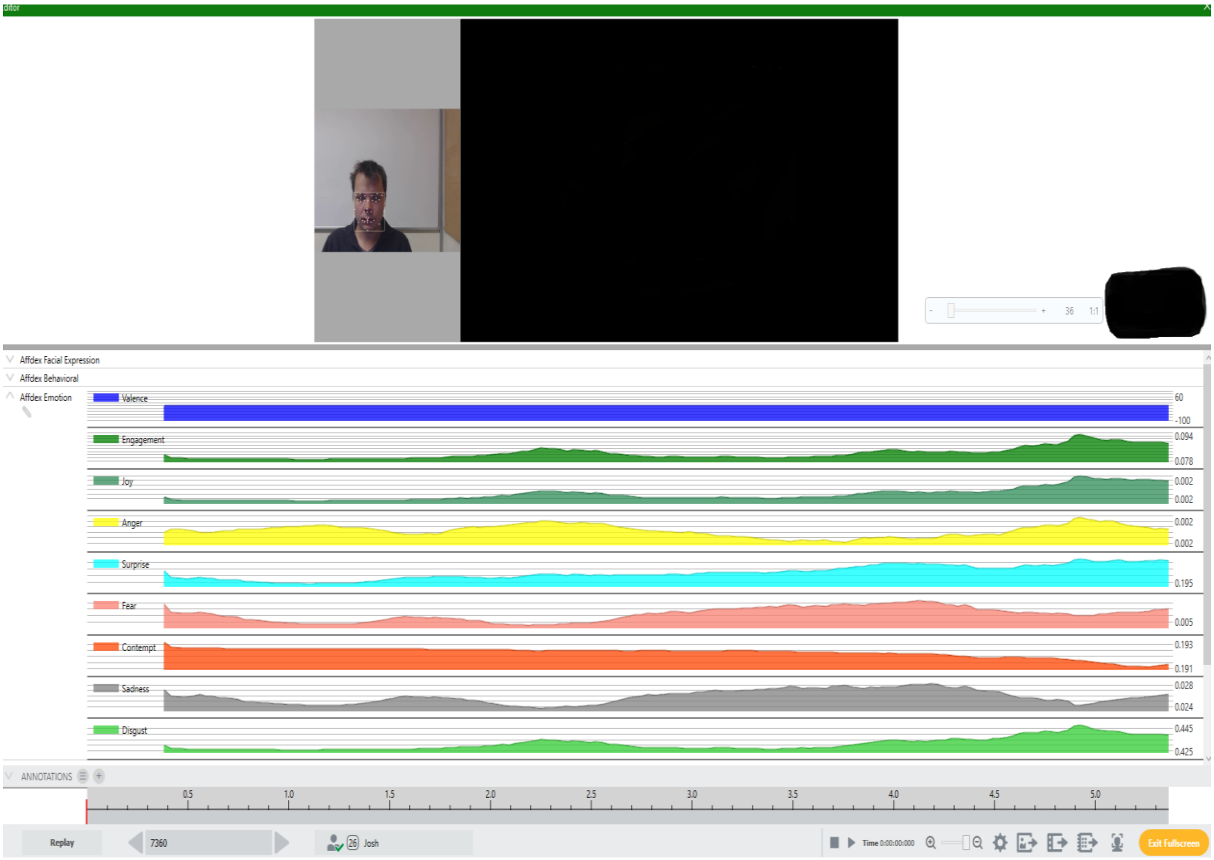


Figure 3.3. A participant viewing a photo with iMotions recording emotions.

Once the participants completed the first phase of the study, they were required to solve problems using SQL-Tutor in the second phase. The web browser took participants to the main interface of SQL-Tutor where they could select problems from the CD-Collection database. Participants were required to solve problems, and help was given in the form of examples and feedback. Whenever the participants clicked on the example button in the interface and saw the example, they needed to answer the mandatory questions at the end of the example explanation. Participants were not required to solve a specific number of problems and free to select any problem. This part of the study took up to 30 minutes to complete. At the end of the session, each participant was awarded a $20 voucher as a note of thanks. The study was conducted from February to March 2019 and approved by the Human Ethics Committee of the University of Canterbury in November 2018 (HEC 2018/47/LR-PS).

3.4 Results

The data obtained for this study comprise demographic and emotion intensity questionnaires, iMotions statistic files, and SQL-Tutor log files.

3.4.1 Demographic and Emotion Intensity Questionnaire

Ten participants aged 18 to 35 years (20% females) participated in the study. Seven of those were undergraduate, and three were postgraduate students. The demographic part of the questionnaire found that all the participants were studying computer science. Six of those participants were domestic students, while the remaining four were international students. All the participants were familiar with SQL; some (6) have worked with SQL-Tutor before the study, and no one knew about IAPS.

The findings on the emotion intensity part of the questionnaire revealed that participants experienced excitement, amusement, and disgust quickly. Particularly for negative emotions, the responses revealed that it was challenging to get fearful but strongly inclined towards disgust. However, it was rather hard for them to be easily saddened, made fearful, or angered, as illustrated in Figure 3.4. This result indicates that participants may have an awareness of their emotions, although I cannot infer the implications of their emotions into the results because of the low sample size.

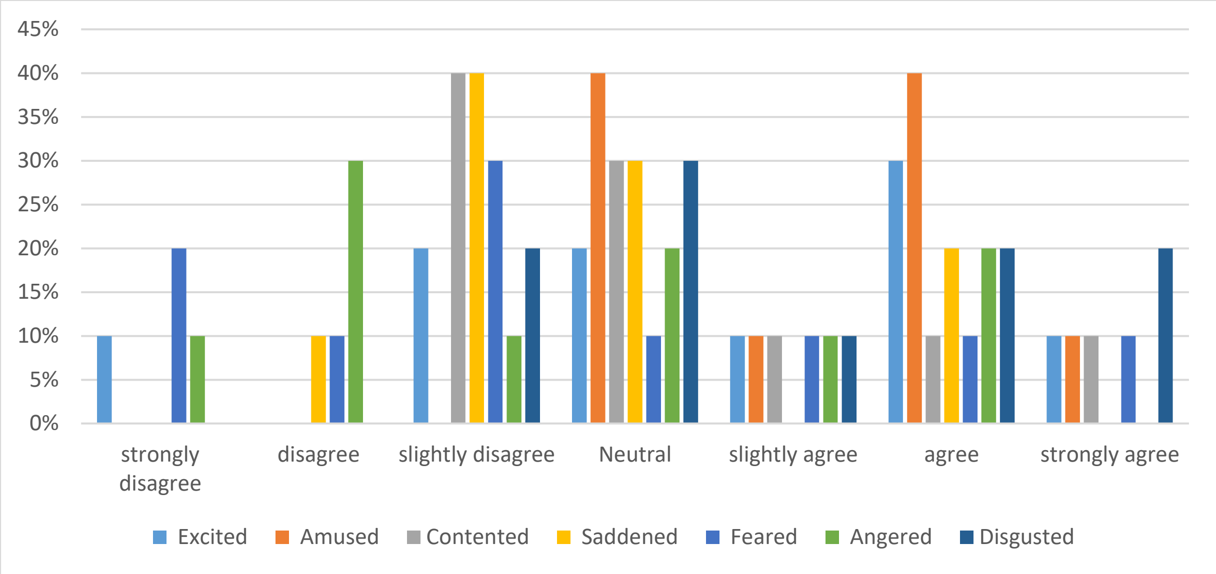


Figure 3.4. Emotion intensity questionnaire responses (N=10)

3.4.2 RQ 1: Does iMotions Accurately Identify Learner Emotions?

To analyse the emotions, I did the post-processing first. As the participants have different heights and were changing positions in front of the screen, it was essential to apply post-processing to increase the quality of results. It applied to each participant's video separately. Video was opened in the edit mode, and the facebox adjusted according to participant's height, and position and AFFDEX analysis was carried again. All the videos have sampling rate quality greater than 80%.

iMotions automatically extracts facial features from videos using the Affectiva facial expression analysis engine (also known as AFFDEX algorithm), based on the facial action coding system (Hamm et al., 2011). This coding system extracts features from images and uses a support vector machine classifier to find the most suitable emotion. AFFDEX generated probabilistic estimates for each of the seven emotions (anger, disgust, surprise, sadness, joy, fear, and contempt) based on each participant's macro-expressions (lasting 0.5-4 seconds). These were the numeric scores of emotions along with their intensity.

In the present analysis, the amplitude-based thresholding was used to find out the strongest emotion. The amplitude-based thresholding can be relative or absolute. For this study, the absolute rates in amplitude thresholding were selected, which showed the frequency with which participants displayed the emotion. Then a threshold of 50% was taken on those numeric scores, thus including only emotions whose values were greater than 0.5. The threshold value remained the same for every emotion. Due to the low occurrence of values, this threshold included the intense occurrences only.

The AFFDEX values showed seven emotions and 20 action units in percentages across a single photo. For example, a photo categorized to represent anger may have AFFDEX values of anger 90%, sadness 30%, and surprise 10%. These percentages were aggregate emotions of all the participants across a single photo. The strongest emotion was the one with the highest percentage value on the photo. That strongest emotion was compared with Mikels et al. (2005) categorized emotion for that photo, and the iMotions accuracy was determined. This process was repeated for all 48 photos.

Table 3.2. *iMotions Prediction Results on IAPS Photos*

Number of Photos	iMotions Prediction	Mikels categorization
37	Same emotions as Mikels	
4	Surprise	Fear
7	Negative emotion	Positive Emotion
48	(77% Accuracy)	

Table 3.2 shows the comparison results, which revealed that iMotions predicted the emotions of 37 (77%) photos same as the Mikels categorization. From the remaining 11 photos, the four photos were predicted as surprise rather than fear; this misinterpretation of emotion has already been discovered and discussed in (Stockli et al., 2018). Another reason for this could be that whenever a photo was presented to the participant, the surprise emotion was highest initially. If the photo content was not fearful enough for that participant, surprise remained the highest emotion. The remaining seven photos showed the content of positive emotions (awe, excitement, amusement), which iMotions found negative (fear, disgust, sadness). The content analysis shows that these photos were either related to river rafting or

mountain biking which could be a fearful or sad adventure for some participants. In addition, the arousal values of these photos were very close to the lower limit in range. This is not an unexpected result because one recent study (Kulke et al., 2020) mentioned that iMotions predicted neutral to negative emotions. However, the possible reasons for this wrong prediction in this study could be because these photos were presented to the participants during the last block, preceded by a negative emotion photo predicted accurately by iMotions. Thus, although iMotions used the automatic baseline correction method, there was a possibility that the photos might become less interesting for the participants at the end of the experiment.

The self-reported scale revealed almost the same emotions as the iMotions predicted. Importantly, none of the results were relevant to the above findings. Taken together, these findings suggest that iMotions successfully (approx. 80%) predicted the emotions in the lab settings under specific circumstances and support **H1**. Although emotion prediction is not perfect, having an ITS which accurately identifies emotions up to 80% and provides strategies to regulate emotions is far better than a system that does not identify or regulate emotions at all. Suppose the prediction refers to the broader categories such as negative and positive emotions rather than specific emotion categories (anger, sadness, happiness, so on). In that case, iMotions may yield more accurate results.

3.4.3 RQ2: Do Worked Examples Help During Problem Solving in SQL-Tutor?

This Section reports the worked examples' effectiveness during problem-solving, the study's second phase. The data files from the SQL-Tutor logs and iMotion’s recorded videos were collected for analysis.

The participants varied in their expertise in the use of SQL-Tutor. Four participants never used SQL-Tutor before the study; two participants used SQL-Tutor a lot, while the remaining three had limited experience with the system, as revealed by the demographic and emotion intensity questionnaire. Table 3.3 shows how many participants attempted and completed each problem, asked for examples, and the time spent on the problems/examples. The Example column specifies the number of participants who viewed examples. The participants mostly solved the problems in the provided order, from the easiest to the hardest.

On average, participants attempted six problems (SD = 1.89). The four most difficult problems were attempted less often, and no one completed problems 9 and 10. The participants completed 62% of the attempted problems, and viewed examples in 59% of the cases. For problems 1-5, as the problem complexity grows, the example use increases. In problems 2, 4, 5, and 10, participants viewed the examples more than once. When they viewed examples for the first time, they spent, on average, a minute viewing them. Upon the second and third viewing, this time decreased to 10-20 seconds only. The average time per example was proportional to the average time on problems, except for the example for problem 1. When the example is longer, this might be because the first-time participants encountered the examples interface, with which they were not familiar previously. The participants' system logs show that participants' use of examples in one problem cannot predict their use of examples in the subsequent problems.

Table 3.3. *Problem, Example and Feedback Use. Time is reported in minutes*

Problem	**Attempted By**	**Completed By**	**Time / Problem (SD)**	**Example**	**Time / Example (SD)**	**Feedback**
1	9	9	1.48 (0.95)	2	40 (7.07)	1
2	8	7	4.18 (2.02)	7	41 (9.72)	3
3	9	7	1.6 (1.34)	1	25 (0)	2
4	10	6	5.58 (3.18)	7	72 (30)	4
5	9	4	6.5 (3.79)	6	46 (54)	5
6	5	3	3.46 (2.62)	3	38 (7.63)	2
7	2	1	2.1 (0.14)	2	16.5 (12)	0
8	3	1	2.05 (0.07)	2	26 (20)	1
9	3	0	2.2 (0.52)	3	30 (16)	0
10	3	0	7.3 (5.23)	3	43 (15.2)	0

The Feedback column of Table 3.3 shows the number of participants who have explicitly required specific levels of feedback (such as hint, partial/complete solution) while solving problems. More participants have used feedback for the easier problems (1-5) than for the rest of the problems. This trend was opposite to how participants used examples. For problems 2, 4, and 5, more participants requested feedback; simultaneously, the time for completing those problems was higher than for other problems, possibly because the participants spent some of the time trying to understand the feedback they received. An interesting finding was that very few participants used both examples and feedback while solving a particular problem. There were only seven cases where both feedback and examples were used, mainly in early problems. This indicates that participants were more inclined towards examples than feedback, particularly in complex problems.

Table 3.4 shows how many participants viewed each example, whether they found the examples useful, and whether they wanted additional examples for a particular problem.

Table 3.4. *Participants' Opinions on Examples*

Example	**Viewed by**	**Useful**	**More examples**
1	2	1	2
2	7	7	1
3	1	1	1
4	7	5	2
5	6	4	2
6	3	2	0
7	2	1	1
8	2	2	0
9	3	3	0
10	3	2	1

In 78% of the cases when examples were viewed, the participants found them helpful, and in 36% of these cases, they needed more examples. For the complex problems (problems 7-10), when the completion rate was low (below 20%), the participants found the examples very useful (80% of the cases), even when they have not completed those problems. This shows that regardless of success in problem-solving, the participants found the examples helpful.

Please note that this study was voluntary, and therefore there was no need for the participant to complete all problems. There were five examples where participants agreed on both usefulness of example and required more examples. Analysing these examples revealed that four of those cases involved a single participant, who used examples extensively.

3.4.3.1 *Eye Gaze Analysis of Worked Examples*

I analysed the eye-tracking data to determine how the participants read worked examples. The Area Of Interest (AOI) was defined to cover the whole example (i.e., title, solution, and explanation). The metrics included in the eye-tracking analysis were: (1) Time in AOI, the total time spent looking at the AOI, (2) Visits, the number of times the participant's eye gaze returns to the AOI, (3) Fixation count, showing the total number of fixations within the AOI (4) Duration of the first fixation on the AOI, and (5) Average fixation duration in AOI.

Table 3.5 shows the metrics for the ten examples (including multiple viewings) averaged over all participants who viewed those examples. The time in the AOI column provides the average time spent by participants while examining a particular example. This includes all viewing an example (please note that the maximum number of times a participant viewed any example was 3). The average number of visits to the AOI seems to increase as problems becoming more complex. As the number of example steps grows in later examples, the participants looked more often towards the problem-solving area and schema.

Table 3.5. *Averages (SD) for Eye Tracking Metrics. Times reported in Seconds*

Example	**Time in AOI**	**Visits**	**Fixation count**	**First fixation duration (s)**	**Fixation duration**
1	40 (2.8)	13 (1.4)	43 (6.3)	.2 (.07)	.19 (0)
2	33 (18.5)	9 (9.06)	38 (30)	.14 (.05)	.21 (.030)
3	24 (0)	5 (0)	7 (0)	.24 (0)	.23 (0)
4	59 (30)	27 (25)	132 (127)	.25 (.073)	.24 (.03)
5	59 (61.9)	18 (18.5)	111 (116.9)	.2 (.094)	.24 (.05)
6	51 (20.5)	23 (2.12)	82 (21)	.23 (.063)	.2 (.04)
7	23 (0)	13 (0)	51 (0)	.22 (0)	.25 (0)
8	49 (24.9)	32 (14)	107 (94)	.46 (.37)	.24 (.05)
9	34 (4.7)	10 (6.5)	55 (32)	.19 (.053)	.22 (.04)
10	47 (16.3)	20 (9.5)	107 (52)	.31 (.31)	.22 (.02)

This was also evident from gaze plots, such as the one shown in Figure 3.5. This gaze plot was generated from the eye gaze data for a participant viewing the example for problem 2 for the first time. Participants read the text of the example and the given solution and compared them to the text of the current problem and the solution they were working on. The eye gaze moved between these two areas often. From the gaze plot, it was also evident that the participant was examining the database schema, which was necessary to understand the provided solution. However, the participant did not spend much time on the worked example explanation.

The average fixation count shown in Table 3.4 was highest in examples 4, 5, 8, and 10. As mentioned above, examples 4 and 5 were viewed by 7 and 6 participants, respectively, and they were viewed multiple times. The highest fixation count shows that these participants did not just glance over those examples, but studied them thoroughly, not only the first time but also for the second or third viewing. The high fixation count on more complicated examples strengthens the above findings that more fixations were recorded as the number of example steps grows.

The average duration of the first fixation on the AOI also shows the same trend. These two results and the number of visits provide evidence that participants were reading, thinking about examples, and comparing the example steps with their solutions. However, a low completion rate on these problems required that examples of these problems must be more than one or should be more detailed and presented to make the step easier to understand.

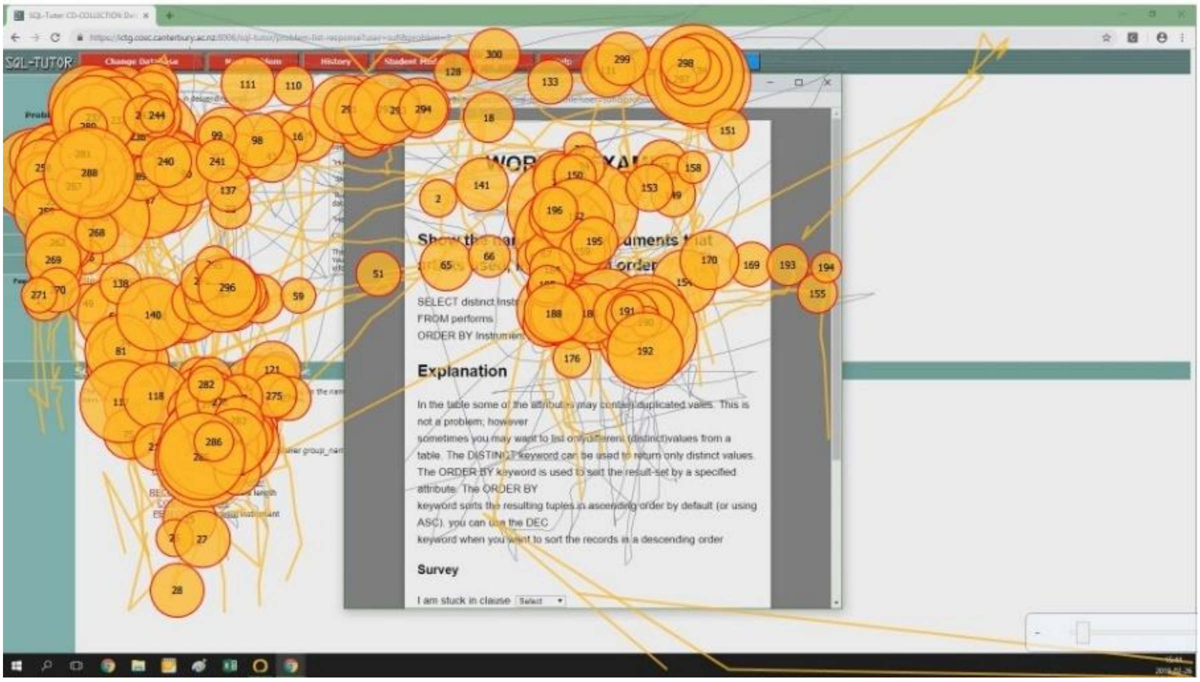


Figure 3.5. Gaze plot for the first viewing of example 2

The gaze plot analysis revealed that participants fixated on the example text, solution steps, and explanations in that order. Most fixations were recorded on the example text and solution steps and less on explanations. Two participants did not read the explanations of any examples they viewed; this kind of behaviour might be the consequence of low motivation to complete problems due to the voluntary nature of the study. These results were consistent with the findings from the study conducted by Najar et al. (2015), who observed that the participants visited explanations less than the solution. Four participants, however, read explanations for all examples they examined. Those four participants attempted and solved more than half of the available problems, which shows their motivation. Not all participants read explanations points to the need to add more interactivity to worked examples. One way to achieve that would be to add self-explanation prompts.

3.4.3.2. Affect Analysis of Worked Examples

iMotions analysed facial expressions and reported the values of seven emotions: anger, sadness, surprise, disgust, joy, contempt, and fear, based on Ekman's (1999) categorization of emotions. However, these are general emotions, not the emotions specific to learning (Baker et al., 2010; Craig et al., 2004). There have been several sets of emotions identified in learning situations; for example, Woolf et al. (2009) considered joy, anger, surprise, and fear as learning emotions. Disgust, sadness, and contempt considered irrelevant for learning (Baker et al., 2010; Craig et al., 2004; Pekrun, 2006; Woolf et al., 2009). Woolf et al. (2009) also suggested mappings between Ekman's basic emotions to learning-related emotions: joy mapped to excitement, anger mapped to frustration, surprise mapped to boredom, and fear mapped to anxiety.

In line with the research mentioned above, I considered anger, joy, fear, and surprise. I additionally included engagement (calculated as separate measure in iMotions similar to the other emotions), which was crucial for learning (Craig et al., 2004; D'Mello et al., 2007). At the beginning of the session, the emotion levels were low across all emotions. At the start of each problem, the dominant emotion was surprise, and once the problem was solved, the dominant emotion was joy. When participants received feedback from SQL-Tutor (upon submitting their solutions), the level of surprise was higher. In those situations, when participants were able to solve the problem after receiving feedback, the level of joy was again increased. However, if they could not solve the problem, I noticed higher levels of anger, showing the participants' frustration.

Table 3.6. *Affective States Before, During or Immediately After Viewing Examples*

Participant	**Segment**	**Anger**	**Fear**	**Surprise**	**Joy**	**Engagement**
1	Before example	0.5	0	1	0	0.5
	During example	1	0	0	0	1
	After example	0.5	0	0.5	0	0.75
2	Before example	0.33	0	1	0.33	0.36
	During example	0.16	0	0.67	0.16	0.58
	After example	0	0.16	0.83	0.16	0.48
3	Before example	0.85	0.42	0.85	0.14	0.89
	During example	0.14	0.21	0.5	0.42	0.71
	After example	0.57	0.21	0.78	0	0.89
4	Before example	0	0.25	0.75	0.25	0.12
	During example	0	0	0	0.25	0.19
	After example	0	0	0.5	0.12	0.25
5	Before example	0	0.5	0.5	0	0
	During example	0	0.25	0.5	0.5	0.5
	After example	0	0	0	0	0.5
6	Before example	0.1	0.04	0.7	0.2	0.15
	During example	0	0.14	0.5	0	0.4
	After example	0	0	0.5	0.2	0.5
7	Before example	0.58	0.083	0.83	0	0.33
	During example	0.33	0.083	0.67	0	0.91
	After example	0.25	0	0.67	0	0.87
8	Before example	0.77	0	0.667	0	0.36
	During example	0.35	0.11	0.33	0.11	0.94
	After example	0.44	0	0.61	0.11	0.94
9	Before example	0.6	0.17	0.92	0	0.78
	During example	0.28	0.07	0.67	0.14	0.71
	After example	0.14	0.14	0.7	0.14	1
10	Before example	0.17	0	0.62	0	0.25
	During example	0	0	0.25	0.25	0.81
	After example	0	0	0.5	0.12	0.87

Another interesting finding was when the participant asked for examples. The emotions were analysed for three different time intervals: (1) one minute before example use, (2) during example use, and (3) one minute after example use. Table 3.6 reports the average of each emotion observed over all examples used by a participant. These values were in the range of 0-1, where 0 shows the complete absence of emotion (see Table 3.6).

Firstly, one minute before participants asked for examples, the dominant emotions were anger and surprise, suggesting that participants asked for examples when they were frustrated. Engagement increased, and surprise decreased during or after working with examples. Fear was the least detected emotion; for four participants, it decreased while and after working with examples, but for three participants, fear slightly increased when working on examples. I observed that joy increased for five participants when they were working with examples, and immediately after, those situations occurred when the participants were able to complete problems after viewing examples.

On the other hand, if the example did not help the participant solve the problem, I observed increased values for anger and surprise. In some of those cases, participants asked the example for the second time, and after that, abandoned the problems. This was consistent with findings reported in the literature (Baker et al., 2010), showing that frustration may lead to boredom, in which case learners lose interest in learning activities.

In short, participants asked for examples when the levels of anger (i.e., frustration) and surprise (i.e., anxiety) were elevated. Working with examples reduced such negative emotions and increased joy. After viewing examples, when participants turned again to problem-solving, the intensity of negative emotions was low but gradually increased if they could not solve the problem. The level of engagement increased for all participants during and after viewing examples. Therefore, examples positively impact participants' affective states, which will help learners with SQL-Tutor and support **H2**.

3.5 Discussion

This study aimed to establish the reliability of iMotions as affect identification software and to understand the behaviour of participants when using examples in SQL-Tutor example mode, Tobii eye tracker, and the iMotions.

The results showed that iMotions could be used as a reliable system to detect human emotions under the AFFDEX algorithm. There is a little evidence in the research about the validity and reliability of the AFFDEX algorithm in natural settings. One study (Stockli et al., 2018) which established the reliability of AFFDEX algorithm on human respondents reported 55% accuracy achieved through iMotions. The current study provides evidence that iMotions can successfully detect emotions in human participants with 77% accuracy. Although, I have achieved a comparatively better accuracy of detecting emotions through iMotions.

Furthermore, the current study also shows that participants used examples extensively in their problem solving, particularly when complexity increased. Both feedback and example viewing increased the total time to complete a problem. Most participants agreed with the usefulness of example, whereas a few required extra examples. Similarly, the eye gaze analysis revealed that participants tried to understand example structure by comparing examples with their solutions. Lastly, the positive impact of examples on participants' emotions is evident from affect analysis; examples reduced participant's negative emotions and increased engagement, and up to some extent, joy. The effectiveness of worked examples has been discussed in the literature (Najar et al., 2014), but its efficacy in reducing negative emotions has not been discussed so far. Therefore, these results are the initial evidence of worked examples supporting learners' cognitive and affective states.

With all the positive results of using examples, the attempt and completion rate for complex problems was meager. Generally, participants attempted the problem a few times; if they could not complete the problem, they tried to understand/use more feedback or examples. They again tried to complete the problem but only up to few more attempts and then switched to another problem. If a participant remained unsuccessful for consecutive two or three problems that demotivated them, they were more likely to abandon the SQL-Tutor. However, this study was voluntary; therefore, most participants left the system after attempting 5 or 6 problems.

These findings on the worked examples might not help further in this research project because the required strategy for providing adaptive worked examples during problem solving could not be devised due to the low sample size. However, this has raised the need for motivation and affect improvement strategies in SQL-Tutor, which can help users learn more and continue using SQL-Tutor and encourage them to complete complex problems.

Chapter 4 Gamification Study

4.1 Purpose of the Study and Research Questions

Gamification is defined as incorporating game elements in a non-gaming environment. The use of term *gamification* in learning environments started a decade ago and positively affected student engagement and motivation (Koivisto & Hamari, 2019). However, its application in intelligent tutoring systems has not been investigated much. In the affect detection study, I observed that when students could not complete problem/s in the first few attempts, they became frustrated, and their motivation decreased; often, in such situations, students either logged off or started switching between problems and playing with the system. In low motivation and boredom, even if the students can solve the problems, they tend to leave the system after a few attempts. On the other hand, students who completed problems in a timely fashion enjoyed working with the system and tried to solve complex problems. However, not completing problems was the most significant disruption in enjoyment.

In light of the above findings, this study aims to determine the impact of gamified SQL-Tutor on students' engagement, motivation, and learning outcomes. Based on this goal, I made the following research questions and related hypotheses.

Research Question 3: What are the effects of gamification on learning? I expect that the experimental group (using gamified features) will be more engaged with SQL-Tutor by spending more time-on-task than the control group **(H1)** and that time-on-task will be positively correlated with their learning outcome **(H2)**. In addition, I expect that badges will have an indirect effect on learning outcomes by influencing time-on-task **(H3)**.

Research Question 4: Do students with different levels of prior knowledge react differently to gamification? Research shows that students with higher background knowledge are more engaged and motivated to learn than students with insufficient background knowledge. I expect that prior knowledge will affect the time students spend in SQL-Tutor and that badges will moderate this relationship **(H4)**.

Research Question 5: What is the effect of gamification on student motivation? As discussed previously, some studies found that gamification increases motivation. Therefore, I expect the experimental group students to report increased self-efficacy levels, perceived competence, and topic interest after the study **(H5)**. In addition, higher motivation will lead to higher learning outcomes **(H6)**.

First, to answer these questions, I have modified the standard version of SQL-Tutor and added gamification through badges. The details of gamifying SQL-Tutor are discussed in Section 4.2. Second, the standard and gamified versions of SQL-Tutor were allocated to the control and experimental groups, respectively. Third, the procedure Section (4.3) provides a complete account of how the study was conducted. Finally, the results obtained from the study are discussed in Section 4.4, and Section 4.5 concludes the Chapter.

4.2 Experiment Design

The experimental design of the study is elaborated in the following sections.

4.2.1 Game Elements and Learning Behaviours

My approach to gamifying SQL-Tutor is based on the theory of gamified learning (Landers et al., 2017). The theory of gamified learning specifies the causal relationship between gamification and learning. This theory elaborated that gamification does not directly impact learning outcomes; however, for gamification to apply successfully, the learning behaviour must be influenced and affected. This modified learning behaviour, in turn, yields the learning outcomes. I selected three game elements from the nine categories discussed by Landers et al. (2017): goals, assessment, and challenges. Challenges can grow competition in students, either in class standing or skill achievement. Munshi et al. (2018) reported that students might become bored or frustrated if they are not challenged enough. Therefore, if done correctly, then complex problems in the form of challenges can be helpful to retain interest.

Goals are also considered a form of challenge; however, the goal-setting theory (Locke & Latham, 1990) states that goals can motivate students when they are SMART (specific, measurable, achievable, realistic, and time-bound). I selected the goals accordingly: they have only one condition (specific), can be measured through completed problems (measurable), are achievable, realistic, and can be achieved within the 4-week study period (time-bound). The difference between challenges and goals lies in the complex and hard-to-achieve nature of challenges. Goals do not consider specific problems or a student's current knowledge. However, challenges consist of problems that are higher in complexity than the ones students have solved. Assessment is another crucial behaviour that is necessary for self-monitoring.

I implemented goals, assessments, and challenges in SQL-Tutor via different types of badges, presented in Table 4.1. First, the goal-setting behaviour is supported by fixing daily and weekly goals stated as winning criteria for badges. Next, the self-testing or assessment behaviour is addressed by providing a quiz. Finally, challenges are implemented via several badges and daily challenges, which consist of complex, unsolved problems. I hypothesize that all these game elements would influence time-on-task, which has been shown in many studies to influence learning outcomes (Landers et al., 2014; Denny et al., 2018).

4.2.2 Badges

In this study, the approach to design badges was adopted from the work presented by Denny et al. (2018), who defined15 different badges divided into three categories for motivating the questioning answering behaviour. However, in this study, I designed 13 badges to increase learners' time-on-task with SQL-Tutor. The badges are categorized into three groups: primary, classic, and elite. The purpose of primary badges is to grab the students' attention at the early stage of using SQL-Tutor, such as awarding a badge for solving the first problem or solving a problem using a complex clause (group by). This category also includes the Activist badge, which discourages the use of "complete solution". This badge checks that the student solved the problem independently rather than copying the complete solution provided by the system.

Table 4.1. *Definitions of Badges and the Relevant Learning Behaviours*

Group	Badge	Criterion	Behaviour	Earned By
Primary	Go getter	Completing the first problem	Goal setting	100%
	Highflyer	3 problems in one session	Goal setting	100%
	Achiever	5 problems in a day	Goal setting	100%
	Activist	5 problems without complete solution	Challenge	16.66%
	Leader	problem with the "Group by" clause	Challenge	16.66%
Classic	Energy house	6 problems in a row	Goal setting	100%
	Scholar	5 problems/day for 5 consecutive days	Goal setting	2.38%
	Fireball	10 problems in one day	Goal setting	92.80%
	Champion	First daily challenge	Challenge	7%
Elite	Genius	Attempting the quiz	Self-testing	38.09%
	Human dynamo	5 problems/day for 10 days	Goal setting	0%
	Einstein	5 daily challenges over 2 weeks	Challenge	0%
	Live-Wire	5 problems per day for 20 days	Goal setting	0%

The classic group contains four badges, which emphasize regularly practicing, for example, completing five consecutive days and solving daily challenges. The last group, elite badges, consists of four badges, and their primary purpose is to keep engaging the student with SQL-Tutor over a more extended period. In this category, badges are awarded when the student completes five problems every day for ten days or solves five daily challenges in two weeks. The last badge was awarded to those extraordinary students who completed five problems every day for 20 consecutive days.

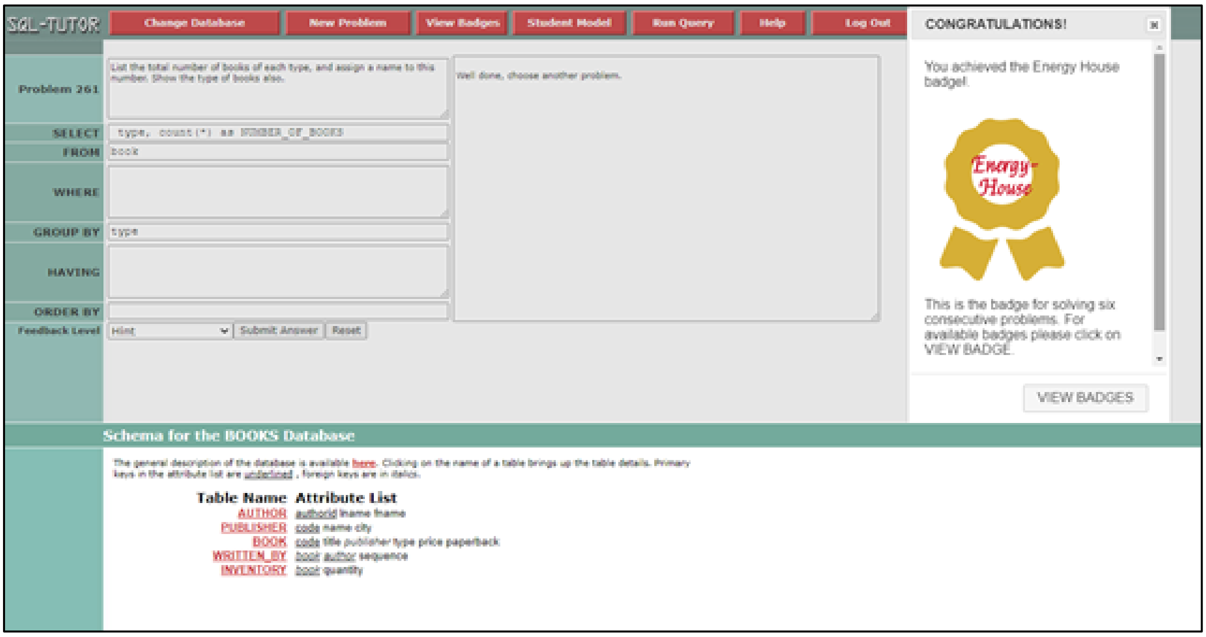


Figure 4.1. Notification of winning a badge

When the student fulfills the condition for a badge, they receive the notification about that badge immediately, as shown in Figure 4.1. Students can view all the badges awarded to them on the badge page, showing the badges that have not been achieved yet. For the study, I modified the OLM page to show the following badge the student could achieve, as shown in Figure 4.2.

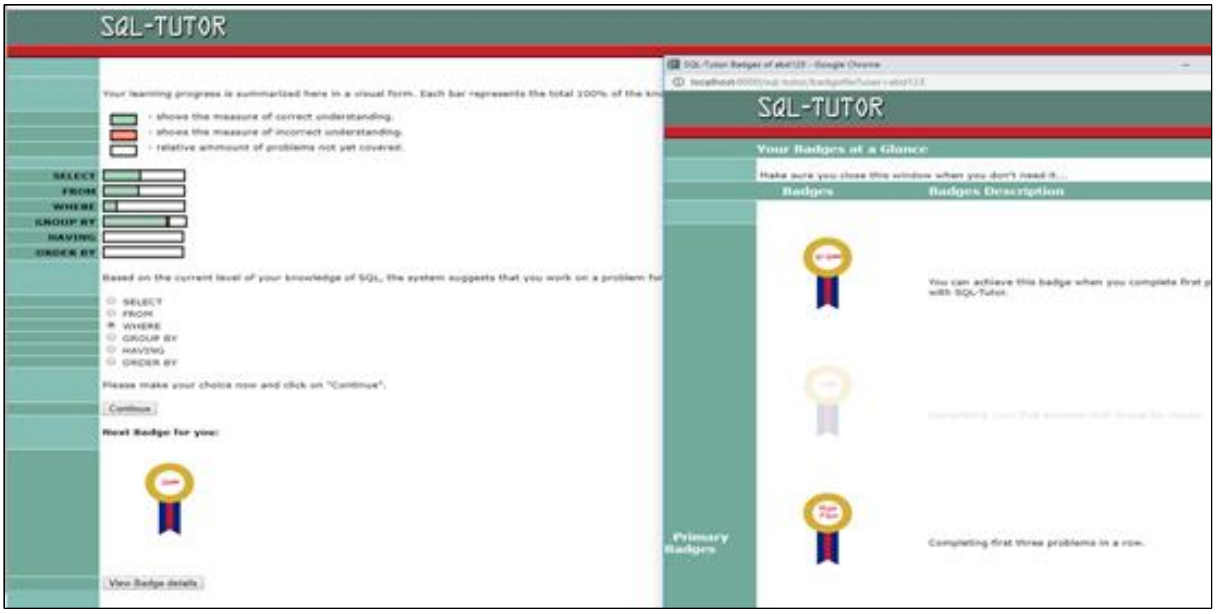


Figure 4.2. The OLM page, illustrating the next badge (left); the badge page (right)

4.2.3 Daily Challenges

Daily challenges are presented to students once they achieve all primary badges, as shown in Figure 4.3. A daily challenge consists of three problems, selected adaptively based on the student model. The problems selected for a daily challenge need to be challenging for the student. SQL-Tutor summarizes the student's learning progress using the student level, which ranges from 1 to 9. Problems in SQL-Tutor also have a complexity level (defined by the teacher) ranging over the same scale. Therefore, the problems selected for the daily challenge are previously unsolved problems, which satisfy two conditions: 1) their level of complexity is equal to the current student level or one level higher, and 2) these problems require the clauses of the SELECT statement that the student needs to practice (as per the student model). Each day, the daily challenge is presented to the student upon logging in and is also available on the problem-selection page. Two badges (*Champion* and *Einstein*) are awarded when the student completes the first daily challenge or completes five daily challenges over two weeks.

Figure 4.3. Introduction to SQL-Tutor (left) and daily challenge (right)

4.2.4 Quiz

I also developed a quiz for students to self-test their knowledge in the middle of the study session. The quiz consists of seven multiple-choice questions and two true/false questions of the same type of questions used in the pre/post-test (see Appendix B). The *Genius* badge is awarded for attempting the quiz, independently of the score achieved. Students received a notification when the quiz was available in the system, as shown in Figure 4.4. When the student completes a quiz, the scores are shown immediately to reflect on his/her knowledge. Thus awarding a badge on attempting the quiz motivated students to maximize their self-testing abilities.

Figure 4.4. Quiz notification in SQL-Tutor

4.2.5 Survey and Questionnaire

I have used a motivation survey in this study to capture the separate effects of various constructs of motivation under gamification. Survey 1 (see Appendix B) contained questions on their previous experiences with gamification, as well as questions on self-efficacy, perceived competence, and topic interest. The survey was adopted from (van Harsel et al., 2019). Self-efficacy was determined by asking the students about the extent of their confidence in writing SQL queries on a 7-point rating scale ranging from 1 (‘not at all confident’) to 7 (‘very confident’). Three items measured perceived confidence: “I feel confident in my ability to learn SQL”, “I am capable of learning SQL querying” and “I feel able to meet the challenge of performing well in SQL”. I reword the ‘course’ in the first and third statements with the ‘SQL querying’ and with ‘SQL’ in the second statement. Participants rated with the same 7-point rating scale from 1 (‘not at all true’) to 7 (‘very true’) to the extent the item applied to them. The adjusted scale had good reliability with my data (Cronbach alpha = .88). Survey 1 also contained seven items on topic-interest in which I referred to ‘SQL Querying’ instead of the original context. The reliability of these items was good, with Cronbach alpha = .83. Survey 1 was administered with the pre-test, and survey 2, which was the same as survey 1, was administered with the post-test.

Another survey (Survey 3), for determining the students’ experience about badges, daily challenges and quiz, was presented after the study. The survey was adopted from the student perception survey by Denny (2013). Survey 3 was provided to students in both groups with different numbers of questions. For the experimental group, the survey consisted of eight questions; the first four questions asked the feedback on the extent to which the badges motivated them, engaged them, or enjoyed working with badges and their intention to use badges in the future. The next two questions were related to daily challenges to determine if students found daily challenges useful and exciting and their future intentions about daily challenges. The last two questions were about the quiz: whether students found the quiz useful and if they would like to see quizzes in future versions of SQL-Tutor. These two questions were the only questions given to the control group. Table 4.2 presents the survey questions. The responses to these questions were recorded on a 5-point Likert scale, from ‘strongly disagree' to 'strongly agree'. This survey was optional and made available to students immediately after the lab test.

Table. 4.2. *Survey 3 for both Experimental and Control Groups*

Q1	Badges motivated me to participate more than I would have otherwise	Experimental group
Q2	I found being able to earn "badges" increased my enjoyment of using SQL-Tutor.	Experimental group
Q3	I would prefer not to see "badges" in SQL-Tutor".	Experimental group
Q4	The badges awarded for solving problems motivated me to solve more problems than I would have otherwise.	Experimental group
Q5	I found daily challenges useful and exciting.	Experimental group
Q6	I prefer to have daily challenges in SQL-Tutor in future.	Experimental group
Q7	I found quizzes useful and enjoy attempting quizzes.	Experimental and control group
Q8	I prefer to have quizzes in SQL-Tutor in future.	Experimental and control group

4.3 Procedure

The participants were recruited from the 198 students enrolled in the second-year course on relational database systems at the University of Canterbury in 2019. Before the study, the students were introduced to SQL in lectures and had two lab sessions, in which they created tables and performed basic SQL queries using Oracle. All enrolled students were randomly allocated to the control group (using the standard version of SQL-Tutor) or the experimental group, who used the gamified version. The students used SQL-Tutor for the first time in a lab session. The use of SQL-Tutor was voluntary; the students did not receive any course credit for solving problems in SQL-Tutor. I obtained informed consent from 77 students (25% female, 62% male, 13% not specified); 42 in the experimental group and 35 in the control group.

The study lasted for four weeks. When students logged into SQL-Tutor for the first time, they received the pre-test and Survey 1. The students could use SQL-Tutor whenever they wanted. The quiz was given at the end of the second week of the study to both groups.

The pre/post-test and the quiz were of similar complexity; each contained seven multiple-choice questions and two true/false questions (worth one mark each). The post-test and Survey 2 were administered at the end of the fourth week. A major piece of the course assessment was the lab test focusing on SQL, worth 20% of the final grade, administered two days after the post-test. After the lab test, the students were invited to complete Survey 3.

4.4 Results

Table 4.3 presents the summary statistics of the study. The average score on the pre-test was 58.73% (SD=26.05). The students interacted with SQL-Tutor on 3.39 days (referred to as *Active Days*) over four weeks (SD = 2.69, min = 1, max = 12), spending 260 min (min = 41, max = 1,441, SD = 243) in the system. During that time, the students solved an average of 37.47 problems (SD = 34.74, min = 3, max = 204). Only 28 students completed the post-test; I believe the reason for the low completion rate was that the post-test was not mandatory. In addition, the post-test was given to the students only two days before the lab test. The average score on the post-test was 69.05% (SD = 25.9), and for the lab test, it was 60.83% (17.07). In addition to defining queries, which students practiced in SQL-Tutor, the lab test covered other SQL topics, and therefore the lab test cannot be considered the direct learning outcome. For those reasons, I use the student level (*slevel*) at the end of the interaction with SQL-Tutor to measure students' learning. The average student level was 3.56 (SD = 1.66, min = 1, max = 8). In the experimental group, 66% of students reported having used some form of gamification before the study, compared to 57% of the control group participants.

Table 4.3. *Summary Statistics of SQL-Tutor Usage*

	Mean	**Standard Deviation**
Pre-test %	58.73	26.05
Active Days	3.39	2.69
Time-on-task (min)	260	243
Solved Problems	37.47	34.74
Student level	3.56	1.66
Post-test % , n=28	69.05	25.9
Lab-test%	60.83	17.07

4.4.1 RQ 3: What is the Effect of Gamification on Student Learning?

Table 4.4 presents statistics for the two groups. There was no significant difference on the pre-test scores of the two groups, showing that the students had comparable levels of pre-existing knowledge. The experimental group spent more time on task, had more sessions, attempted and solved more problems, and attempted more complex problems in SQL-Tutor in comparison to the control group, although the differences are not statistically significant. Therefore, my hypothesis **H1** is not supported. There was also no significant difference between the groups on the number of active days, student levels, the post-test and lab test scores. The possible reasons for no significant difference are low sample size, design of badges (which did not motivate or entice students), and less interest in games and gamification. These results prompted me to investigate the effects of gamification on learning of the experimental group, therefore, H2 and H3 are evaluated.

Table 4.4. *Summary Statistics of SQL-Tutor Usage: Mean (SD)*

	Experimental (42)	**Control (35)**
Pre-test %	59.52 (24.02)	57.78 (28.62)
Time-on-task (min)	288.40 (302.02)	225.94 (143.44)
Sessions	7.29 (7.84)	6.11 (4.49)
Active Days	3.33 (3.09)	3.46 (2.13)
Attempted problems	42.26 (42.75)	37.34 (26.94)
Solved Problems	39.33(40.99)	35.23 (25.72)
Max Problem Complexity	6.95 (1.78)	6.71 (2.02)
Student level	3.31 (1.62)	3.86 (1.68)
Post-test %	n = 17, 67.97 (26.32)	n = 11, 70.71 (26.42)
Lab test %	60.43 (16.49)	61.31 (17.97)

To evaluate H2**,** I regressed the student level on time-on-task. The time-on-task strongly predicts the student level ($\beta = .536$), and was statistically significant ($t = 5.5$, $p < .001$). Variance in student level explained by time-on-task was 28.7%. Therefore, hypothesis **H2** was supported.

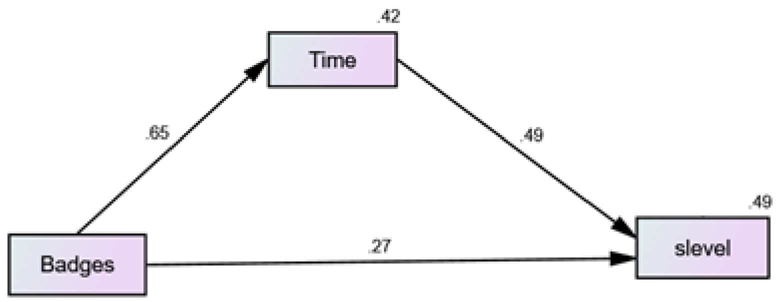


Figure 4.5. The mediation model, with standardized coefficients

To evaluate H3, I used the data for the experimental group only. I analysed the mediation effect using the Process macro, version 3.5 software for SPSS (Hayes, 2017), with the student level as the dependent variable. Figure 4.5 shows the standardized regression coefficients for the mediation model. The direct effect of badges on the student level is not significant ($p = .08$), but the significant relationship in this first step is not a requirement for mediation (Shrout & Bolger, 2002). The direct effect of badges on time is significant ($p < .001$), as is the direct effect of time on the student level ($p < .005$). The indirect and total effects in the model are tested using bootstrap samples and 95% confidence intervals. Results show that the standardized, indirect effect of badges on the student level is $\beta = 0.32$. The confidence interval for the estimate of the indirect effect [.165, .501] does not include zero; therefore, the null hypothesis is rejected. 52.26% of the total effect is mediated. Therefore, hypothesis **H3** is confirmed.

The results of H1, H2, and H3 indicate that though there were no significant differences between the experimental and the control groups on time-on-task, problems solved and attempted, and so on. Gamification has an impact on learners' time-on-task which positively affects their learning. In the control group no such relationship can be established, and their time-on-task can be the result of their intrinsic motivation. Therefore, I can conclude that badges positively affect learners' behaviour when provided in the SQL-Tutor but a similar effect can be achieved when learners are intrinsically motivated to use the system.

4.4.2 RQ 4: Do Students with Different Levels of Prior Knowledge React Differently to Gamification?

I also investigated the relationships between students' prior knowledge (using the pre-test score), the number of badges achieved as the moderating variable, time-on-task as the mediating variable, and student level as the outcome variable (Figure 4.6). The direct effect of pre-test score on time on task was not significant ($p = .48$). However, the interaction variable (pre-test x badges) significantly ($p < .01$) and positively affects ($\beta = 46.4$) time-on-task, which shows that badges moderate the effects of pre-test over time-on-task. R^2 change due to moderation effect is .098, indicating the interaction effect accounted for 9.8% added variation in time-on-task. Moreover, time-on-task significantly affects the student level ($\beta = .0044, p < .001$) confirming it as a mediator in the relation. Therefore, Hypothesis **H4** is confirmed. The total effect in the model again shows no direct relation between predictor (pre-test) and outcome (slevel) variables; however, the index of moderated mediation tested against bootstrap sample and 95% confidence interval confirmed the moderated mediation effect [.0032, 0.3282] of badges in the indirect relation between pretest and student level (zero does not fall between the upper and lower interval) mediated by time-on-task. This indicates that the mediation effect of time-on-task between pre-test and student level is conditional on the levels of badges. The more badges students are presented with the more time they will spend on the task, regardless of their pre-test score.

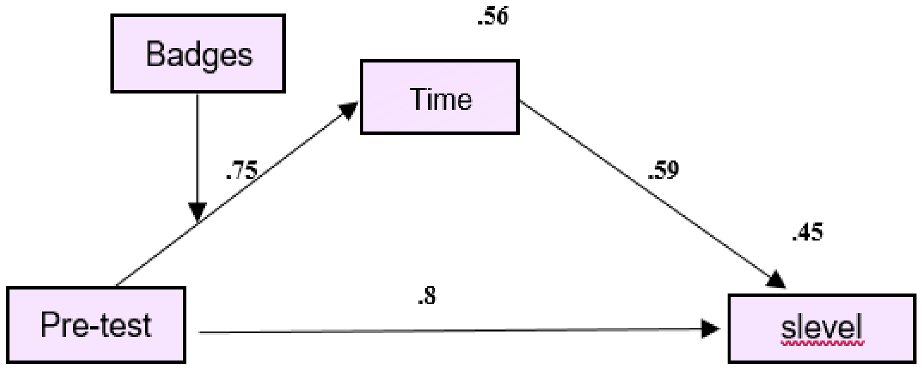


Figure 4.6. The moderated-mediation model, with badges as moderator

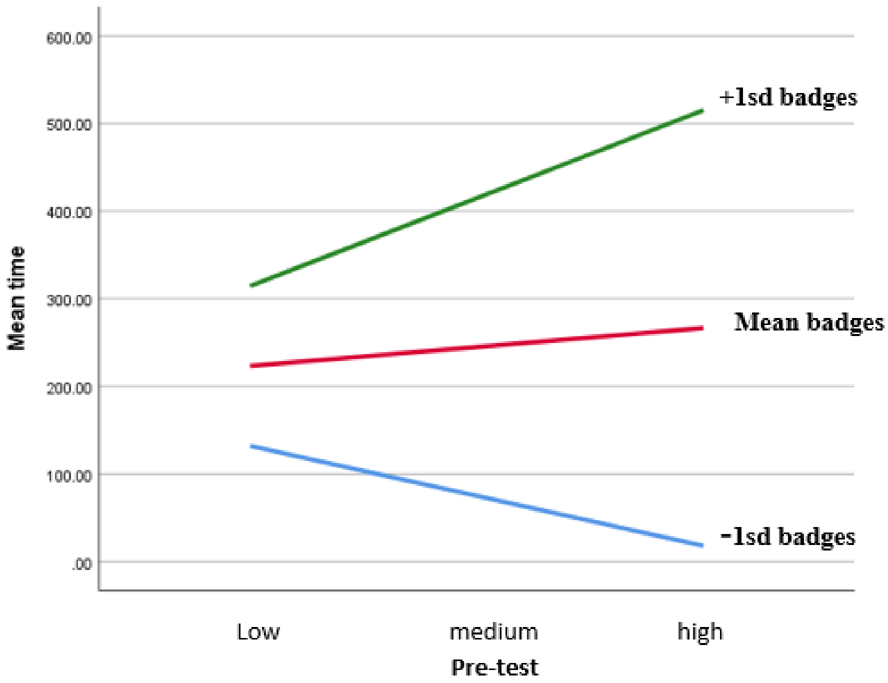


Figure 4.7. The conditional effects of pre-test score over time-on-task, moderated by badges

As the moderation effect of badges was significant, it is important to investigate different levels of the conditional effects. Figure 4.7 shows the moderation effects at +1SD (.86), mean (0), and -1SD (-.86) of badges. The significant moderation effects ($\beta = 46.4, p < .005$) were found only on higher badges (+1SD). It means that students who achieved more badges invested significantly higher time-on-task, particularly those who had higher levels of prior knowledge. However, those students who achieved fewer badges (mean or -1SD) invested less time-on-task (mean time in Figure 4.7) regardless of their prior knowledge scores. In fact, higher prior knowledge students experienced the worst case as their mean time approached zero.

4.4.3 RQ 5: What Would be the Effect of Gamification on Student Motivation?

The effect of badges on student motivation was measured by the motivational questionnaire in Surveys 1 and 2. I analysed the responses of 34 students who completed both surveys. This data set comprised 16 (46%) responses from the control group and 18 (43%) from the experimental group. I analysed the scores for each group separately to comprehend the independent effects on student motivation.

Self-efficacy increased after using SQL-Tutor as shown in Table 4.5, although the differences are not statistically significant for either group. Perceived competence results revealed that students from both groups were confident in their learning and performance skills at the time of the pre-test. However, at the end of the study, this confidence remains intact in the control group only and slightly decreased in the experimental group. The students' responses on the topic-interest items show the same pattern, with no differences between Survey 1 and Survey 2. As none of the differences were significant, Hypothesis **H5** is not supported.

Table 4.5. *Self-efficacy, Perceived Competence, and Topic-Interest Statistics: mean (SD)*

	Self-efficacy		**Perceived Competence**		**Topic-Interest**	
	Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test
Control (16)	3.56 (0.99)	4.37 (1.86)	5.33 (0.20)	5.41 (0.11)	4.56 (0.48)	4.62 (0.39)
Experimental (18)	3.55 (1.57)	3.88 (1.91)	4.96 (0.24)	4.5 (0.21)	5.04 (0.47)	4.47 (0.47)

To evaluate Hypothesis H6, I took topic-interest scores of each student in the experimental group from Survey 1 and tested its effects on my mediation model. The model is shown in Figure 4.8, where path A tests the effects of topic-interest as a moderator in the relationship between badges and time-on-task. Path B tests the effects of topic-interest as a moderator between time-on-task and student level. I selected model 58 in the Process macro to evaluate the two paths (Hayes, 2017).

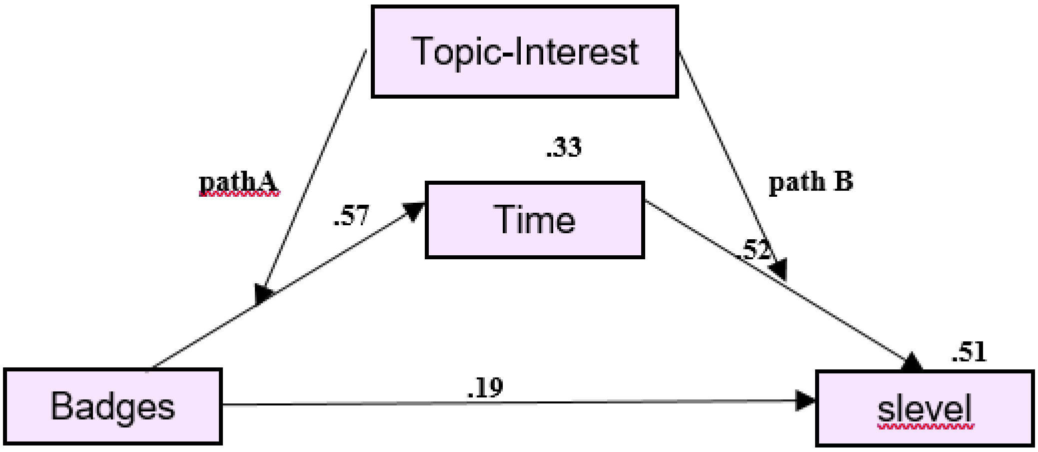


Figure 4.8. The moderation-mediation model, with topic-interest as moderator

The results of path A revealed a significant positive relation between badges and time-on-task ($\beta = 122, p < .0005$) and no significant relation between topic-interest and time-on-task ($p = .2$). However, the interaction variable (badges x topic-interest) has a significant positive effect ($\beta = 60.86, p = .05$) over time-on-task. This infers that topic-interest moderates the effects of badges on time-on-task. R2 change due to moderation effect is .0674, indicating the interaction effect accounted for 6.74% added variation in time-on-task. Since the moderation is symmetric, I can also interpret my results as the badges moderate students' interest in the topic and the time they spent on SQL-tutor.

As the interaction term in path A was found significant, I want to probe the interaction to better comprehend the moderated relationship between badges and time-on-task, as shown in Figure 4.9. At +1 SD on the topic-interest, which indicates the higher topic-interest, the relationship was positive and significant ($\beta = 175, p < .0005$). Similarly, at the mean (0), which represents the medium topic-interest, the relationship was positive and significant ($\beta = 122, p < .0005$). Finally, at -1SD of topic-interest, representing the low topic-interest, the relationship was negative and insignificant ($\beta = 69.8, p > .05$). Figure 4.9 shows that students with greater topic interest earned more badges that motivated them to spend more time-on-task.

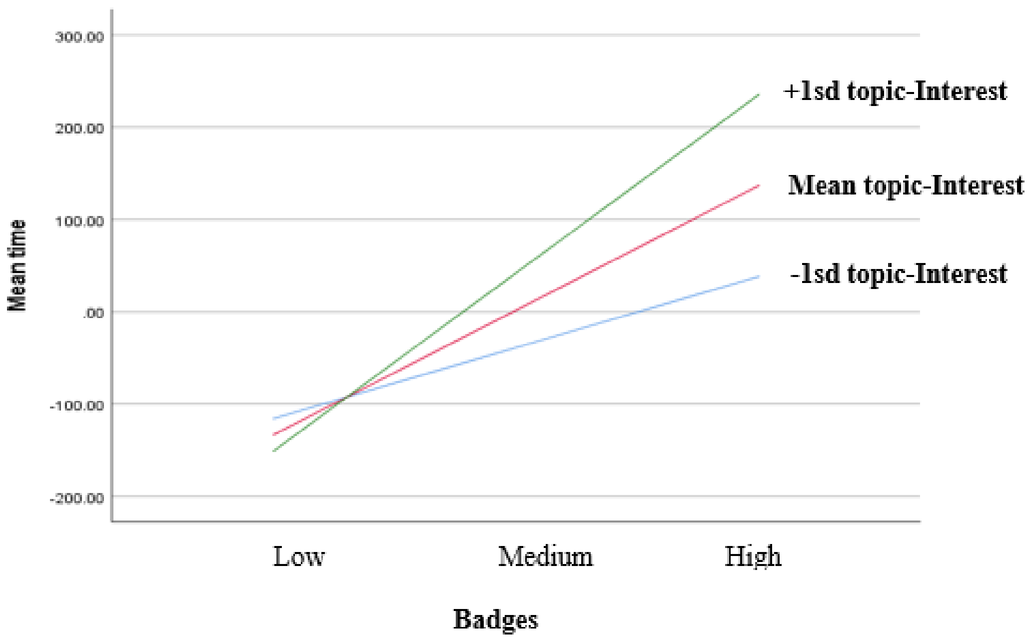


Figure 4.9. Relationship between badges and time-on-task moderated by topic-interest

The analysis of path B (Figure 4.8) revealed no significant relationship between badges and student level ($p > .1$) but a significant positive relation between time-on-task and student level ($\beta = .005, p < .0001$). However, the interaction effect between time-on-task and student's topic-interest is negative ($\beta = -.001$) at $p = .09$. R^2 change was .0398 indicating the interaction effect accounted for 3.98% added variation in student level. Therefore, the student's topic-interest marginally moderates the relationship between time-on-task and learning outcomes.

As the moderation effect of topic-interest was found significant, it is important to investigate the conditional effects at different levels. Figure 4.10 shows the moderation effects at high topic-interest (+1SD = .87), medium topic-interest (mean = 0), and low topic-interest (-1SD = -.86). It is evident that students who have the lowest interest in the topic but spent more time-on-task, significantly (*$\beta = .004, p < .001$*) improved their student level. On the other hand, students with higher interest in the topic also achieved the highest student level by spending more time-on-task. These results partially supported **H6**.

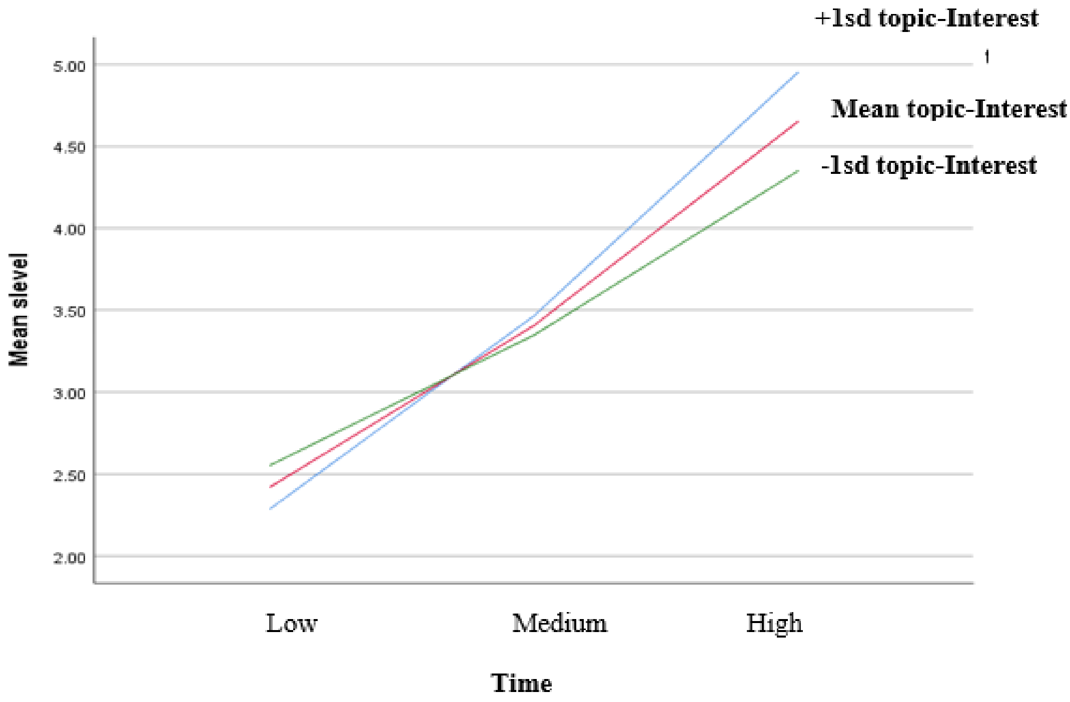


Figure 4.10. Relationship between time-on-task and slevel moderated by topic interest

From H5 and H6, I can state that the topic interest did not directly motivate students to spend more time on SQL-Tutor. However, badges as an external motivator indirectly motivated only those students with a higher interest in the topic. In order to motivate those students who are less motivated or have less interest in the topic, interventions that raise interest in the topic and increase their motivation are needed.

4.5 Further Investigation of the Experimental Group

Overall, the experimental group students achieved from 4 to 7 badges, with a mean of 5.43 (SD = .86). The percentage of students from the experimental group who earned various badges is shown in the last column of Table 4.1. On the very first day of interacting with SQL-Tutor, the students achieved an average of 4.60 badges (SD = .76). Only seven students achieved all primary badges; therefore, they were the only ones who were given daily challenges. For that reason, it is not possible to make any conclusions about the daily challenges.

The literature review shows that, in some cases, students are not interested in badges when they are not directly related to course credit. To investigate whether there is a difference in how much the experimental group students were interested in badges, I divided the experimental group students into two subgroups: those who visited the badge page at least once (23 students), and those who have never visited that page (19 students). Table 4.6 presents the differences found between the two subgroups.

Table 4.6 *Students Who Visited the Badge Page or not: mean (SD)*

	Seen badge page (23)	Not seen (19)	Significant
Pre-test %	54.59 (25.05)	65.49 (21.88)	p = .22
Time-on-task (min)	365.30 (272.27)	195.32 (316.96)	U = 348.5, p < .001
Sessions	9.48 (7.69)	4.63 (7.37)	U = 334.5, p < .005
Active Days	4.13 (3.22)	2.37 (2.71)	U = 312.5, p < .05
Attempted problems	51.91 (39.51)	30.58 (44.62)	U = 332, p < .005
Solved Problems	47.48 (36.86)	29.47 (44.49)	U = 326.5, p < .01
Constraints	287.74 (60.98)	247.84 (75.82)	U = 299.5, p < .05
Badges	5.74 (.81)	5.05 (.78)	U = 317, p < .01
Student level	3.70 (1.72)	2.84 (1.39)	p = .07
Post-test %	n = 13; 4.38 (2.93)	n = 8; 5.88 (3.72)	p = .34
Lab test %	59.74 (13.90)	61.26 (19.55)	p = .81

There was no significant difference between the two subgroups on the pre-test scores. The students who visited the badge page have interacted with SQL-Tutor significantly more, measured either as the total time ($p < .001$), the number of sessions ($p < .005$), or the number of active days ($p < .05$). Those students attempted/solved more problems (*p < .005 and p < .01* respectively) than their peers and achieved significantly more badges ($p < .01$). The students who have seen more badges have used significantly more constraints than their peers. In SQL-Tutor, domain knowledge is represented in terms of more than 700 constraints. Therefore, the students who visited the badge page covered a higher proportion of the domain in comparison to their peers. Therefore, there is evidence that visiting the badge page is correlated with more time-on-task and engagement. However, there was no significant difference between the two subgroups in terms of learning, measured either by the student level achieved ($p = .07$), post-test scores ($p = .34$) or the lab test score ($p = .81$).

4.6 Self-testing Behaviour

As mentioned in Section 4.3, the quiz was completely optional and provided to both experimental and control groups. To analyse students' self-testing behaviour, I investigated whether there is a difference in the student level achieved based on whether the students took the quiz and the group they were in (Table 4.7). I introduced a dummy QuizTaken variable, with values of 0 (quiz not taken) or 1 (quiz taken). In the control group, 12 students attempted the quiz while 23 did not. For the experimental group, 14 out of 42 students attempted the quiz. A two-way ANOVA (*F = 3.07, p < .05*, partial η^2 = .11) revealed neither a significant interaction between group and QuizTaken, nor the main effect of group, but there was a significant effect of the self-testing behaviour (*p = .01*, partial η^2 = .09) Students who attempted the quiz achieved a significantly higher student level.

Table 4.7. *Student Level*

Group	QuizTaken	Students	Student Level
Control	0	23	3.48 (1.38)
	1	12	4.58 (2.02)
Experiment	0	28	3.00 (1.47)
	1	14	3.93 (1.77)

Table 4.8 presents the statistics for students who attempted or did not attempt the quiz. There was no significant difference on the pre-/post-test scores and the lab test scores. The students who attempted the quiz interacted with SQL-Tutor significantly more, measured in terms of time, sessions, active days and attempted/solved problems. They used more constraints and solved more complex problems, thus achieving higher student levels.

Table 4.8. *Comparing Students Who Attempted/ Not Attempted Quiz: mean (SD)*

	Not attempted (51)	Attempted (26)	Significant
Pre-test %	56.65 (25.75)	62.82 (26.66)	p = .33
Time-on-task (min)	189.73 (153.89)	397.88 (321.47)	t = 3.85, p < .001
Sessions	5.20 (5.43)	9.81 (7.46)	t = 3.09, p < .005
Active Days	2.39 (1.86)	5.35 (3.01)	t = 5.32, p < .001
Attempted problems	28.27 (21.37)	63.08 (47.47)	t = 4.44, p < .001
Solved Problems	25.98 (19.09)	60.00 (46.28)	t = 4.56, p < .001
Max Problem Complexity	6.37 (1.93)	7.77 (1.42)	t = 3.26, p < .005
Constraints	244.24 (62.44)	317.23 (63.09)	t = 4.83, p < .001
Student level	3.22 (1.43)	4.23 (1.88)	t = 2.64, p < .05
Post-test %	n = 13; 4.38 (2.93)	n = 8; 5.88 (3.72)	p = .08
Lab test %	59.74 (13.90)	61.26 (19.55)	p = .10

4.7 Previous Gamification Experience

To investigate the relationship of students' previous experience in gamification (GE) with the badges and its subsequent effects on time-on-task and slevel, I developed the model shown in Figure 4.11. I took badges as an independent variable, time-on-task as a mediator, and slevel as the dependent variable in the model. I added GE (yes=1, no=0) as a dichotomous moderating variable.

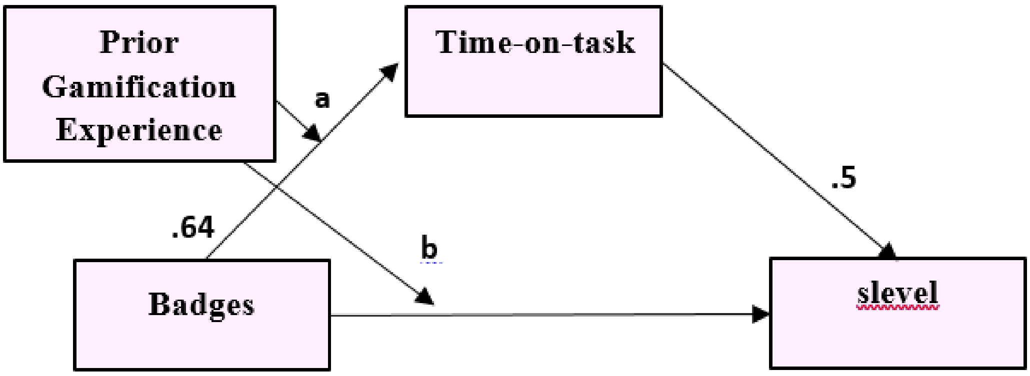


Figure 4.11. The moderated mediation model with gamification experience as a moderator

To evaluate the model, first, I regressed slevel on both time-on-task and badges. Results showed that time-on-task is a significant predictor of slevel (*β = .50, p = .01*), but the number of badges was not a significant predictor (*p* = *.06*). However, the number of badges was a significant predictor of time-on-task (*β = .64, p < .001*). Therefore, I can infer those badges indirectly and significantly predict slevel by mediating the time on task.

To examine the effects of GE in the established mediating model, I investigated the effects on two paths: a) GE moderates the effects of badges on time-on-task, and b) GE moderates the relationship between badges and slevel. I re-evaluated this model by using the Process macro for SPPS. The result of the path a shows the interaction term between badges and GE significantly and positively influences the time on task (*t = 2.33, p = .02*). Since the interaction term is found significant, it is important to probe the conditional relationship. That relationship was significant only when GE = 1 (*t = 5.59*, $p < .001$), but not when GE = 0. This indicates that badges helped only those students who had prior gamification experience. As evident from Figure 4.12, those who had no prior gamification experience spent maximum 200 minutes (mean time-on-task) with the system. On the other hand, those with previous gamification experience spent on average 370 minutes with the system.

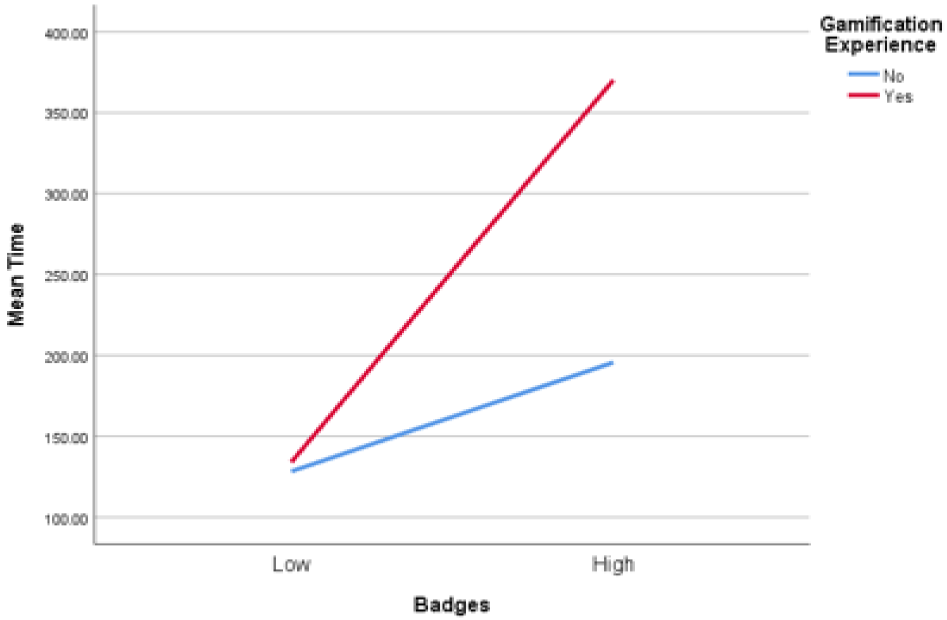


Figure 4.12. Relationship between badges and time-on-task when student had GE or no GE

The result of path b shows the interaction term did not influence slevel significantly. Therefore, GE did not moderate the relationship.

The total effect model shows both moderation and mediation effects in the model. The indirect effect shows that time-on-task is a significant mediator between badges and slevel as zero does not lie between upper and lower confidence interval limits [.0272, .4108]. This means that although GE increases time-on-task when combined with badges, in the absence of GE badges still influences student time-on-task, which influences their slevel. The index of moderated mediation, tested against the bootstrap sample and 95% confidence interval, confirmed the moderated-mediation effect [.0864, .7859] of GE in the model.

4.8 Survey 3 Analysis

I received 21 responses from the experimental group and 22 responses from the control group students. Table 4.9 summarizes the responses to the four questions on badges from the experimental group students. The Cronbach alpha for those questions is 0.88.

Table 4.9. *Responses from the Experimental Group (1 - Strongly Disagree to 5 - Strongly Agree)*

Question	**1**	**2**	**3**	**4**	**5**
Badges motivated me to participate more than I would have otherwise.	22%	26%	39%	4%	9%
I found being able to earn badges increased my enjoyment of using SQL-Tutor	9%	35%	26%	26%	4%
I would prefer not to see badges in SQL-Tutor.	0%	39%	35%	17%	9%
The badges awarded for solving problems motivated me to solve more problems than I would have otherwise.	17%	31%	39%	13%	0%

The responses of the experimental group indicate that students did not find badges very motivating. However, their trace data revealed that badges did motivate students to spend more time with the system. Therefore, the responses of students were regarded as indifferent. Students were indifferent in their responses about the enjoyment when they received badges. However, 39% of students stated they wanted to see the badges. I do not discuss the questions on daily challenges, as only seven students received them during the study. Almost 62% of students wanted to see the daily challenges in SQL-Tutor; this Figure reveals that students were interested in daily challenges in principle. The students from both groups enjoyed attempting quiz (control = 68%, experimental = 62%) and prefer to see them in SQL-Tutor (control= 86%, experimental = 62%).

4.9 Discussion

This paper presents a classroom study in which I analysed the effect of gamification in the context of SQL-Tutor. These findings highlight the effects of gamification in the context of an ITS, under realistic conditions, in a study that lasted four weeks.

Starting from Landers' theory of gamified learning (2014), I designed badges that supported goal setting, assessment, and challenges—three common categories of game elements. I hypothesized that the badges would motivate students to spend more time on tasks (solving problems in SQL-Tutor). The goal-setting behaviour is supported by setting SMART goals/criteria for achieving each badge. Challenges motivate students to perform more complex tasks, and the quiz allowed students to test their knowledge.

This study provides initial evidence that badges can increase student learning in ITSs (measured as the student level in SQL-Tutor). This relation can be mediated by the time participants spend on the task. The results show the impact of gamification on learning through behavioural change, supporting the theory of gamified learning with the time-on-task as a valid behaviour target for gamification. I determined that time-on-task correlated and predicted learning outcomes. I did not find a difference between gamified and non-gamified groups regarding time spent in SQL-Tutor, problems completed, and learning outcomes. A possible explanation for this finding is that the students were already highly motivated and used SQL-Tutor to prepare for the lab test. However, I found evidence that goal setting, challenges, and self-testing behaviours implemented as badges indirectly and significantly affected learning outcomes through the time-on-task as the mediator.

The second finding of the study is that prior knowledge did not directly affect time on task; however, it yielded significant effects when combined with badges. The detailed investigation of this moderation effect revealed that students who achieved more badges spent more time on SQL-Tutor, particularly those who had higher prior knowledge. However, students who achieved an average number of badges spent their mean time regardless of their prior knowledge. Those students who received fewer badges spent little time, especially the higher prior knowledge group who spent the least time. These findings further elaborate on the dynamics of badges in my study.

As mentioned in the literature review, badges do not only engage students but also affect their motivation. In this study, I evaluated student motivation by measuring their self-efficacy, perceived competence, and topic-interest. I found no differences in these three motivational constructs between the two groups. I found that those badges enticed students to spend more time-on-task; for that reason, I further investigated the indirect effects of these motivational constructs in the study. The scores on topic interest from Survey 1 provide an insight into how much students valued this part of the course. The statistical analysis revealed that topic interest moderated the effect of badges on time on task but marginally moderated the effect of time on task on the student level. As the moderation relationship is symmetric, it can be stated that badges moderated the relationship between topic interest and time on task. The detailed investigation on the moderation relationship indicated that higher interest in SQL strengthened the relationship between badges and time on task by influencing students to achieve more badges. When combined with achieved badges, lower interest in SQL motivated students to spend more time but not as much as higher interest did. Similarly, the student's interest in SQL slightly influences their time on task and learning outcome (student level) relationship.

In the literature review, I pointed out a few methodological gaps in the educational gamification research. In this study, I tried to fill those gaps by following the gamified theory of learning, analyzing the effects of a particular game mechanic (badges) on specific student behaviour (time on task), and most importantly, conducting a controlled experiment by following most of the design guidelines. Another contribution of this research is to provide separate and combined effects of different motivational constructs through the gamified system.

From the discussion above, I can conclude that gamification influences students' learning behavior, affecting their learning outcomes. It affects both higher prior knowledge and low prior knowledge students; the more badges they achieve, the more time they spend interacting with SQL-Tutor. Finally, the student's interest in SQL influenced the time on task when combined with badges. This provides evidence of both engagement and motivation dynamics of gamification in the context of ITSs even though the surveys did not show that badges were directly motivating.

There are two major limitations of my study, the first being the small sample size. The second limitation was the design of the badges, which could be designed in a more visually attractive manner; also having other extrinsic rewards associated with badges would be useful, for example high scores. As discussed, almost 46% of students in the experimental group did not access the badge page despite receiving badge notifications. This shows that the design of badges was not attractive enough to entice some learners and motivate them to achieve.

Chapter 5 Self-regulated Learning Support

5.1 Introduction

Self-regulated learning (SRL) is a combination of processes and motivational beliefs which empower learners to set their goals, select strategies, and monitor performance while learning (Zimmerman, 2000). SRL accounts for considerable improvement in learning, particularly in higher education settings (Sitzmann & Ely, 2011). A few learning environments support SRL processes to help students adopt self-regulated learning (Duffy & Azevedo, 2015; Winne & Hadwin, 2013). The researchers advocate that for SRL to yield utmost benefits, the learning environments should (a) follow a model or framework for implementing SRL-based activities and processes and (b) accurately measure these processes. As these processes have cyclical effects (Winne & Hadwin, 2008; Zimmerman, 1990), implementation under a specific framework is critical for the success of SRL. MetaTutor (Duffy & Azevedo, 2015) is one such example, which provides support for many of the SRL processes based on the COPES model (Winne & Hadwin, 2008).

In Chapter 4, I reported that only 54% of students were interested in gamification, primarily those who had a higher interest in the topic. One possible reason for these results is the nature of gamification as an external motivator that might not excite university students. SRL is a strategy that invokes the internal motivation of learners to reach their goals by applying various processes and strategies and continuously monitoring their progress. These processes and strategies are more helpful in higher education settings.

The study presented in this Chapter evaluates the effects of SRL support on learners' motivation and learning. To provide this support, I introduced three interventions: goal-setting support, dashboard, and self-reflection prompts in SQL-Tutor. These interventions support specific processes in the three phases of Zimmerman's (1990) self-regulated framework. The first intervention, *goal-setting support*, targets the goal-setting process of the forethought phase. The dashboard provides support to the self-monitoring process of the performance phase, and self-reflection prompts facilitate self-evaluation processes of the self-reflection phase. This study aims to determine *the effects of SRL support provided in interventions on students' learning, motivation, and SRL skills.* In the light of the aims of this study, I propose three research questions and six hypotheses. The first research question addresses learning. The second research question focuses on the different effects of each intervention. The last research question evaluates the students' SRL skills and motivation part of the study. Following are the research questions and related hypotheses:

Research Question 6: What are the effects of SRL support on learning? In the research literature, little evidence was found about the influence of SRL support on students' learning outcomes (Broadbent et al., 2020). In this context, I expect that the *experimental group which received the SRL support would achieve higher learning outcomes than the control group who did not receive SRL support* **(H1)**.

Research Question 7: What are the (separate) effects of the three interventions on students' learning behaviors? The experimental group students were encouraged to select more challenging goals in goal-setting support intervention but could decide to go with less challenging goals. This strategy was based on the goal-setting theory (Locke, 2019), which claimed that difficult goals lead to higher performance. The effectiveness of the goal-setting theory has been shown in more than 1,000 studies. The theory focused on goal difficulty and discussed the greater effects of task-specific over non-task related goals (Chen et al., 2014; Latham et al., 2012) and subsequent effects of selecting challenging goals (Latham et al., 2017). As mentioned in the meta-review of achievement (Collins, 2004), meeting a standard or goal is not enough; one should struggle for excellence. The following hypothesis addresses such a situation to determine what happens when students do not follow the system's suggestions. *Setting challenging goals affects students' motivation, engagement, and* *performance positively* **(H2)**. The research found that students selected problems effectively in the presence of OLM (Long et al., 2015). SQL-Tutor presented OLM to students when they required it or when they selected a new problem and found that OLM positively affects the learning of less-able students and improves problem-selection skills (Mitrovic & Martin, 2007). However, the OLM provides limited support for behavioural metrics (Bodily et al., 2018), which are vital for self-regulated learning (Verbert et al., 2014). In this study, I combined OLM with the learning analytics dashboard to provide a more complete picture of learners' problem-solving progress. Furthermore, the dashboard provided in this study presents two problem selection options to learners. Research shows that problem selection from the dashboard improves learners' metacognitive abilities, leading to better problem selection decisions (Chen et al., 2019). Therefore, the following hypothesis would address such a situation: *Selecting problems after viewing the dashboard improves learners' problem selection behaviour, leading to higher engagement and motivation* **(H3)**. Carpenter et al. (2020) report that the learners' reflection affect their problem-solving progress and learning outcomes; particularly, time spent reflecting is associated with higher learning gains (Dever et al., 2021). In this study, self-reflection prompts support self-reflection in the form of questions. Therefore, *I would expect more time spent on self-reflection prompts is associated with increased learning outcomes* **(H4).** *I also expect that learners' responses to these prompts are predicted by their problem selection and problem-solving in SQL-Tutor* **(H5)**.

Research Question 8: Do SRL interventions affect learners' SRL skills and motivation? Based on the previous research which advocated that technological scaffolds should increase SRL skills of learners (Broadbent et al., 2020) even outside the learning environments, I hypothesize that the *experimental group which received support for goal setting, support for monitoring by the dashboard, and facility for reflecting their thoughts and emotion on self-reflection prompts will improve their SRL skills and motivation* ***(H6)***.

To address these research questions, first, I have modified SQL-Tutor with SRL interventions. The details of intervention implementation are discussed in Section 5.2. The standard and enhanced versions of SQL-Tutor were allocated to the control and experimental groups, respectively. Section 5.3 provides a complete account of how the study was conducted. The results obtained from the study are elaborated in Section 5.4. Section 5.5 presents a small lab experiment for participants' eye gaze and affective state determination, followed by discussion and conclusions in Section 5.6.

5.2 Experiment Design

SQL-Tutor was enhanced with three interventions, each depicting one phase of Zimmerman's SRL framework. These interventions are goal-setting support for the forethought phase, the dashboard for the performance phase, and self-reflection prompts for self-reflection. The following are the implementation details of each intervention.

5.2.1 Goal-Setting Support

The goal-setting process allows students to set their own goals and plan actions to attain those goals. SQL-Tutor provides eight task-specific goals to students, where each goal represents a problem template in SQL. A problem template covers a set of problems, which require the same problem-solving strategy. SQL-Tutor contains a set of 300 problems for practicing multiple types of queries. These problems and their ideal solution are classified using 38 different problem templates (Matthews, 2006). The classification has been done based on a range of criteria, such as the number of tables used in the query, nested statements, number, and type of attributes and relations involved, aggregate functions, joins, and other SQL keywords. The 38 problem templates are then grouped into eight categories to provide a higher level of abstraction; I refer to those categories as task-based goals in this study.

The goals are arranged by increasing complexity (Figure 5.2). The student needs to select a goal at the start of each session. The system provides support for selecting challenging goals. In the situation illustrated in Figure 5.1, the student selected the first goal (Goal 1), and SQL-Tutor provided a message, encouraging the student to select a more challenging goal (goal 4 or 5). The student is free to select one of the suggested goals or any other goal. At the start of interaction with SQL-Tutor, the system uses the student's score on the pre-test to suggest a challenging goal. The pre-test score ranges from 0 to 9. However, in the subsequent sessions, it considers the student's current level (*slevel*) only. The student level ranges from 1 to 9, and it is determined dynamically based on the student's success during problem-solving (Mitrovic, 2003). The proposed goal is determined heuristically, as presented in Table 5.1. For example, if the student scored six (i.e., the median pre-test score in the gamification study) or more on the pre-test, the challenging goal should be 8. The goal-setting page shows the total number of problems for each goal and the number of problems the student has solved. Previously achieved goals are highlighted on this page (please note that in the situation shown in Figure 5.1, the student has not yet achieved any goals). If the student with a low pre-test score selects a very challenging goal, the system would suggest a less challenging goal. To achieve a goal, the student needs to complete at least half of the problems or solve the five most complex problems for *the current goal.*

Table 5.1. *Rules for Recommending Goals*

Rules	Pre-Test	S-level	Recommended Goal
1	0, 1	1	4
2	2, 3	1	6
3	4, 5	1	7
4	>5	1	8
5	na	2, 3	4, 5
6	na	4, 5	6, 7
7	na	>5	8

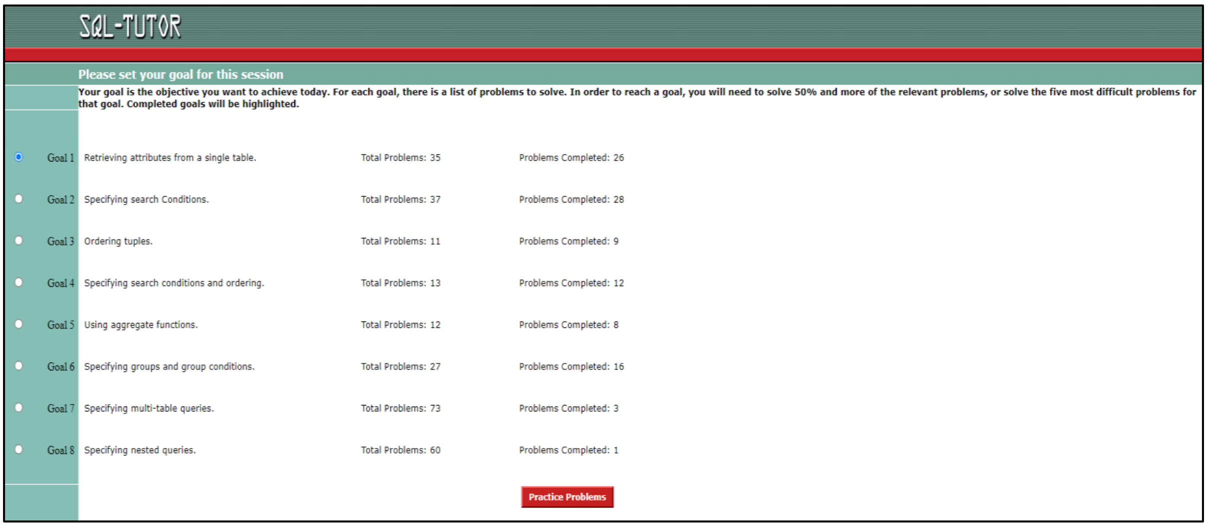


Figure 5.1. Goal-setting page for control group

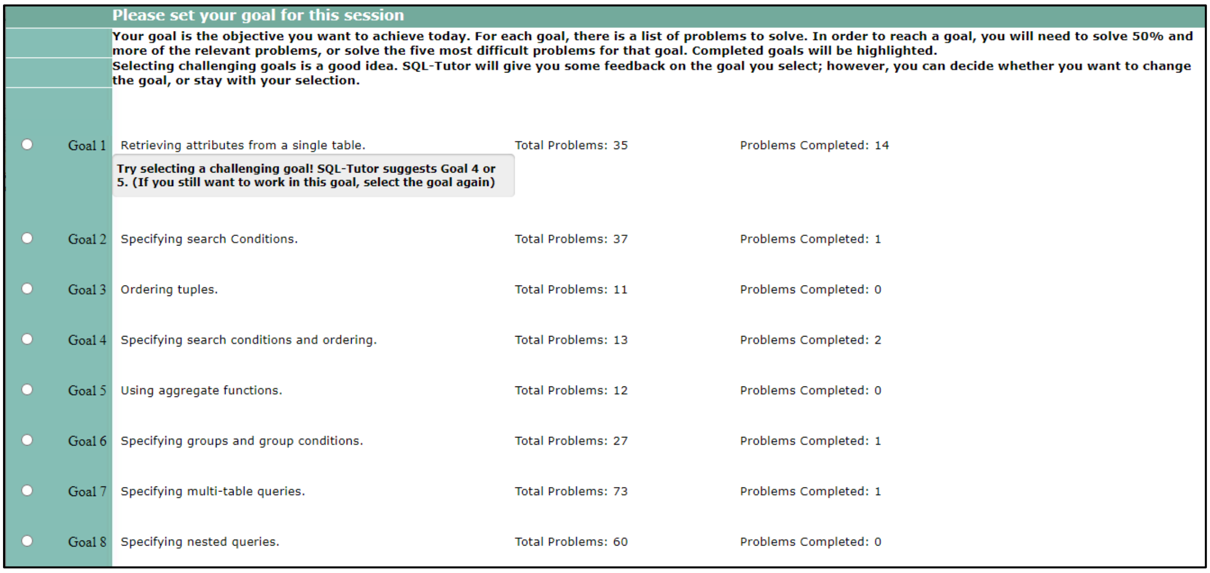


Figure 5.2. Goal-setting page with the message set challenging goals for experimental group

5.2.2 Progress Bar

A student progress on the selected goal was presented on the problem-solving interface by the progress bar. The progress bar and the goal description showed how many problems the learner has completed in that goal, as shown in Figure 5.3. The progress button on the top of the problem-solving interface has a similar purpose as the progress bar. When the student clicked on the button, a second window emerged, as shown in Figure 5.4, which consisted of goal description, goal progress, and the number of solved/unsolved problems in the goal.

The experimental group received the progress bar on the interface, whereas the control group received the progress button. The reason to provide the progress bar is to facilitate the monitoring skills of the experimental group. However, for the control group, their need to monitor their progress was recorded explicitly when they clicked on the progress button.

5.2.3 Self-reflection Prompts

Self-reflection prompts presented four questions to the learners, as shown in Figure 5.5. The first two questions were related to the student's self-evaluation on the given task (Q.1 *I think I have understood the principle in the problem*, Q.2 *While working with SQL-Tutor, I intensely reflected on the subject matter*). The third question accounted for student's self-efficacy (Q.3 *I am satisfied with my performance*), and the last question records their current feelings (Q.4 *I am feeling*). The scale used for the first three questions ranges from Not very True (1) to Very True (7). Five emojis were shown (smiling, happy, neutral, sad, and angry faces) as a scale to depict their current emotion in the last question. These questions were presented every time the student completed a problem. Responding to all four questions was mandatory to proceed to the next step.

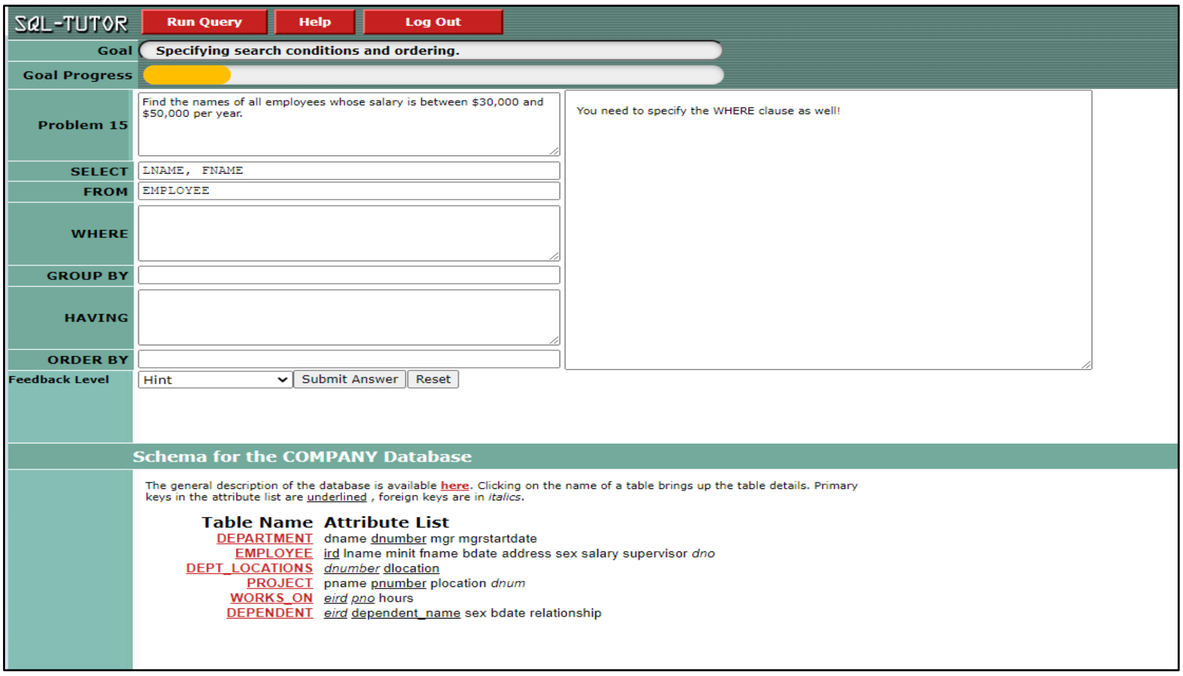


Figure 5.3. Progress bar on the problem-solving interface

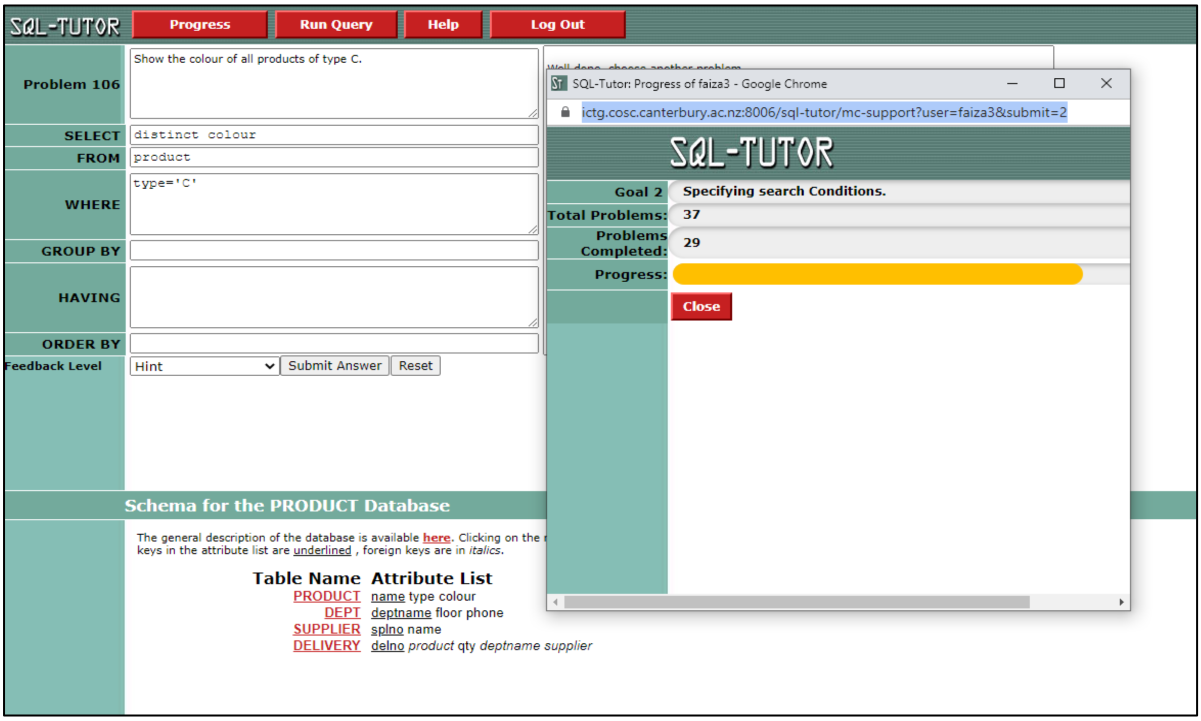


Figure 5.4. Progress button and page showing explicit progress of a student on the selected goal

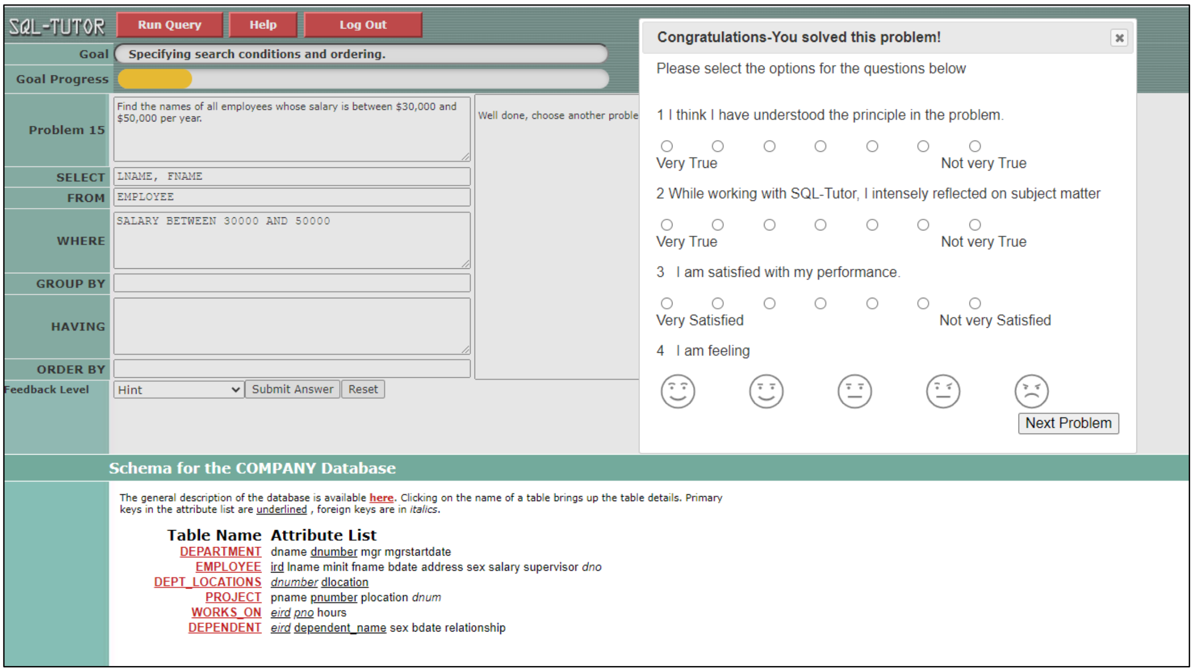


Figure 5.5. An example of the self-reflection prompts for the experimental group

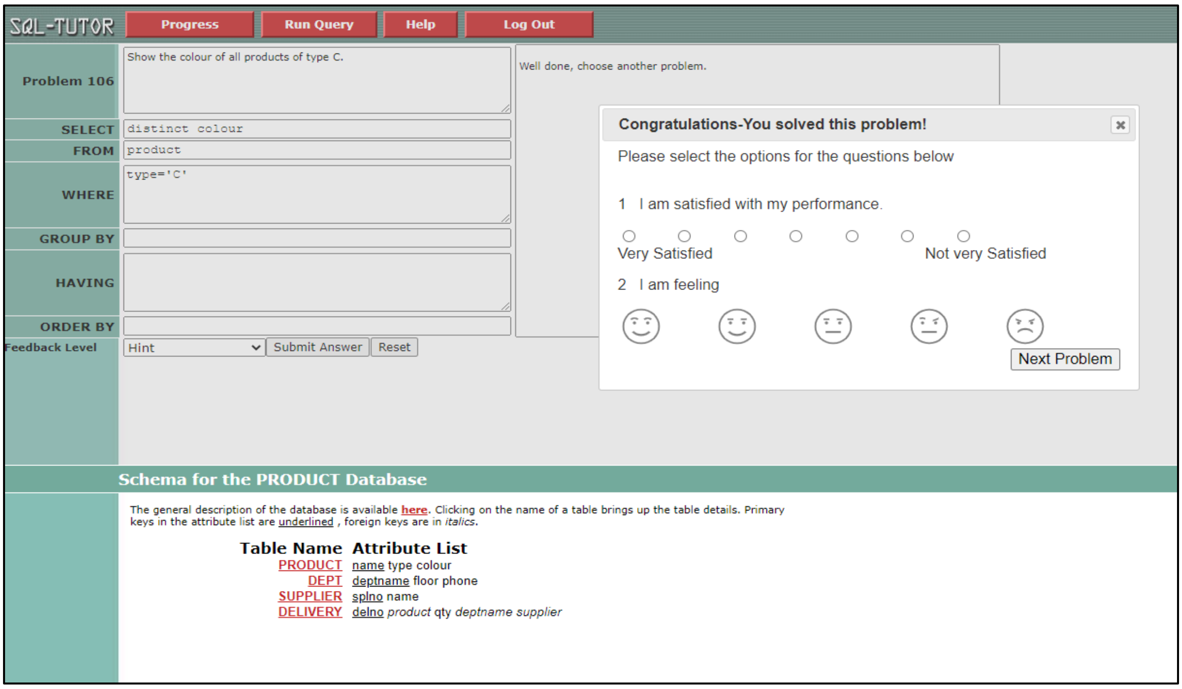


Figure 5.6. An example of the self-reflection prompts for the control group

5.2.4 Dashboard

The research literature shows that only one-third of the OLMs show behavioural metrics to students (Bodily et al., 2018). These behavioural metrics are vital to inform the learning analytics dashboard and to increase metacognitive abilities and self-regulated learning (Verbert et al., 2014). Therefore, I presented a comprehensive dashboard in this study, combining OLM and learning analytics dashboard (LAD) approaches.

The dashboard is presented to the experimental group students upon completing a problem and the self-reflection prompt, as shown in Figure 5.7. The top section of the dashboard provides the general information about the student's history, such as his/her pre-test score, current knowledge level, total time spent with SQL-Tutor, total problems solved with SQL-Tutor, the highest problem complexity, and the percentage of attempts on which the student required to see the complete solution. The second section of the dashboard visualizes the student's progress and the average class progress on each goal in skill meters. When a student achieves the goal, the dashboard shows an appreciation message and the next goal selection option; otherwise, it shows two options to select the following problem: practice problems and challenge me. The practice problems strategy leads the student to the problem selection page to select an appropriate problem for the selected goal. The "challenge me" option selects one of the most complex problems for the current goal to the student. The bottom section of the dashboard presents two graphs, which track the problems completed and time spent with SQL-Tutor per week. The third graph is the open learner model, i.e., the visualization of the student's knowledge in terms of six clauses of the SQL Select statement (select, from, where, order by, having, and group by).

Figure.5.7. Dashboard of SQL-Tutor

5.2.5 Surveys 1 and 2

The SRL instrument at the start (Survey 1) and end (Survey 2) of the study were adopted from (Kizilcec et al., 2017). The survey contained 24 questions divided into six subscales: goal setting (4), strategic planning (4), help-seeking (4), elaboration (3), self-evaluation (3), and task strategies (6). In addition to these questions, five questions from the self-efficacy dimension of the Motivated Strategies for Learning Questionnaire (Pintrich & De Groot, 1990) were also added to the survey. The self-efficacy dimension consists of nine questions that have two major types, (1) self-efficacy in comparison to other classmates (n=4) and (2) self-efficacy of oneself (n=5). I have selected the later ones in which an individual's self-efficacy was measured. A five-point Likert scale, ranging from "Not at all true for me" (1), "True for me" (2), "neither true nor untrue for me" (3), "True for me" (4), and "Very true for me" (5) was used for responses.

5.3 Procedure

The participants were volunteers from the second-year relational database systems course at the University of Canterbury in 2020. Students were introduced to the system in a scheduled lab, and the system's use was completely voluntary. In other words, students did not receive any credit for using the system. The students were divided into control and experimental groups. 57 students (female = 19%, male = 81%) in experimental group and 42 in control group (female = 29%, male = 71%) gave their consent to participate in the study.

Students received the pre-test (discussed in Section 4.3, Chapter 4) immediately after logging in, followed by the survey1. After that, students were required to set a goal for the session. In the experimental group, students received support during goal setting, as discussed earlier. On the other hand, the control group participants did not receive any suggestions upon selecting a goal.

After selecting a goal, the list of all problems for the selected goal was shown, arranged by the complexity. The student could choose any problem and work on it. The experimental group could view their progress on the selected goal in the progress bar presented at the top of the page. The progress bar was not available for the control group; instead, the participant clicked the "progress" button to see the current goal and the progress bar. After achieving a goal, the student needed to select the next goal.

The experimental group students were directed toward the dashboard to view the progress and select the following problem. Once the student completed a problem, self-reflection prompts were presented on the screen. The experimental group received all four questions, while the control group received only two questions, as discussed in Section 5.2.3. At the same time, the control group was diverted towards the problem selection page.

The study lasted for four weeks. At the end of the study, students completed the post-test of similar structure and complexity as the pre-test and survey 2 (identical to Survey 1).

5.4 Results

The average score on the pre-test was 63.98% (SD = 27.14), and post-test was 66.92% (SD = 28.26). However, only 46 students completed the post-test. I believe the low completion rate for the post-test was due to its voluntary nature; in addition, the post-test was administered only one day before a high-stake scheduled test in the course. On average, students completed 71.46 (SD = 53.36) problems, with the average number of attempted solution being 273.85 (SD = 234.84). The participants interacted with the system on four distinct days on average (SD= 2.7, max=12, min=1) and spent on average 323 minutes (SD=281, max=1526, min=16) with the system.

Table 5.2. *Pre-/Post-test Scores for the Two Groups*

	Experimental (28)	**Control (18)**
Pre-test (%)	55.56 (29.18)	59.88 (28.82)
Post-test (%)	71.04 (27.45)	60.51 (29.09)
Normalized learning gain	0.34 (.59)	0.12 (0.5)
Effect size (d)	0.47	0.03

Table 5.3. *Summary of Major Statistics: Mean (SD)*

	Control (42)	**Experimental (57)**	**Significance**
Attempted Problems	92.98 (61.86)	57.46 (41.33)	U = 783, p = .003
Completed Problems	91.86 (61.33)	56.44 (41.09)	U = 783, p = .003
Attempts	326.95 (303.73)	234.72 (159.60)	p = .23
Problem Complexity	2.92 (0.96)	3.32 (1.08)	U = 1465.5, p =.057
Time (min)	360.19 (335.33)	296.71 (233.22)	p = .58

5.4.1 RQ 6: What are the Effects of the SRL Support on Learning?

To evaluate H1, I compared the pre/post-test scores of those participants who completed both tests (Table 5.2). There is no significant difference between the two groups on the pre-test scores, showing that participants had the same level of background knowledge. The experimental group improved significantly from pre- to post-test ($W = 298, p = .03$), but the control group students did not ($p = .74$). No significant difference ($p = .2$) was found between the post-test scores. The normalized gain for the experimental group was 34% with a medium effect size ($d = 0.47$), whereas the control group had the normalized learning gain of 12% and a small effect size ($d = 0.03$). Comparing normalized gains revealed no significant difference. These results partially support hypothesis **H1**.

The Mann-Whitney-U test was used to compare the means of the two groups, as reported in Table 5.3. The reason to select this test is that the data were not normally distributed. The control group attempted/completed significantly more problems than the experimental group and made significantly more attempts on problems. There was no significant difference between the two groups on time spent in the system. The experimental group completed significantly more complex problems than the control group. These findings show that the experimental group achieved higher learning gains by completing fewer but more complex problems than the control group.

These findings indicate that completing more problems and spending more time with the system may not help in learning until learners make a conscious decision about their learning effort. Students who received suggestions on goals started thinking about appropriate goals, leading to specific problem selection.

5.4.2 RQ7: What are the (separate) Effects of the Three Interventions on Students' Learning Behaviours?

To evaluate hypothesis H2, I analyzed data from the experimental group only. I divided participants' post-hoc into three subgroups based on how they selected goals. 14 students always accepted the goals suggested by the system (referred to as SG). Eighteen students worked on the goals in the sequential order (SEQ), disregarding the system's suggestions. The remaining 25 students used a mixed strategy; they accepted the goal the system suggested in some cases, and in others ignored those suggestions (Mix). Table 5.4 presents the statistics for the three subgroups. Using the Kruskal-Wallis test, I found no significant differences between the subgroups on the pre-/post-test scores and the time spent in the system, but there were statistically significant differences on the number of attempted goals ($H = 8.12, p = .017$), achieved goals ($H = 10.13, p = .006$), the number of attempted/solved problems ($H = 13.88, p = .001$ and $H = 14.41, p = .001$ respectively), and problem complexity ($H = 12.20, p = .002$).

I then analyzed the differences between the subgroups using the Mann-Whitney U test with the Bonferroni correction for multiple comparisons, with statistical significance accepted at the $p < .017$ level. The post-hoc analyses revealed no significant differences between the SEQ and Mix groups. The SG subgroup attempted significantly fewer goals in comparison to the SEQ ($U = 55, p = .006$) and Mix groups ($U = 94, p = .016$), and achieved significantly fewer goals in comparison to the Mix group ($U = 77, p = .003$). The SG group also attempted and solved significantly fewer problems in comparison to the SEQ ($U = 44, p = .002$ in both cases) and Mix groups ($U = 74.5, p = .003$ and $U = 71, p = .002$ respectively). However, the average problem complexity of solved problems for the SG group was significantly higher in comparison to the SEQ ($U = 40.5, p = .001$) and Mix ($U = 77.5, p = .004$) groups. Therefore, those students who accepted the system's suggestions solved fewer but more complex problems.

I also analyzed the data using the structural equation model shown in Figure 5.8. A positive correlation between pre-existing knowledge and learning outcomes is commonly found in the literature, e.g. (Brusilovsky et al., 2018). Another predictor is the number of attempted problems since students get personalized feedback from the system on incorrect attempts, which helps them improve their knowledge. The variable labelled "Accepted goals" shows how many times students accepted the system's suggestion for the goal. Because not all students completed the post-test, I used a different measure of learning: the Highest Achieved Goal (HAG).

Table 5.4. *Summary Statistics for the Three Subgroups: Mean (SD)*

	SEQ (18)	**Mix (25)**	**SG (14)**
Pre-test %	62.97 (27.75)	64.46 (24.01)	61.11 (33.98)
Post-Test %	n=9, 64.21 (31.81)	n=12, 77.78 (28.55)	n=8, 69.45 (21.23)
Time (min)	346.17 (290.49)	283.36 (163.41)	257.0 (263.09)
Attempted goals	6.39 (2.62)	7.04 (1.14)	5.00 (2.18)
Achieved goals	4.72 (2.54)	3.60 (2.43)	1.64 (2.34)
Attempted Problems	78.28 (44.84)	58.96 (38.18)	28.0 (22.34)
Problem Solved	77.50 (44.23)	57.88 (38.02)	26.79 (21.92)
Problem Complexity	2.85 (.74)	3.11 (.86)	4.31 (1.20)

To evaluate the model against the data, I used the Hayes (2017) Process macro for SPSS, version 3.5. Figure 5.8 shows the standardized regression coefficients for the model. All the path coefficients are significant at $p < .05$ except the effect of Pre-Test on HAG ($p = .11$), and the covariance between Accepted Goals and Pre-Test. There is a significant negative effect of Accepted goals on Attempted problems. The effect of Pre-Test on Attempted problems is also negative. Accepted goals have a significant positive effect on HAG, and Attempted problems also have a significant positive effect on HAG ($p < .001$). The indirect and total effects in the model are tested using the bootstrap sample and 95% confidence intervals.

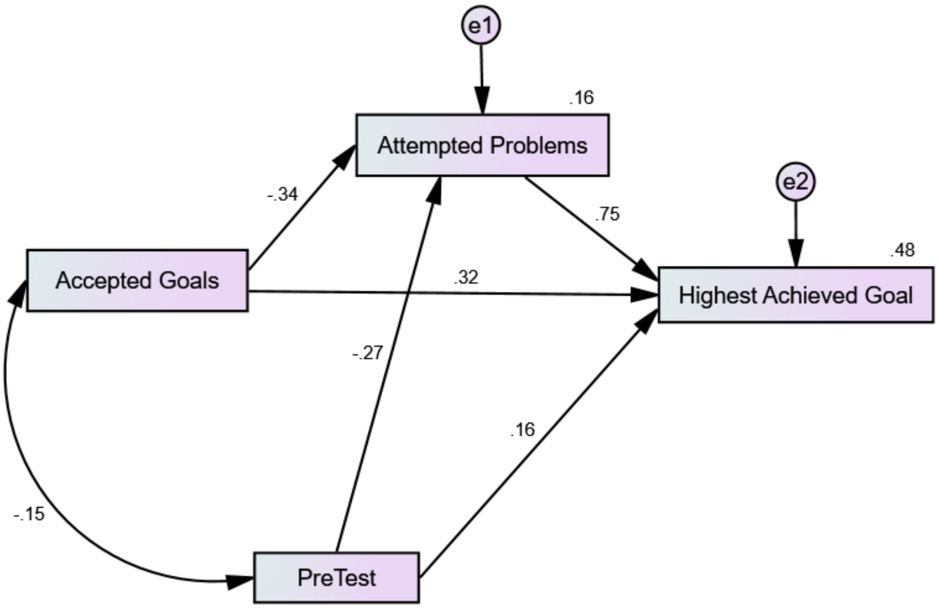


Figure. 5.8. Multiple mediation model with standardized coefficients

These findings suggest that (1) students who accepted system-suggested goals tended to achieve higher goals (the confidence interval [.1345, .7074] does not include zero), and (2) students who accepted suggested goals despite attempting fewer problems achieved higher goals (the confidence interval [-5903, -.1133] does not include zero). Students who scored less on the pre-test achieved higher goals when they accepted system suggestions. These findings support **H2**.

The experimental group solved significantly fewer problems than the control group (Table 5.3). Remember that experimental group students selected problems on the dashboard, and control group students selected from the practice problem's page. Table 5.5 shows the average percentages (number of completed problems in a level / total number of problems completed) of problems completed in each complexity level for the two groups. In both groups, students selected problems mostly from complexity levels 1, 2, and 3. The differences between mean values were tested using an independent sample t-test. The experimental group students selected and completed significantly fewer problems than the control group in complexity levels 1 and 2. For complexity levels 3, 5, 7, and 9, experimental group students completed a higher number of problems than the control group, and the differences are statistically significant.

Table 5.5. *Means of Percentages of Problems Completed in each Complexity*

Complexity Levels	Experimental (n=57)	Control (n=42)	Significance
1	21.76 (108.08)	32.48 (56.04)	t (73) = -2.29, p < .05
2	21.25 (71.77)	28.54 (144.13)	t (68) = -3.24, p < .001
3	23.49 (108.08)	19.21 (56.04)	t (91) = 2.31, p < .05
4	17.04 (56.6)	14.01 (57.02)	t (84) = 1.81, p = .07
5	12.46 (71.39)	8.06 (35.06)	t (78) = 2.74, p < .001
6	6.57 (11.75)	6.79 (129.72)	
7	11.22 (95.29)	4.82 (12.75)	t (56) = 3.75, p < .001
8	1.65 (2.76)	2.13 (1.55)	
9	14.59 (40.79)	0.73 (0.08)	t (12) = 2.45, p < .05

These results shed some light on how students selected SQL-Tutor problems with the dashboard and without the dashboard. Students in the experimental group selected fewer easier problems and more complex problems than the control group. It is evident that these students did not solve the problems in the presented order but thoughtfully selected problems. Another interesting finding from the literature is that students tend to select fewer complex problems (Long et al., 2015); however, the results reported that 13 students in the experimental group completed the problems in the highest complexity level (level 9) compared to only three students in the control group as shown is the last row of Table 5.5. It is also observed that students in the experimental group did not overstay on the low complexity problems, and the average between complexity 1 and 4 did not vary much. On the other hand, the average complexity levels of the control group dropped significantly after complexity level 2 and kept declining for the rest of the levels.

The post-hoc analysis of the experimental group revealed that students selected 30% of problems using the *challenge-me* option and 70% of problems from the practice problems option. These results show that students in the experimental group took better problem selection decisions than the control group. Therefore, it is essential to explore further this decision to figure out whether these decisions improve their engagement with the ITS.

The challenge-me option presented the most complex problem of the goal; therefore, if the learner completed five problems using the challenge-me option, they can achieve that goal (one of the conditions for achieving a goal). Figure 5.9 shows goals achieved by the number of students using the challenge-me option. Almost 50% of the students achieved goals 1, 4, 5, and 6 using the challenge-me option. Goal 8 (the highest goal) was achieved by 11 students. These figures are evidence that students used the challenge-me option and achieved the goals by solving complex problems.

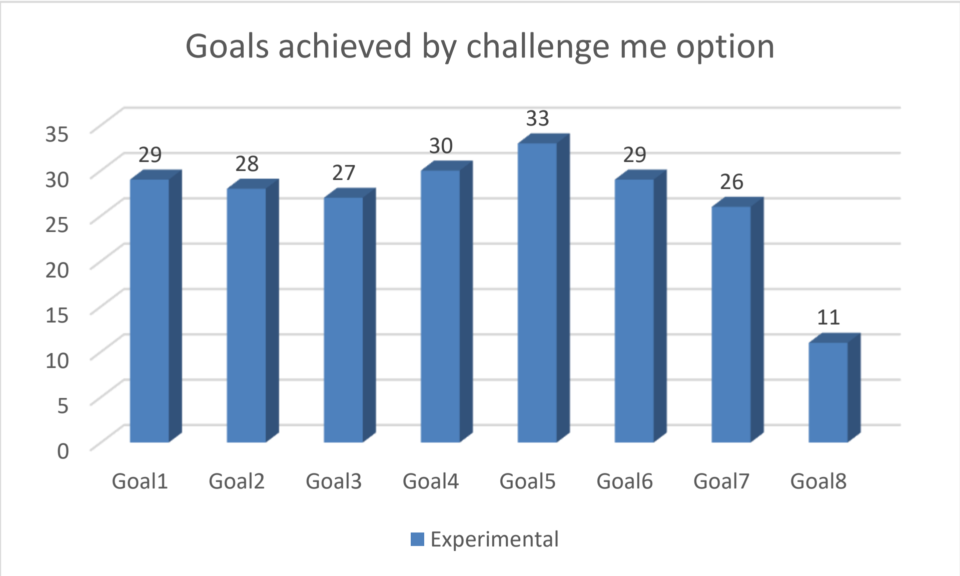


Figure. 5.9. Number of students achieving goals using challenge me option on the dashboard

To evaluate the effects of selecting complex problems on the learners' engagement and motivation, I divided the experimental group into two subgroups based on their medium complexity scores (3.2). The high challenge accepting group (HCAG) has an average complexity of more than 3.2, and the low challenge accepting group (LCAG) has an average complexity of less than 3.2. The differences between the two groups on problem-solving features are presented in Table 5.6

Table 5.6. *Comparison between HCAG and LCAG*

	LCAG (30)	HCAG (27)	Significance
Pre-test %	65.88 (24.3)	60.11 (25.1)	U = 351, p = .3
Post-test %	43.22 (38.3)	63.67 (29.9)	U = 133, p =. 1
Attempted problems	54.67 (46.61)	60.56 (35.17)	
Solved problems	54.03 (46.33)	59.11 (35.03)	
Number of attempts	190.77 (140.63)	283.56 (166.51)	U = 262.5, p < .05
Number of Constraints	253.23 (81.49)	324.48 (45.94)	U = 186, p < .001
Number of messages seen	320.63 (224.78)	535.41 (342.28)	U = 245, p < .05
Number of days working	3.03 (2.14)	4.11 (2.91)	
Number of logins	11.77 (14.94)	11.11 (7.09)	
Total Time (mins)	227.2 (177.89)	373.96 (264.73)	U = 252.5, p < .05

To compare the means between the two groups, I conducted a non-parametric Mann-Whitney U test. Table 5.6 presents the mean differences between HCAG and LCAG. The results show no significant difference between both groups at the pre-test level ($p = .30$), meaning that both groups have comparable prior knowledge. No difference ($p = .10$) was found between the post-test of the two groups. Both groups have attempted and solved a similar number of problems. However, HCAG had more attempts, covered more constraints, saw more feedback messages, and finally spent more time on SQL-Tutor than LCAG. Moreover, all these differences are statistically significant. These results provide insights into the learners' problem selection decisions. (1) The learners in the experimental group have improved their problem selection behaviour, (2) The dashboard motivated students to select complex problems, and (3) selecting more complex problems increased learners' engagement with the system in terms of time spent, number of attempts, feedback messages used, and constraints covered. These results confirmed the positive effects of the dashboard on learners' problem selection decisions and confirmed **H3**.

To evaluate hypothesis 4, I extracted the students' responses and time spent on the self-reflection prompts throughout their interaction with SQL-Tutor. I performed linear regression analysis with the average reflection time spent on the reflection prompts to predict the post-test scores. The results found that students' average self-reflection time significantly predicts their post-test scores *(β = 0.413, p < .01*). These results are aligned with Carpenter et al. (2020), who incorporated the students' reflections in a game-based learning environment and found that the students who spent more time reflecting through self-reflection prompts received higher post-test scores.

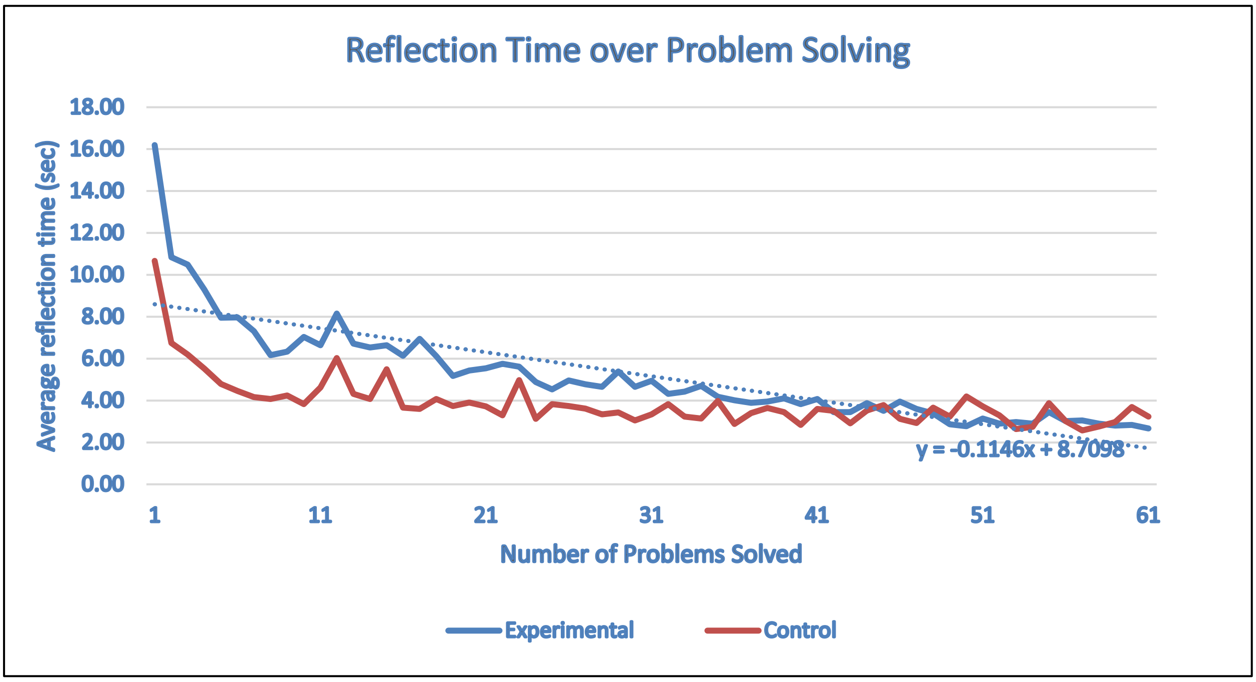


Figure 5.10. Time spent on self-reflection prompts for up to 60 problems

To find the change in learners' reflection time over the four weeks (while they interacted with SQL-Tutor), I have plotted the reflection time learning curve (Figure 5.10). Figure 5.10 illustrates that students spent time in the range of 10-16 seconds on reflection when they received the self-reflection prompt for the first time. However, this time decreased and reached 4-8 seconds after the first ten problems. From 20-40 problems, the time remained between 4-6 seconds, and after 40 problems, the student spent only 2-4 seconds on prompts. The slope (m= - 0.1) shows that time spent gradually decreased and remained steady after the student solved a few problems. Both experimental and control groups have similar models, and no significant difference was found. On the problem-solving features, learners' average reflection time is significantly related to attempted problems *(r = - 0.272, p<.01)* and solved problems (*r = - 0.247, p <.05*).

These findings summarize in two points: (1) Learners time on reflecting is associated with their learning outcomes, and (2) as the learners solved problems, their reflection time significantly decreased. These results support **H4**.

To evaluate this hypothesis, I divided my analysis into two Sections. The first Section (A) analyses students' understanding (Q1) and self-reflection (Q2) responses, and the second section deals with the student's satisfaction responses (Q3). The reason to divide the analysis is because Q1 and Q2 were given to experimental group only, but Q3 was given to both groups. . The three questions are positively correlated with the high-reliability coefficient (alpha = 0.918). To evaluate the hypothesis, I used the hierarchical multiple regression analysis in SPSS. The reason to select this test is its stepwise nature, which coincides with the problem-solving in SQL-Tutor.

The dataset for the first section (A) consisted of all the solved problems (n = 2,976) for the experimental group students, and their responses on the self-reflection prompt (2,976 x 4). The problem's complexity, number of attempts, and time spent on each problem were extracted as predicting variables. The problem complexity was regarded as the task analysis feature (forethought phase); the number of attempts and time spent on the problem were considered problem-solving features (performance phase).

Table 5.7 presents the hierarchical linear regression analysis of students' understanding. First, the students' understanding of the problem was negatively and significantly (*β = - 0.141, p < .001*) predicted by a problem's complexity and in the second step it was predicted by number of attempts (*β = .094, p < .001*), and time spent on problem (*β = - 0.087, p < .001*). The negative coefficient of problem complexity explains that students might have difficulty in understanding as the problem complexity grew. The same is the case with the time on the problem; more time led to less understanding. On the other hand, number of attempts positively affects student understanding.

Table 5.7. *Students' Understanding Predicted from Attempts, Time, and Problem Complexity*

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	β		
1	(Constant)	6.072	0.073		83.259	0.001
	Problem Complexity	-0.144	0.019	-0.141	-7.576	0.001
2	(Constant)	5.977	0.08		75.076	0.001
	Problem Complexity	-0.135	0.02	-0.132	-6.906	0.001
	Number of Attempts	0.005	0.001	0.094	5.04	0.001
	Time on problem	-0.001	0	-0.087	-4.596	0.001

a. Dependent Variable: Understanding

Table 5.8 presents the hierarchical linear regression analysis of student reflection. First, the student reflection on the topic was negatively and significantly (*β = - 0.112, p < .001*) predicted by problem's complexity, and in the second step, it was only predicted by the time spent on the problem (*β = -0.092, p < .001*). The negative coefficients of problem complexity and time on problem explain that students did not intensely reflect on the subject matter when the problem complexity and time increased.

Table 5.8. *Students' Reflection Predicted from Attempts, Time and Problem Complexity*

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	β		
1	(Constant)	5.548	0.08		69.122	0.001
	Problem Complexity	-0.126	0.021	-0.112	-6.007	0.001
2	(Constant)	5.558	0.088		63.174	0.001
	Problem Complexity	-0.106	0.022	-0.095	-4.91	0.001
	Number of Attempts	0.002	0.001	0.031	1.635	0.102
	Time on problem	-0.001	0	-0.092	-4.8	.001

a. Dependent Variable: Reflection

For section B, the data collection consisted of experimental and control groups' solved problems (control = 3,732 and experimental = 2,976) and their features. Table 5.9 presents the results of hierarchical linear regression analysis of student satisfaction as an outcome variable. First, the student's satisfaction was negatively and significantly (*β = - 0.153, $p < .001$*) predicted by problem's complexity and in the second step it was significantly predicted by number of attempts (*β = 0.108, $p < .001$*), and the time spent on problem (*β = -0.118, $p < .001$*). The negative coefficients of problem complexity and time on problem explain that students did not satisfy with their performance when the problem complexity and time on the problem increased.

Table 5.9. *Student's Satisfaction Predicted from Attempts, Time, and Problem Complexity*

Model		**Unstandardized Coefficients**		**Standardized Coefficients**	**t**	**Sig.**
		B	**Std. Error**	**β**		
1	(Constant)	5.993	0.073		82.097	0.001
	Problem Complexity	-0.157	0.019	-0.153	-8.238	0.001
2	(Constant)	5.895	0.079		74.357	0.001
	Problem Complexity	-0.143	0.02	-0.138	-7.298	0.001
	Number of Attempts	0.006	0.001	0.108	5.82	0.001
	Time on problem	-0.001	0	-0.118	-6.275	0.001

a. Dependent Variable: Satisfaction

These results provide insights into the reasons for specific responses to the self-reflection prompts. The reasons why students did not understand the principle of the problem, did not reflect on the subject matter, and were not satisfied with their progress were the selected problem complexity and time spent solving those problems. However, the number of attempts on the problem positively affects the student's understanding and motivation. These findings highlight the need for support when the problem is complex, or the learner takes more time solving it and indicates that decisions of each phase of the SRL framework (Zimmerman, 2000) affect the other phases. These results support **H5**. Moreover, no difference was found between the two groups on the "feelings" question responses.

5.4.3 RQ 8: Do SRL Interventions Affect Learners' SRL Skills and Motivation?

I received 77 (experimental = 44, control = 33) responses for Survey 1, and 35 (experimental = 21, control = 14) responses for Survey 2. I considered only those students who completed both surveys (n=35). Cronbach's alpha for both Survey 1 (alpha = 0.78) and Survey 2 (alpha = 0.74) are acceptable.

To test that SRL support interventions improved SRL skills of learners, first, I compared experimental and control groups on Survey 1 by using the Mann-Whitney U test. No differences were found between the two groups at the time of Survey 1 on goal setting ($U = 172, p > .05$), self-efficacy ($U = 132.5, p > .05$), strategic planning ($U = 134, p > .05$), task strategies ($U = 110.5, p > .05$), elaboration ($U = 143, p > .05$), self-evaluation ($U = 160, p > .05$), and help seeking ($U = 184, p > .05$).

As the second step, I compared the scores from Surveys 1 and 2 using the Wilcoxon Signed Rank test (Table 5.10). In the experimental group, goal setting improved significantly ($z = 1.93, p = .05$) but not in control group. For the control group, strategic planning ($z = -1.97, p = .05$), elaboration ($z = -2.45, p < .05$), and self-evaluation ($z = -2.44, p < .05$) significantly decreased from Survey 1 to Survey 2.

To compare Survey 2 responses of both groups, I again used the Mann-Whitney U test. There is a significant difference in goal setting ($U = 194, p < 0.05$), self-efficacy ($U = 211.5, p < 0.05$) and marginally significant difference in self-evaluation ($z = 1.60, p < 0.1$) between experimental and control groups.

The findings may help to explain the dynamics at play. Specifically, goal setting, self-efficacy, elaboration, and self-evaluation differed both as a function of condition group and time. These findings suggest that (a) students who would typically complete the tasks (in the absence of the intervention) tend to report lower strategic planning, self-evaluation and elaboration over time; and (b) the interventions may lead to considerable improvement in goal setting and self-efficacy, *especially* for students who started the tasks with less confidence. However, the intervention did not affect the other SRL skills. These findings only support the efficacy of goal-setting intervention and student's self-efficacy. However, the results remain inconclusive because of the low turnout on survey 2 and may support hypothesis 6 partially.

Table 5.10. *Comparison of Student Responses on Survey 1 & 2: mean (SD)*

SRL Skills	**Groups**	**Survey 1**	**Survey 2**
Goal Setting	*Experimental. (21)	3.56 (0.63)	3.95 (0.65)
	Control (14)	3.39 (0.64)	3.28 (0.65)
Self-Efficacy	Experimental. (21)	3.38 (0.65)	3.74 (0.65)
	Control (14)	3.5 (0.66)	2.98 (0.67)
Strategic Planning	Experimental. (21)	3.20 (0.88)	2.96 (0.87)
	*Control (14)	3.32 (0.86)	2.76 (0.86)
Task Strategies	Experimental. (21)	3.27 (0.70)	2.96 (0.69)
	Control (14)	3.40 (0.67)	3.05 (0.67)
Elaboration	*Experimental. (21)	3.98 (0.81)	3.24 (0.81)
	*Control (14)	4.05 (0.85)	3.00 (0.85)
Self-evaluation	Experimental. (21)	3.68 (0.75)	3.44 (0.75)
	*Control (14)	3.57 (0.72)	2.98 (0.72)
Help Seeking	Experimental. (21)	3.56 (0.78)	3.12 (0.78)
	Control (14)	3.34 (0.77)	2.98 (0.78)

*$p < .05$

5.5 Eye tracking and Emotion Detection on SRL Phases

Eye-tracking and emotion detection is another important dimension of this project. I conducted a lab study discussed in Chapter 3 with Tobii and iMotions to detect the emotions when students solved problems with SQL-Tutor, using worked examples. The study revealed that when students were stuck during problem solving, they experienced frustration and tended to leave SQL-Tutor. These findings paved the way to introduce motivational strategies like gamification (Chapter 4) and self-regulation learning (Chapter 5) when students experienced frustration and boredom during problem solving.

This experiment aims to detect and analyze students' emotions on three SRL interventions. In addition to the emotions, eye gaze analysis was also performed on these interventions. At the end of the experiment, participants filled in a survey about their experience of the dashboard.

The findings of this experiment contribute to the confidence in the interventions by determining whether they are producing any negative emotions in students. Moreover, the results facilitate the process of refining these interventions in the future. The research questions addressed in this section are the following:

RQ 9: What are the major emotions stimulated by SRL interventions?

RQ 10: What information did students find more useful on the self-reflection prompt and dashboard?

5.5.1 Study Design and Procedure

The SQL-Tutor version given to the participant in this experiment has the support for all three SRL interventions i-e., goal-setting support, dashboard, and self-reflection prompts. The recruited participants either had worked with the current version of SQL-Tutor or had previous experience with SQL-Tutor. Each participant had a separate session in which they worked with SQL-Tutor while the iMotions recorded their facial expression and eye gaze data. At the beginning of the session, the participants provided informed consent. They filled the demographic and emotional intensity questionnaire (the same questionnaire used in Chapter 3, Section 3.2) to collect demographic information and their familiarity with SQL-Tutor. The participants sat in front of the Tobii screen, and the standard Tobii calibration was completed. During the calibration test, each participant was requested to track a ball's movement on the screen. This calibration took 6 seconds, and the experiment started when the results were excellent. Otherwise, the participant's position was adjusted in front of the camera, and recalibration took place. After the calibration was completed, participants started problem solving with SQL-Tutor. Students were free to solve as many problems as they liked, and the sessions lasted 35-40 mins.

At the end of the session, participants were asked to fill in a questionnaire related to their experience with the dashboard (see Appendix). The questionnaire was composed of 12 questions evaluated on the five-point Likert scale ranging from strongly disagree (1) to strongly agree (5). The first four questions of the questionnaire asked about their experience with various dashboard elements, such as how well students understand various dashboard elements, which elements on the dashboard are most useful in learning. The following seven questions asked about the logical construction of various elements on the dashboard and their effectiveness on student's time management and other skills. The last open question asked the suggestions to improve the dashboard.

5.5.2 Findings

Ten undergraduate and post-graduate students participated in the study, of which 50% were females. Figure 5.11 shows the responses to the emotional intensity questionnaire. These responses revealed that participants got easily contented (mean = 5.9, SD = 1.29), excited (mean = 5.5, SD =1.65), and amused (mean =5.5, SD = 1.27), but it was rather hard for them to be easily saddened (mean= 3.8, SD =1.32), feared (mean =4.3, SD = 1.64), and angered (mean= 4.0, SD =1.70), as illustrated in Figure 5.11. Particularly for negative emotions, the participants stated that it was difficult to get fearful but strongly inclined towards contented and exciting emotions. These results are specific to the participants in the study and cannot be generalize due to low dataset.

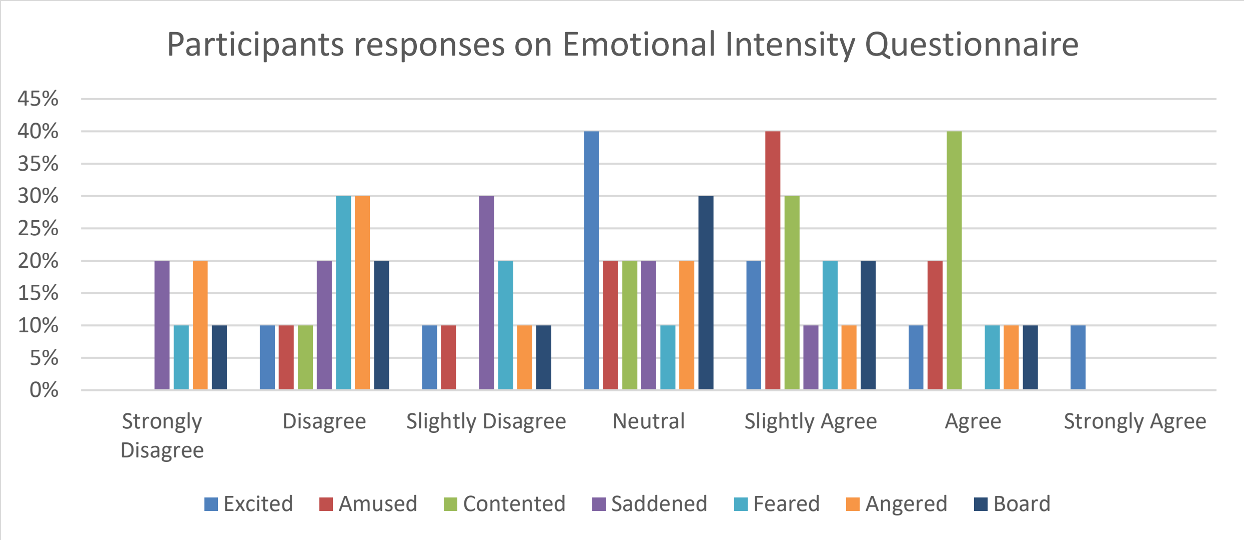


Figure 5.11. Seven emotions marked by ten participants

Affectiva AFFDEX generated probabilistic estimates for each of the seven emotions (anger, disgust, surprise, sadness, joy, fear, and contempt) based on each participant's micro-expressions (lasting 0.5-4 seconds). These estimates are called evidence scores which provide evidence of the presence of the emotion and its intensity. In the analysis, I have selected the amplitude-based thresholding to find out the strongest emotion. I selected absolute rates in amplitude thresholding, which showed the frequency with which participants displayed the emotion. I applied a threshold of 50% on these numeric scores, thus including only those emotions whose values were greater than 0.5. Due to the low occurrence of values, I have selected this threshold value to include more emotions. This threshold value was identical for each emotion and action unit.

5.5.2.1 RQ 9: What are the Strongest Emotions on SRL Interventions?

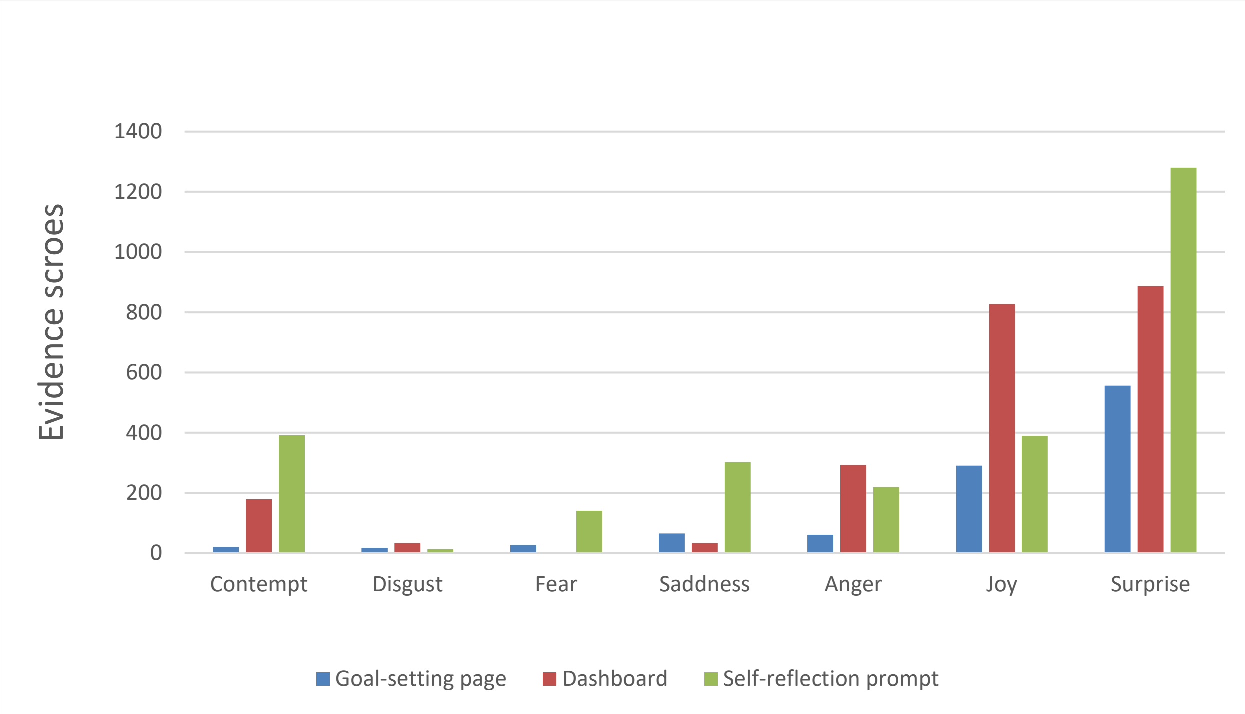


Figure 5.12. Emotions experienced by ten participants on three interventions

For this research question, I examined the participants' emotions while engaging with the SRL interventions. I have conducted three separate analyses for each intervention. For the goal-setting intervention, the participants' emotions on the goal-setting screen were examined. The list of goals and the messages underneath were marked as an area of interest (AOI). The whole screen was marked as an area of interest for the dashboard, and for the self-reflection prompt, the entire prompt box showing all four questions were considered an AOI. Participants' videos were segmented for three AOIs, and evidence scores for each emotion were recorded. Finally, the evidence scores of each AOI were summed to find the final evidence score of each emotion. Figure 5.12 shows the final evidence score of each emotion on three interventions.

Surprise, joy, and anger were the most experienced emotions across all the interventions. On the other hand, disgust and fear were the least experienced emotions. On the goal-setting screen, participants did not show many emotions while setting goals and reading system suggestions. The two powerful emotions were surprise and joy. The surprise was shown by more than 50% of participants. The second most potent emotion was joy, followed by sadness. Contempt, disgust, and fear were the least experienced emotions. Participants showed almost equal evidence scores for surprise and joy on the dashboard, followed by anger and contempt. Lastly, on self-reflection prompts, the highest emotion experienced was surprise, followed by joy, contempt, and sadness. Anger and fear were also reported with insufficient evidence scores on these prompts.

In summary, I found that goal setting and the dashboard did not cause any negative emotions. The reported happiness about the dashboard suggests that participants were pleased with their progress but were occasionally dissatisfied, resulting in anger (i.e., frustration). On the self-reflection prompts, participants felt surprised which can be a sign of confusion and sadness, fear, or even joy. Therefore, I can conclude that interventions increased the enjoyment with SQL-Tutor and did not ignite the negative emotions except for self-reflection prompts.

5.5.2.2 RQ 10: Which information do students find useful on dashboard and self-reflection prompts?

Participants' eye-tracking data were analyzed to determine where they looked at the self-reflection prompts and dashboard, how long they looked at them, and any changes in eye gaze patterns during the session. I started analyzing the videos in three stages: when participants looked at the interventions the first time (stage 1), when participants looked at them after solving 5-6 problems (stage 2), and finally when students looked at the interventions on their last problem (stage 3) before finishing their session. I focused on two differences in these three stages, 1) time and 2) gaze patterns.

Table 5.11. *Time (in seconds) Spent on the Dashboard in Three Stages*

Participant	Stage 1 (s)	Stage 2 (s)	Stage 3 (s)
1	45	4	3
2	7	8	3
3	21	17	3
4	1	3	4
5	4	2	4
6	25	23	12
7	4	2	2
8	26	14	4
9	18	7	4
10	15	9	2

Table 5.11 shows that students spent an average of 16.6s (SD = 16) on looking at the dashboard when it was first presented to them. However, this time spent on the dashboard started declining and reached half (mean = 8.9s, SD = 7.06) when they solved 5-6 problems and quarter (mean = 4.1s, SD = 2.89) at the end of the session. I conducted the heat map analysis on three stages for the gaze pattern, as shown in Figures 5.13, 5.14, and 5.15. Students looked at all three sections of the dashboard when they looked at the first time, but subsequently they only looked at their open learner model and progress on goal on completing average problems. At the end of their session, they focused again on their learner model and the total problems completed graph.

Figure 5.13. Eye gaze pattern for stage 1

Figure 5.14. Eye gaze pattern in stage 2

Figure 5.15. Eye gaze pattern in stage 3

As discussed in Section 5.4.4, the average reflection time varied between 10-16 seconds initially, and then it decreased as the number of problems increased. In these eye-tracking sessions, I have observed similar trends. The participants spent an average of 10-15 seconds in stage 1, 6-8 seconds in stage two, and merely 2-4 seconds in the last self-reflection prompt. The heat map analysis shows the eye gaze patterns in three stages: Figures 5.16, 5.17, and 5.18. The analysis revealed that participants looked at all the self-reflection questions when they first viewed the prompt. However, after solving a few problems, they only glanced at questions. At the end of the session, participants mainly focused on the performance-related question and only looked at the response radio buttons. Many prompts appear to have decreased learners' interest, as seen by these gaze patterns or they may get familiar with the prompts' wordings and therefore, only focused on the response buttons. However, their abiding interest in the performance question indicates that participants were already motivated for learning. These results give some insights into how students looked at the dashboard and self-reflection prompts during their course of study. Further analysis of these results is helpful to reveal the reasons for these behaviors and help refine these interventions.

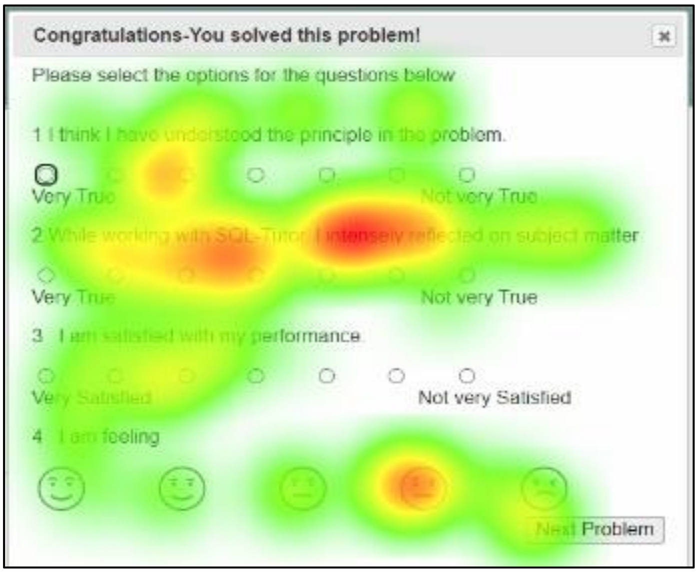


Figure 5.16. Eye gaze pattern for stage 1

Figure 5.17. Eye gaze pattern for stage 2

Figure 5.18. Eye gaze pattern for stage 3

5.5.2.3 Questionnaire Analysis

Out of ten students, seven students filled the after-study questionnaire. The first question asked about their understanding of the dashboard. The responses revealed that students understood time and problem-related information, for example, total time, total problems solved, session time, and weekly graphs very well. Students also understood the problem selection strategies (challenge me and practice problem), open learner model, complete solution option, and their progress on each goal. The student knowledge level, suggestions, and class progress were the least understood information.

In the second question, students were asked which information on the dashboard was most beneficial for their learning. The responses indicated that students found the open learner model, the progress bar for goals, and class progress the most helpful information for their learning. Whereas pre-test scores, total time, strategy, and suggestions provided the least helpful information for their learning.

In the third question, students were asked about which dashboard element motivated them to improve their learning. The responses to this question revealed that the progress bar for each goal, highest problem complexity, strategy, and student learner model as the most motivating elements on the dashboard. However, student knowledge level, total time, session time, complete solution used, and problem graphs were not motivating elements.

In the fourth question, students were asked about the elements which helped them reflect on their progress. Students indicated that the learner model and progress bar for goals were the most helpful elements in their reflection. They also indicated that the class progress bar, highest problem complexity, total and session problems were helpful in reflection. However, students were indifferent about the student knowledge level. Strategy, suggestion, and the complete solution used were the minor elements that helped in reflection. Furthermore, total and session time and weekly graphs were also not much help for students.

Table 5.12. *Percentage of Responses on Questions 5-10*

	Question Description	**Strongly Disagree**				**Strongly Agree**
Q5	Do you think the dashboard supported you in managing your time?	0	14%	29%	57%	0
Q6	Do you think the dashboard supported your problem selection strategy?	0	14%	29%	43%	14%
Q7	Does seeing the class progress assist your learning?	0	14%	71%	14%	0
Q8	Do you think there should be more suggestions on your learning strategy on dashboard?	0	0	86%	0	14%
Q9	Do you think the information on dashboard is constructed logically?	0	0	43%	57%	0
Q10	Do you think the information is delivered effectively?	0	0	29%	71%	0

Table 5.12 presents the responses (in percentages) on questions 5-10. Four out of seven students strongly agreed on the dashboard facilitating time management and problem selection (Q5 & 6). However, students were indifferent to the usefulness of suggestions and class progress bar in their learning (Q7 & 8). All students agreed that information on the dashboard

was constructed logically and delivered effectively (Q9 & 10). In the final open question, most students suggested adding a short explanation of a few elements such as student knowledge and the highest problem complexity. One student suggested that a bit of explanation of how these elements contribute to student learning may increase the student interest and help in reflection. The other recommendation was about the dashboard suggestions. Students found them not very effective for their learning and wanted to receive diverse "suggestions" on the dashboard.

5.6 Discussion

This Chapter presented a classroom study in which I analyzed the effects of self-regulated learning in the context of an intelligent tutoring system SQL-Tutor. Our findings highlight the effects of providing goal-setting support, dashboard, and self-reflection prompts under realistic conditions in a study that lasted four weeks.

Starting from Zimmerman's (2000) framework of self-regulated learning, I designed three interventions: goal-setting support, dashboard, and self-reflection prompts representing each framework phase. I hypothesized that these interventions motivate students and increase their SRL skills and learning outcomes. In the goal-setting support, I presented eight specific task-based goals to students. The experimental group received support messages to select a challenging goal during goal setting, whereas the control group students were free to choose any goal. The first finding revealed that the experimental group significantly improved their knowledge from pre- to post-test, but the control group did not. The evaluation of goal-setting intervention revealed that students who accepted the system suggested goals achieved more challenging goals and solved more complex problems by attempting and solving significantly fewer problems, particularly those who have lower prior knowledge. These results provide evidence that system suggestions invoked the metacognitive abilities within the students, and they started thinking about their selection of goals and progress. The post-hoc analysis further strengthens this claim that the students who selected the goals by analyzing their progress and system suggestions achieved greater student-level and problem-solving results. On the other hand, all the students in the non-intervention group selected goals in sequential order, starting from the easiest one, even those who scored highest on their pre-tests.

The second intervention was a dashboard which presented to experimental group students after solving each problem. The dashboard provided information about the student's problem-solving progress and two problem selection options. The findings revealed that the dashboard improved the problem selection behavior of students by attracting them to solve complex problems, which increased learners' engagement with the system. These results support the research that claims that dashboards helped improve the problem selection behavior of learners (Long et al.,2015).

Similar to the dashboard, self-reflection prompts were provided to the experimental group only. However, the control group students received the performance and feeling questions only. The intervention evaluation revealed that students' reflection time affects their learning outcomes. The possible explanation for this finding is that students received the prompts after completing every problem. Some of these questions might have been solved in less than a minute. This strategy might decrease the effectiveness and impartiality of the responses. I also found no difference between the negative and positive feelings and satisfaction of both groups.

The third finding of the study is predicting learners' response to self-reflection prompts from their problem selection and problem-solving behavior. The findings indicate that learners did not understand the problem fully when it was complex, and more time was required to solve the problem. This finding highlights the effects of problem selection and problem-solving actions on learners' thoughts. In addition, it focuses on the need for scaffolding in complex problems.

As mentioned in the literature review, the SRL affects learning and supports learners' SRL skills and motivation. I measured students' SRL skills and motivation with the help of surveys 1 and 2 administered at the start and end of the study. The findings revealed that the intervention group increased their goal setting and self-efficacy. However, self-evaluation, strategic planning, and elaboration have statistically decreased in the control group. The main reason for the inconclusion is the small sample size, as very few students attempted survey 2 which was entirely voluntary.

After this study, I conducted a small lab experiment to track learners' eye gaze and facial expressions on three interventions. The findings revealed that participants were delighted and surprised when presented with the goal setting and dashboard. However, self-reflection prompts caused negative emotions in them. One reason to experience these negative emotions could be many prompts that a participant had to respond to, and to accept that they did not understand the problem increased the sadness and fear. The higher levels of surprise could result from intervention novelty, as discussed in Graesser (2020).

The eye-tracking results provide evidence that participants looked at the interventions (dashboard and self-reflection prompt) in detail when first presented to them. This analysis revealed the vital information for the participants were the goals' progress and OLM on the dashboard and questions related to self-efficacy questions on the self-reflection prompt. These results have strengthened with the questionnaire responses, in which the most valuable and motivating elements on the dashboard were goal progress, OLM, and the highest problem complexity. The least motivating and helpful in learning were the percentage of complete solutions used, time-on-task and problems completed. However, this interest declined with the increased number of problems.

The most significant limitations of our study are the small sample size and the low completion rates for survey two and post-test. The second limitation is the study design, as the dashboard was not interactive, and not much data have been logged from the learners. Due to the unavailability of data, I cannot evaluate the effects of other elements of the dashboard beyond problem selection. The third limitation of the study is the poor estimation of when self-reflection prompt should be presented to the learners.

Chapter 6 Conclusions and Future Work

This Chapter begins with a summary of the thesis as well as my Ph.D. research questions. Following that, I discuss the contributions of my research endeavour. Last, I outline the research limitations, draw conclusions, and propose future directions for research.

6.1 Summary

This PhD research project aimed to influence learners' motivation and engagement with SQL-Tutor to improve their learning outcomes. Many versions of SQL-Tutor have been released, each with new features and support, such as a probabilistic student model (Mayo & Mitrovic, 2000), different problem selection strategies (Mathews, 2012; Mayo, 2001; Mitrovic et al., 2004), an animated pedagogical agent (Mitrovic & Suraweera, 2000), positive feedback (Barrow, Mitrovic, Ohlsson, & Grimley, 2008), worked examples (Najar, Mitrovic, & McLaren, 2014), and erroneous examples support (Chen, Mitrovic, & Mathews, 2016). However, most of these features were focused on building the cognitive skills of learners. Affective and motivational support, on the other hand, has been overlooked. The ITS provides metacognitive help in the form of an open learner model; however, it was not informed by the learners' behavioural analytics.

Chapter 1 discussed my motivation for this research project and the research questions followed by the proposed solution and research framework in Chapter 2. In the literature review (Chapter 2), I first presented a comprehensive overview of various ITSs, including their main characteristics and accomplished learning outcomes. Following that, I gave an account of SQL-Tutor, including its architecture and feature enhancements along with the major challenges. After that, I reviewed the most recent research on each of the proposed solution strands. There are two main strands of the proposed solution, accurately detecting learners' affective states and increasing motivation by using different strategies.

In Chapter 3, I explained the first study of the project. I identified the learners' affective states using automatic emotion detection software iMotions and the Tobii eye tracker while working on SQL-Tutor. The results provided the reasons to introduce our first motivational strategy, gamification. Chapter 4 presented a study that evaluated the game mechanics and gamified SQL-Tutor based on the theory of gamified learning (Landers, 2014). The findings of the gamification study paved the way to another strategy: self-regulated learning. Another study, reported in Chapter 5, evaluated the benefits of SRL support in SQL-Tutor by incorporating three interventions: goal-setting support, dashboards, and self-reflection prompts. In the second phase of this chapter, I presented a lab experiment that identified various affective states and gaze patterns of learners on the three interventions.

Following this brief overview of the study's organization, I conclude my findings for each research question.

RQ1: *Does iMotions accurately identify learners' emotions?* The purpose of this research question is to establish confidence on the results generated through iMotions. I recruited ten volunteers for this study and presented them with 48 IAPS photos while capturing and recording their emotions and eye gaze using iMotions and the Tobii eye tracker. The recorded emotions were then compared with an already established emotion categorization system. The findings revealed that iMotions could be used as reliable software to detect human emotions up to 77% accurately. However, these results can be improved by combining other techniques.

RQ2: *Do examples help problem solving in SQL-Tutor*? This research question was addressed in the second phase of this study. In this phase, the worked examples were provided to learners during problem solving in SQL-Tutor. While learners worked on the system, their facial expressions and eye gazing were recorded, especially when using worked examples. The results revealed that worked examples helped students mainly when the problems were complex and alleviated negative emotions (anger and fear) while increasing engagement and, up to some extent, joy.

The findings of this study motivated me to introduce the motivational strategy of gamification in SQL-Tutor. Thus, I added three game mechanics: goal setting, challenges, and self-assessment. All three-game mechanics were implemented in SQL-Tutor in the form of 13 badges and evaluated in a classroom study. Research questions 3-5 were addressed in this study.

RQ3: *What are the effects of gamification on learning?* The findings of this research question indicated that badges did not increase students' engagement and learning directly. However, the number of achieved badges positively affected time on task which influence the student's learning.

RQ4: *Do students with different levels of prior knowledge react differently to gamification?* The study's second research question showed that prior knowledge did not improve learners' time on task; however, it yielded significant positive effects when combined with badges. In other words, when learners achieved more badges, they spent more time on the system regardless of their prior knowledge.

RQ5: *What is the effect of gamification on student motivation?* In the third research question of this study, I evaluated student motivation by measuring their self-efficacy, perceived competence, and topic-interest. The students who received badges showed no significant improvement in their self-efficacy, perceived competence, and topic interest when compared to those who did not receive badges. However, higher topic interest enhanced the association between badges and time on task by encouraging students to earn more badges, according to the path analysis. Lower interest in SQL mixed with many badges pushed students to spend more time, but not as much as higher interest did.

The findings of the gamification study indicated no direct improvement in the motivation, engagement, and learning outcomes of students. These results raised the need to introduce another strategy that directly affects students' learning outcomes and influences learners' internal motivation. Therefore, in a subsequent study, I introduced the support of SRL in SQL-Tutor using three interventions (goal-setting support, dashboard, self-reflection prompts) based on Zimmerman's (2000) SRL framework. The effects of SRL support were evaluated in another classroom study, and research questions 6-8 were addressed.

RQ6: What are the effects of SRL support on student learning? The findings revealed that the students who received SRL support interventions significantly improved their learning outcomes by solving fewer but more complex problems than the students who did not receive SRL support.

The second finding of this study addressed in *RQ7: What are the (separate) effects of three interventions on students' learning behaviours*? The effects of the goal-setting support intervention revealed that students who accepted the system-suggested goals attained maximum complex goals while attempting and solving significantly fewer problems, especially those with lower prior knowledge. The dashboard analysis reported the improved problem selection behaviour of students by attracting them to solve complex problems, which increased learners' engagement with the system. Lastly, the time the learner spent on reflecting their thoughts on self-reflection prompts positively predicted their learning outcomes.

Additionally, the responses on self-reflection prompts showed that when complexity and time on a problem increased, the learners' understanding and satisfaction with their progress decreased. However, more number of attempts on the problems positively predicts their understanding and satisfaction.

RQ8: What are the effects of SRL interventions on learners' SRL skills and motivation? The findings revealed that students who received SRL support interventions increased their goal setting and self-efficacy. Students in the non-intervention group, showed a decline in strategic planning, self-evaluation, and elaboration at the end of the study. The analysis did not reach a conclusive result for the remaining SRL skills because of the low response rate on survey 2.

After this study, I conducted a lab experiment with the SRL intervention-based SQL-Tutor. This study aimed to determine how learners' affective states and gaze patterns changed upon interacting with SRL interventions.

RQ9: What are the major emotions on each SRL intervention? The experiment's findings revealed that participants were delighted and surprised when presented with the goal setting and dashboard but felt surprised, sadness and fear on self-reflection prompts.

The findings related to research question 10 (*R Q 10: What information did students find more valuable on the self-reflection prompt and dashboard?)* evident that participants looked at the interventions (dashboard and self-reflection prompt) in detail when first presented to them. However, this time decreased as the number of problems solved increased. After solving a few problems, the participants only looked at their OLM and goal progress bar on the dashboard and response radio buttons on the self-reflection prompts. After study questionnaire responses strengthened these results where students explicitly mentioned the OLM, goal progress bar, and highest problem complexity as the most useful elements on the dashboard.

6.2 Contributions

One of the significant contributions of this Ph.D. research project is identifying and evaluating learners' affective states. CBM tutors, for example, SQL-Tutor, developed with the focus of improving the cognitive skills of learners provides limited metacognitive support (in the form of an OLM) and no direct support for and regulation of learners' affective states, which is the most crucial ingredient in the research of engagement and motivation. Therefore, in this project, I have taken the first step towards affective state support in the CBM-based SQL-Tutor.

The second contribution of this research project is investigating the effects of gamification on the learning and motivation of undergraduate students. Gamification has gained much popularity in learning environments for increasing learners' engagement and motivation. However, with ITSs being serious and mature learning environments, its effects have not been studied on the undergraduate level (the age group that is difficult to entice with intangible rewards). To my knowledge, this is the first empirical study that evaluated the effects of gamification in ITS for this age group under realistic conditions.

The reason for the adoption of SRL in learning environments is that SRL's effect on learning is well established. However, not many learning environments have accommodated all the SRL framework phases, primarily encompassing only one or two of those phases. Another contribution of this project is to extend SQL-Tutor to facilitate students' self-regulated learning by designing and implementing activities based on the complete SRL framework. The separate and combined effects of each activity/intervention have been evaluated on student learning and motivation. These are the three most significant contributions of this research project.

6.3 Limitations

In all the classroom studies and lab experiments of this research project, the biggest problem was the low sample size. Affect identification studies usually need a rich dataset for avoiding noise and other factors such as novelty effects. However, the participation rate in the affect detection studies did not get more than ten participants. Another limitation in these studies were that iMotions and the Tobii eye tracker were attached only to a single machine, and separate sessions had to be conducted for each participant. Although the lab setup conforms to the standard of conducting such studies, the amount of light and other individual differences affected the results. The accuracy of iMotions' results was found to be 77% in my dataset, though this percentage is better than those reported in previous studies. Nevertheless, this accuracy is another limitation of results.

Beyond the affect detection studies, small sample size is the most important limitation of all the studies in this research project. In the pre-and post-test setup, the post-tests were always voluntary in these studies and administered only two days before the final lab test, making the students' participation very low. Similarly, the surveys in both studies received inadequate responses from learners. In the gamification study, daily challenges were implemented to increase student motivation. However, students should not use the complete solution option extensively to activate daily challenges. Only seven students fulfilled the condition to receive daily challenges.

Another limitation was the design of the SRL study, which is based on the SRL framework. As mentioned in the literature review (Chapter 2), the challenge in SRL research is finding and collecting the data that measure the exact SRL processes. The dashboard intervention provided much information about students' progress in the study, but I did not collect the data beyond the student problem selection strategy. Moreover, survey 1 in the study consists of 29 questions that usually took around 20 minutes to complete. However, the users' session logged out automatically after 15 mins of inactivity. This underestimation of the session time lost the data of approximately 20 participants on survey 1. I did not receive many responses on survey 2, and therefore, hypothesis 7 remained inconclusive.

6.4 Future Directions

Like any other research project, there are countless possibilities for future directions from this research project. This project detected the affective state of learners for the first time in SQL-Tutor while problem solving. Therefore, the first future direction could be to extend this study with more participants that will increase the accuracy of emotions detected by iMotions and provide strong evidence of the affective state of learners, which will help formulate the affect regulation strategy. Furthermore, affective states determined in this study can be used to develop affect regulation strategies. For example, the worked examples, interactive examples, or topic-related tutorials can be used as regulation strategies when the learners are experiencing anger or sadness. As mentioned in the analysis of the Chapter 5, that students selected the negative emoji face when they had spent much time on problem solving, or they were dealing with complex problems. Therefore, a cognitive strategy that helps students in both these situations can be an immediate next step in affect detection and regulation direction.

In the gamification study, students' opinion was not very favorable about the badges. In the future, the design and texture of the badges can be changed to make them more attractive and enticing. Badges can replace or be combined with other game mechanics, for example, a leaderboard for increasing competition within the class, various milestones, or challenges in the form of complex problems. According to the gamification study's findings, the students who had previous gamification experience and a higher interest in SQL won more badges, than those who either had no gamification experience or had a lower interest in SQL. These findings can be used to create adaptive gamification based on the user's profile and interests.

The findings of the SRL support study indicated the positive effects of SRL interventions on learning outcomes. Therefore, one of the possible future directions could be introducing interventions for other SRL processes, such as time management or help-seeking. The dashboard developed for this study showed a comprehensive overview of learners’ effort and progress in SQL-Tutor. However, the responses on the questionnaire revealed that the dashboard needs refinement and should be more diverse and interactive. In the light of these opinions, the next step is to refine the dashboard with more valuable and motivating elements combining with interactivity and predictive analytics. In the future, I can compare various learning behaviours between two techniques, for instance, how students selected problems in the presence of badges versus when the dashboard. Additionally, the SRL support study can be repeated with the different populations for analysing the cultural, demographic, previous knowledge, and ability effects on the learners' motivation and learning.

References

Abramovich, S., Schunn, C., & Higashi, R. M. (2013). Are badges useful in education?: It depends upon the type of badge and expertise of learner. *Educational Technology Research and Development, 61*(2), 217-232.

Ainley, M., Hidi, S., & Berndorff, D. (2002). Interest, learning, and the psychological processes that mediate their relationship. *Journal of Educational Psychology, 94*(3), 545.

Aleven, V., Roll, I., McLaren, B. M., & Koedinger, K. R. (2010). Automated, unobtrusive, action-by-action assessment of self-regulation during learning with an intelligent tutoring system. *Educational Psychologist, 45*(4), 224-233.

Alhammad, M. M., & Moreno, A. M. (2018). Gamification in software engineering education: A systematic mapping. *Journal of Systems and Software, 141*, 131-150.

Anderson, J. R. (1996). *The Architecture of Cognition* (Vol. 5): Psychology Press.

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

Arroyo, I., Beal, C., Murray, T., Walles, R., & Woolf, B. P. (2004). Web-based intelligent multimedia tutoring for high stakes achievement tests. International Conference on Intelligent Tutoring Systems (pp. 468-477). Springer, Berlin, Heidelberg.

Arroyo, I., Murray, T., Woolf, B. P., & Beal, C. (2004). Inferring unobservable learning variables from students' help seeking behavior. In *International Conference on Intelligent Tutoring Systems* (pp. 782-784). Springer, Berlin, Heidelberg.

Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Meheranian, H., Fisher, D., ... & Woolf, B. P. (2007). Repairing disengagement with non-invasive interventions. International Conference on Intelligent Tutoring Systems (pp. 195-202).

Arroyo, I., Woolf, B. P., Burelson, W., Muldner, K., Rai, D., & Tai, M. (2014). A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. *International Journal of Artificial Intelligence in Education, 24*(4), 387-426.

Atkinson, R. K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology, 94*(2), 416.

Azevedo, R. (2002). Beyond intelligent tutoring systems: Using computers as METAcognitive tools to enhance learning? *Instructional Science, 30*(1), 31-45.

Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *Human-Computer Studies, 68*(4), 223-241.

Bandura, A. (1986). Social foundations of thought and action. *Englewood Cliffs, NJ, 1986*, 23-28.

Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of personality and social psychology, 41*(3), 586.

Barrett, L. F., & Bliss-Moreau, E. (2009). Affect as a psychological primitive. *Advances in Experimental Social Psychology, 41*, 167-218.

Barrow, D., Mitrovic, A., Ohlsson, S., & Grimley, M. (2008). Assessing the impact of positive feedback in constraint-based tutors. In *International Conference on Intelligent Tutoring Systems* (pp. 250-259). Springer, Berlin, Heidelberg.

Biswas, G., Leelawong, K., Schwartz, D., Vye, N., & Vanderbilt, T. (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence, 19*(3-4), 363-392.

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.

Bodily, R., Kay, J., Aleven, V., Jivet, I., Davis, D., Xhakaj, F., & Verbert, K. (2018). Open learner models and learning analytics dashboards: a systematic review. In Proceedings of the 8th international conference on learning analytics and knowledge (pp. 41-50).

Borracci, G., Gauthier, E., Jennings, J., Sale, K., & Muldner, K. (2020). The effect of assistance on learning and affect in an algebra tutor. *Journal of Educational Computing Research, 57*(8), 2032-2052.

Broadbent, J. (2017). Comparing online and blended learner's self-regulated learning strategies and academic performance. *The Internet and Higher Education, 33*, 24-32.

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education, 27*, 1-13.

Broadbent, J., Panadero, E., Lodge, J. M., & de Barba, P. (2020). Technologies to enhance self-regulated learning in online and computer-mediated learning environments. In *Handbook of Research in Educational Communications and Technology* (pp. 37-52). Springer, Cham.

Carpenter, D., Geden, M., Rowe, J., Azevedo, R., & Lester, J. (2020). Automated analysis of middle school students' written reflections during game-based learning. In International Conference on Artificial Intelligence in Education (pp. 67-78). Springer, Cham.

Chen, L., Lu, M., Goda, Y., & Yamada, M. (2019). Design of learning analytics dashboard supporting metacognition. In Proceedings of 16th International Conference Cognition and Exploratory Learning in Digital Age (pp. 175-182).

Chen, X., Mitrovic, A., & Mathews, M. (2016). Do Erroneous Examples Improve Learning in Addition to Problem Solving and Worked Examples? In International Conference on Intelligent Tutoring Systems (pp. 13-22). Springer, Cham.

Collins, A., Brown, J. S., & Newman, S. E. (1988). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. *Thinking: The Journal of Philosophy for Children, 8*(1), 2-10.

Conati, C., & Vanlehn, K. (2000). Toward computer-based support of meta-cognitive skills: A computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education, 11*, 389-415.

Corti, K. (2006). Games-based Learning; a serious business application. *Informe de Pixel Learning, 34*(6), 1-20.

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

Csikszentmihalyi, M., & Csikzentmihaly, M. (1990). *Flow: The Psychology of Optimal Experience* (Vol. 1990): Harper & Row New York.

Denny, P. (2013, April). The effect of virtual achievements on student engagement. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 763-772).

Dever, D. A., Cloude, E. B., & Azevedo, R. (2021). Examining Learners' Reflections over Time During Game-Based Learning. In International Conference on Artificial Intelligence in Education (pp. 129-133). Springer, Cham.

Diana, N., Stamper, J., & Koedinger, K. (2018). An instructional factors analysis of an online logical fallacy tutoring system. In International Conference on Artificial Intelligence in Education (pp. 86-97). Springer, Cham.

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction, 22*(2), 145-157.

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction, 29*, 153-170.

D'Mello, S., Lehman, B., Sullins, J., Daigle, R., Combs, R., Vogt, K., ... & Graesser, A. (2010). A time for emoting: When affect-sensitivity is and isn't effective at promoting deep learning. In International conference on intelligent tutoring systems (pp. 245-254). Springer, Berlin, Heidelberg.

D'Mello, S. K., & Graesser, A. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-Adapted Interaction, 20*(2), 147-187.

D'Mello, S. K., Lehman, B., & Graesser, A. (2011). A Motivationally Supportive AffectSensitive AutoTutor. In Calvo, R. A., & D'Mello, S. K. (Eds.), New Perpectives on Affect and Learning Technologies (pp. 113-126). New York: Springer.

Deci, E. L., & Ryan, R. M. (2010). Intrinsic motivation. *The Corsini Encyclopedia of Psychology,* 1-2.

Denny, P., McDonald, F., Empson, R., Kelly, P., & Petersen, A. (2018). Empirical support for a causal relationship between gamification and learning outcomes. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (pp. 1-13).

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining" gamification". In Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments (pp. 9-15).

del Soldato, T., & du Boulay, B. (1995). Implementation of Motivational Tactics in Tutoring Systems. *International Journal of Artificial Intelligence in Education*, 6(4), 337-378.

Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in education: A systematic mapping study. *Educational Technology & Society, 18*(3), 75-88.

Dörrenbächer, L., & Perels, F. (2016). More is more? Evaluation of interventions to foster self-regulated learning in college. *International Journal of Educational Research, 78,* 50-65.

Duffy, M. C., & Azevedo, R. (2015). Motivation matters: Interactions between achievement goals and agent scaffolding for self-regulated learning within an intelligent tutoring system. *Computers in Human Behavior, 52*, 338-348.

Ekkekakis, P. (2012). Affect, mood, and emotion. *Measurement in Sport and Exercise Psychology, 321*.

Ekman, P. (1999). Facial expressions. *Handbook of Cognition and Emotion, 16*(301), e320.

Ellis, H., Heppell, S., Kirriemuir, J., Krotoski, A., & McFarlane, A. (2006). Unlimited learning: Computer games in the learning landscape. London: Entertainment and Leisure Software Publishers Association. Retrieved August 30, 2007

Geisinger, K. F. (2016). 21st century skills: What are they and how do we assess them?. *Applied Measurement in Education*, *29*(4), 245-249.

Graesser, A. C. (2020). Emotions are the experiential glue of learning environments in the 21st century. *Learning and Instruction*, *70*, 101212.

Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, *48*(4), 612-618.

Graesser, A. C., Conley, M. W., & Olney, A. (2012). Intelligent Tutoring Systems. *APA Educational Psychology Handbook, Vol 3: Application to Learning and Teaching.*, 451-473.

Graesser, A. C., & D'Mello, S. K. (2019). *U.S. Patent No. 10,388,178*. Washington, DC: U.S. Patent and Trademark Office.

Graesser, A. C., D'Mello, S., & Person, N. (2009). 19 meta-knowledge in tutoring. *Handbook of metacognition in education,* 361.

Graesser, A. C., Hu, X., Nye, B. D., VanLehn, K., Kumar, R., Heffernan, C., . . . Rus, V. (2018). ElectronixTutor: an intelligent tutoring system with multiple learning resources for electronics. *International Journal of STEM Education*, *5*(1), 15.

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

Graesser, A. C., VanLehn, K., Rosé, C. P., Jordan, P. W., & Harter, D. (2001). Intelligent tutoring systems with conversational dialogue. *AI Magazine*, *22*(4), 39-39.

Haaranen, L., Ihantola, P., Hakulinen, L., & Korhonen, A. (2014). How (not) to introduce badges to online exercises. In Proceedings of the 45th ACM Technical Symposium on Computer Science Education (pp. 33-38).

Halverson, L. R., & Graham, C. R. (2019). Learner engagement in blended learning environments: A conceptual framework. *Online Learning*, *23*(2), 145-178.

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work?--a literature review of empirical studies on gamification. In 47th Hawaii international conference on system sciences (pp. 3025-3034).

Hamm, J., Kohler, C.G. Gur, R.C., & Verma, R. (2011). Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders. *Journal of neuroscience methods*, 200(2), 237-256.

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education*, *80*, 152-161.

Hayes, A. F. (2017). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach.* Guilford publications.

Heffernan, N. T., & Heffernan, C. L. (2014). The ASSISTments Ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education, 24*(4), 470-497.

Helmefalk, M. (2019). An interdisciplinary perspective on gamification: Mechanics, psychological mediators and outcomes. *International Journal of Serious Games, 6*(1), 3-26.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational psychologist, 41*(2), 111-127.

Hooshyar, D., Kori, K., Pedaste, M., & Bardone, E. (2019). The potential of open learner models to promote active thinking by enhancing self-regulated learning in online higher education learning environments. *British Journal of Educational Technology, 50*(5), 2365-2386.

Hooshyar, D., Pedaste, M., Saks, K., Leijen, Ä., Bardone, E., & Wang, M. (2020). Open learner models in supporting self-regulated learning in higher education: A systematic literature review. *Computers & Education, 154*, 103878.

Houlfort, N., Koestner, R., Joussemet, M., Nantel-Vivier, A., & Lekes, N. (2002). The impact of performance-contingent rewards on perceived autonomy and competence. *Motivation and Emotion, 26*(4), 279-295.

Johnson, W. L., & Lester, J. C. (2016). Face-to-face interaction with pedagogical agents, twenty years later. *International Journal of Artificial Intelligence in Education, 26*(1), 25-36.

Kinnebrew, J. S., Segedy, J. R., & Biswas, G. (2014). Analyzing the temporal evolution of students' behaviors in open-ended learning environments. *Metacognition and learning, 9*(2), 187-215.

Kirsh, D. (2005). Metacognition, distributed cognition and visual design. *Cognition, Education, and Communication Technology*, 147-180.

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

Klock, A. C. T., Gasparini, I., Pimenta, M. S., & Hamari, J. (2020). Tailored gamification: A review of literature. *International Journal of Human-Computer Studies, 144*, 102495.

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education, 8*(1), 30-43.

Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. *International Journal of Information Management, 45*, 191-210.

Kort, B., Reilly, R., & Picard, R. W. (2001). An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion. In Proceedings IEEE international conference on advanced learning technologies (pp. 43-46).

Kulke, L., Feyerabend, D., & Schacht, A. (2020). A comparison of the Affectiva iMotions Facial Expression Analysis Software with EMG for identifying facial expressions of emotion. *Frontiers in Psychology, 11*, 329.

Kwang Sik, C., & Sungho, C. (2018). Affective Learning Reaction Model Based on Smartphone and Sensors. *Journal of Physics: Conference Series, 1060*(1), 012040.

Lachman, S. J. (1997). Learning is a process: Toward an improved definition of learning. *The Journal of Psychology, 131*(5), 477-480.

Landers, R. N. (2014). Developing a theory of gamified learning: Linking serious games and gamification of learning. *Simulation & Gaming, 45*(6), 752-768.

Landers, R. N., Bauer, K. N., & Callan, R. C. (2017). Gamification of task performance with leaderboards: A goal setting experiment. *Computers in Human Behavior, 71*, 508-515.

Landers, R. N., & Landers, A. K. (2014). An empirical test of the theory of gamified learning: The effect of leaderboards on time-on-task and academic performance. *Simulation & Gaming, 45*(6), 769-785.

Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (2008). International affective picture system(IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8. University of Florida,Gainesville, FL.

Lee, Y., & Choi, J. (2011). A review of online course dropout research: Implications for practice and future research. *Educational Technology Research and Development*, 59, 593–618.

Leelawong, K., & Biswas, G. (2008). Designing learning by teaching agents: The Betty's Brain system. *International Journal of Artificial Intelligence in Education, 18*(3), 181-208.

Legaki, N.-Z., Xi, N., Hamari, J., Karpouzis, K., & Assimakopoulos, V. (2020). The effect of challenge-based gamification on learning: An experiment in the context of statistics education. *International Journal of Human-Computer Studies, 144*, 102496

Long, Y., & Aleven, V. (2014). Gamification of joint student/system control over problem selection in a linear equation tutor. In International Conference on Intelligent Tutoring Systems (pp. 378-387). Springer, Cham.

Long, Y., Aman, Z., & Aleven, V. (2015). Motivational design in an intelligent tutoring system that helps students make good task selection decisions. In International Conference on Artificial Intelligence in Education (pp. 226-236). Springer, Cham.

Long, Y., & Aleven, V. (2016). Mastery-oriented shared student/system control over problem selection in a linear equation tutor. In International conference on intelligent tutoring systems (pp. 90-100). Springer, Cham.

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, *106*(4), 901.

Marder, A. (2015). Stack overflow badges and user behavior: an econometric approach. In 2015 IEEE/ACM 12th Working Conference on Mining Software Repositories (pp. 450-453). IEEE.

Mathews, M. M. (2012). A Framework for Multiple Adaptable Pedagogical Strategies in Intelligent Tutoring Systems.

Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science*, *26*(1-2), 49-63.

Mayer, R. E. (2011). Multimedia learning and games.

Mayo, M., & Mitrovic, A. (2000). Using a probabilistic student model to control problem difficulty. In International conference on intelligent tutoring systems (pp. 524-533). Springer, Berlin, Heidelberg.

Mayo, M. J. (2001). Bayesian student modelling and decision-theoretic selection of tutorial actions in intelligent tutoring systems.

Martinez-Miron, E., Harris, A., du Boulay, B., Luckin, R., & Yuill, N. (2005). The role of learning goals in the design of ILEs: Some issues to consider. In Looi, C. K., McCalla, G., Bredeweg, B., & Breuker, J. (Eds.), Artificial intelligence in education, 2005.

McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy training for active reading and thinking. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 222-233.

Mikels, J. A., Fredrickson, B. L., Larkin, G. R., Lindberg, C. M., Maglio, S. J., & Reuter-Lorenz, P. A. (2005). Emotional category data on images from the International Affective Picture System. *Behavior research methods*, *37*(4), 626-630.

Mitchell, A., & Savill-Smith, C. (2004). The use of computer and video games for learning. *A review of the literature*.

Mitchell, T. R. (1982). Motivation: New directions for theory, research, and practice. *Academy of management review*, *7*(1), 80-88.

Mitrovic, A. (1998). A knowledge-based teaching system for SQL. In *Proceedings of ED-MEDIA* (Vol. 98, pp. 1027-1032).

Mitrovic, A. (2002). NORMIT: A web-enabled tutor for database normalization. In International Conference on Computers in Education, 2002 (pp. 1276-1280).

Mitrovic, A. (2003). An intelligent SQL tutor on the web. *International Journal of Artificial Intelligence in Education*, *13*(2-4), 173-197.

Mitrovic, A., Koedinger, K. R., & Martin, B. (2003). A comparative analysis of cognitive tutoring and constraint-based modeling. In International Conference on User Modeling (pp. 313-322). Springer, Berlin, Heidelberg.

Mitrovic, A., & Ohlsson, S. (1999). Evaluation of a constraint-based tutor for a database language.

Mitrovic, A., Ohlsson, S., & Barrow, D (2013). The effect of positive feedback in a constraint-based intelligent tutoring system. *Computers & Education*, *60*(1), 264-272.

Mitrovic, A., & Martin, B. (2000). Evaluating the effectiveness of feedback in SQL-Tutor. In Proceedings International Workshop on Advanced Learning Technologies. Advanced Learning Technology: Design and Development Issues (pp. 143-144).

Mitrovic, A., & Martin, B. (2002). Evaluating the effects of open student models on learning. In International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (pp. 296-305). Springer, Berlin, Heidelberg.

Mitrovic, A., & Martin, B. (2007). Evaluating the effect of open student models on self-assessment. *International Journal of Artificial Intelligence in Education*, *17*(2), 121-144.

Mitrovic, A., Mayo, M., Suraweera, P., & Martin, B. (2001). Constraint-based tutors: a success story. In International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (pp. 931-940). Springer, Berlin, Heidelberg.

Mitrovic, A., & Ohlsson, S. (2016). Implementing CBM: SQL-Tutor After Fifteen Years. *International Journal of Artificial Intelligence in Education, 26*(1), 150-159.

Mitrovic, A., & Suraweera, P. (2000). Evaluating an animated pedagogical agent. In International Conference on Intelligent Tutoring Systems (pp. 73-82). Springer, Berlin, Heidelberg.

Mitrovic, A., Suraweera, P., Martin, B., & Weerasinghe, A. (2004). DB-suite: Experiences with three intelligent, web-based database tutors. *Journal of Interactive Learning Research, 15*(4), 409-432.

Munshi, A., Rajendran, R., Ocumpaugh, J., Biswas, G., Baker, R. S., & Paquette, L. (2018). Modeling learners' cognitive and affective states to scaffold SRL in open-ended learning environments. In Proceedings of the 26th conference on user modeling, adaptation and personalization (pp. 131-138).

Najar, A. S., Mitrovic, A., & McLaren, B. M. (2014). Adaptive support versus alternating worked examples and tutored problems: which leads to better learning? In International Conference on User Modeling, Adaptation, and Personalization (pp. 171-182). Springer, Cham.

Najar, A.S., Mitrovic, A., & Neshatian, K. (2015). Eye tracking and studying examples: how novices and advanced learners study SQL examples. *Journal of computing and information technology, 23*(2), 171-190

Nalepa, G. J., Kutt, K., & Bobek, S. (2019). Mobile platform for affective context-aware systems. *Future Generation Computer Systems, 92*, 490-503.

Nicholson, S. (2015). A recipe for meaningful gamification. In *Gamification in Education and Business* (pp. 1-20): Springer.

Nye, B. D., Karumbaiah, S., Tokel, S. T., Core, M. G., Stratou, G., Auerbach, D., & Georgila, K. (2018). Engaging with the scenario: Affect and facial patterns from a scenario-based intelligent tutoring system. In *International Conference on Artificial Intelligence in Education* (pp. 352-366). Springer, Cham.

O'Donovan, S., Gain, J., & Marais, P. (2013). A case study in the gamification of a university-level games development course. In Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference (pp. 242-251).

Ocumpaugh, J. (2015). *Baker Rodrigo Ocumpaugh monitoring protocol (BROMP) 2.0 technical and training manual*. Technical Report. New York, NY: Teachers College, Columbia University. Manila, Philippines: Ateneo Laboratory for the Learning Sciences.

Ohlsson, S. (1996). Learning from Performance errors. *Psychological Review*, *103*(2), 241.

Palthepu, S., Greer, J. and McCalla, G. (1991), Learning by teaching, The International Conference on the Learning Sciences, Evanston, Illinois, AACE, 357-363.

Panadero, E., Klug, J., & Järvelä, S. (2016). Third wave of measurement in the self-regulated learning field: When measurement and intervention come hand in hand. *Scandinavian Journal of Educational Research*, *60*(6), 723-735.

Pekrun, R. (2006). The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educational Psychology Review, 18*(4), 315-341.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational psychologist*, *37*(2), 91-105.

Pekrun, R., & Perry, R. P. (2014). Control-value theory of achievement emotions. *International handbook of emotions in education*, 120-141.

Picard, R. W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D. Strohecker, C. (2004). Affective learning—a manifesto. *BT Technology Journal*, *22*(4), 253-269.

Prensky, M. (2003). Digital game-based learning. *Computers in Entertainment, 1*(1), 21-21.

Rieber, L. P. (1996). Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational Technology Research and Development, 44*(2), 43-58.

Rowe, J. P., Shores, L. R., Mott, B. W., & Lester, J. C. (2011). Integrating learning, problem solving, and engagement in narrative-centered learning environments. *International Journal of Artificial Intelligence in Education, 21*(1-2), 115-133.

Sanford, K., & Madill, L. (2006). Resistance through video game play: It's a boy thing. *Canadian Journal of Education/Revue canadienne de l'éducation*, 287-306.

Schunk, D. H. (1984). Enhancing self-efficacy and achievement through rewards and goals: Motivational and informational effects. *The Journal of Educational Research, 78*(1), 29-34.

Schunk, D. H. (1991). Self-efficacy and academic motivation. *Educational Psychologist, 26*(3-4), 207-231.

Schunk, D. H., & DiBenedetto, M. K. (2020). Motivation and social cognitive theory. *Contemporary Educational Psychology, 60*, 101832.

Schunk, D. H., & Ertmer, P. A. (2000). Self-regulation and academic learning: Self-efficacy enhancing interventions. In *Handbook of self-regulation* (pp. 631-649). Academic Press.

Schunk, D. H., & Zimmerman, B. J. (1998). *Self-regulated learning: From teaching to self-reflective practice*: Guilford Press.

Sitzmann, T., & Ely, K. (2011). A meta-analysis of self-regulated learning in work-related training and educational attainment: what we know and where we need to go. *Psychological Bulletin, 137*(3), 421.

Shneyderman (2001), "Evaluation of the Cognitive Tutor Algebra I Program," Miami-Dade County Public Schools Office of Evaluation and Research 1500, 33132.

Stöckli, S., Schulte-Mecklenbeck, M., Borer, S., & Samson, A. C. (2018). Facial expression analysis with AFFDEX and FACET: A validation study. *Behavior research methods*, *50*(4), 1446-1460

Suh, A., Wagner, C., & Liu, L. (2018). Enhancing user engagement through gamification. *Journal of Computer Information Systems, 58*(3), 204-213.

Suraweera, P., & Mitrovic, A. (2002). KERMIT: A constraint-based tutor for database modeling. In International Conference on Intelligent Tutoring Systems (pp. 377-387). Springer, Berlin, Heidelberg.

Susi, T., Johannesson, M., & Backlund, P. (2007). Serious games: An overview

Taggart, R. W., Dressler, M., Kumar, P., Khan, S., & Coppola, J. F. (2016). Determining emotions via facial expression analysis software. *Proceedings of Student-Faculty Research Day, CSIS, Pace University*.

Toda, A. M., Valle, P. H., & Isotani, S. (2017). The dark side of gamification: An overview of negative effects of gamification in education. In *Researcher links workshop: higher education for all* (pp. 143-156). Springer, Cham.

Van Eck, R. (2006). Digital game-based learning: It's not just the digital natives who are restless. *EDUCAUSE review, 41*(2), 16.

van Roy, R., Deterding, S., & Zaman, B. (2018). Uses and gratifications of initiating use of gamified learning platforms. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-6).

VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education, 16*(3), 227-265.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197-221.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rosé, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive science, 31*(1), 3-62.

Verbert, K., Govaerts, S., Duval, E., Santos Odriozola, J., Van Assche, F., Parra Chico, G., and Klerkx, J. (2014). Learning dashboards: an overview and future research opportunities. *Personal and Ubiquitous Computing*, 18, 6, 1499- 1514

Waalkens, M., Aleven, V., & Taatgen, N. (2013). Does supporting multiple student strategies lead to greater learning and motivation? Investigating a source of complexity in the architecture of intelligent tutoring systems. *Computers & Education, 60*(1), 159-171.

Weber G. (2012) Adaptive Learning Systems. In: Seel N.M. (eds) *Encyclopedia of the Sciences of Learning*. Springer, Boston, MA.

Weerasinghe, A., Mitrovic, A. Facilitating Deep Learning through Self-Explanation in an Open-ended Domain. *Int. J. of Knowledge-based and Intelligent Engineering Systems* (KES), IOS Press, vol. 10, no 1, 3-19, 2006.

Wetzel, J., VanLehn, K., Butler, D., Chaudhari, P., Desai, A., Feng, J., . . . Patade, V. (2017). The design and development of the dragoon intelligent tutoring system for model construction: lessons learned. *Interactive Learning Environments, 25*(3), 361-381.

Winne, P. H., & Hadwin, A. F. (2013). nStudy: Tracing and supporting self-regulated learning in the Internet. In *International Handbook of Metacognition and Learning Technologies* (pp. 293-308): Springer.

Winne, P.H. & Hadwin, A.F. (2008) The Weave of Motivation and Self-Regulated Learning. In Schunk, D.H., & Zimmerman, B.J. , Motivation and Self-Regulated Learning: Theory, Research, and Application (pp. 297–314). New York, NY: Routledge.

Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: recognising and responding to student affect. *International Journal of Learning Technology, 4*(3-4), 129-164.

Zakharov, K., Mitrovic, A., & Johnston, L. (2008). Towards emotionally-intelligent pedagogical agents. In International conference on intelligent tutoring systems (pp. 19-28). Springer, Berlin, Heidelberg.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In *Handbook of self-regulation* (pp. 13-39): Elsevier.

Zimmerman, B.J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, 25, 3-17.

Appendix A. Pre-test and Post-test

The pre/post-tests administered to students during the studies. (Correct answers arehighlighted).

Pre-Test For Gamification and SRL Studies

1. What clause of the SELECT statement allows tuples to be selected?

    a. SELECT

    b. FROM

    c. WHERE

    d. GROUP BY

2. Which clause needs to be used together with HAVING?

    a. GROUP BY

    b. ORDER BY

    c. COUNT

    d. DISTINCT

3. What does EXISTS in general do?

a. The EXISTS condition is satisfied when the nested query does not return any tuples.

b. Acts like the AND operator.

c. The EXISTS condition is satisfied when the nested query returns at least one tuple.

d. Sorts the nested query result.

4. All attributes listed in the ORDER BY clause of a SELECT statement must also appear in the SELECT clause.

    True False

5. Attribute names used in subqueries are assumed to come from tables used in the outer query.

    True False

6. Which of the following would allow all tuples of table R to be kept in the resulting table, but only those tuples from S that have matching tuples in R?

a. from R join S on R.A=S.B

b. from R right outer join S on R.A=S.B

c. from S left outer join R on R.A=S.B

d. from R left outer join S on R.A=S.B

e. from R full outer join S on R.A=S.B

7. Which clauses can contain a nested query?

A. WHERE

B. GROUP BY

C. ORDER BY

D. WHERE and HAVING

8. Two tables are given:

STUDENT(StudNo, Name, Department)

GRADES(StudNo, *Course*, Grade)

What is the effect of the following statement:

SELECT StudNo, Name

FROM student

WHERE StudNo IN (select StudNo from grades where Course='COSC265');

a. Find students who have failed COSC265

b. Find students who have passed some courses.

c. Find students who have taken COSC265.

d. Find students who have passed COSC265

9. The BOOK table is defined as follows:

BOOK (Book_No, Title, Genre, Year, Price, No_of_Pages)

We need to find the number of books for each genre. The query below is incorrect because:

a. TITLE should be added to the GROUP BY clause

b. The GROUP BY clause is not needed

c. TITLE should be removed from the SELECT clause

d. count(*) should be replaced with count(Book_No)

[Faiza Tahir]

SELECT genre, title, count(*)

FROM book

GROUP BY genre;

Post-Test For Gamification and SRL Studies

1. Which aggregate function can be used to return the number of tuples?

   a. SUM
   b. COUNT
   c. MAX
   d. AVG

2. Which of the following clauses is not allowed in a nested query?

   a. ORDER BY

   b. WHERE
   c. SELECT
   d. GROUP BY
   e. FROM

3. Which of the options below is correct?
   a. DISTINCT is always used with COUNT.
   b. **COUNT can be used without DISTINCT**.
   c. DISTINCT is an attribute type.
   d. DISTINCT can be specified in ORDER BY.

4. NOT IN allows you to specify a condition on an attribute checking that the value of the attribute does not appear in the enumerated set of values.
   True False

5 We need to find the titles of all movies other than comedies. The following SQL statement achieves that.

```
SELECT TITLE
FROM MOVIE
WHERE TYPE = 'comedy' or 'drama';
```

True **False**

6. How do we specify a numeric constant and a string constant?
   a. **Strings between apostrophes (single quotes), and numbers without delimiters**
   b. Numbers between two apostrophes, and strings without delimiters
   c. Number and strings should come between two apostrophes
   d. Number and string should not be enclosed by any symbols

7. Which option is equivalent with this condition?

title IN ('Someone to watch over me','Summertime')

a. title = 'Someone to watch over me'
b. **(title = 'Someone to watch over me' or title= 'Summertime')**
c. (title = 'Someone to watch over me' and title= 'Summertime')
d. (or (title = 'Someone to watch over me', title= 'Summertime'))

8. Two tables are given:
STUDENT(StudNo, Name, Department)
GRADES(StudNo, *Course*, Grade)
What is the effect of the following statement:

```
SELECT StudNo, Course,count(*)
FROM grades
GROUP BY StudNo,Course
HAVING count(*)=2;
```

a. **Show students who have taken some courses twice**.
b. For students who have repeated courses, show courses they have taken more than once.
c. Show courses students have passed.
d. Show courses students have failed.

9. The BOOK table is defined as follows:

BOOK (Book_No, Title, Genre, Year, Price, No_of_Pages)

We need to find the mean number of pages for books of each genre. The query below is incorrect because:

a. TITLE should be added to the GROUP BY clause
b. The GROUP BY clause is not needed
c. TITLE should be added to the GROUP BY and the SELECT clauses
d. sum(no_of_pages) should be replaced with avg(no_of_pages)

```
SELECT genre, sum(no_of_pages)
FROM book
GROUP BY genre;
```

Appendix B. Questionnaire

Questionnaires and survey administered to students during studies.

Questionnaire for Affect detection Study

1. Demographic and Emotion Intensity Questionnaire

UNIVERSITY OF CANTERBURY

Questionnaire

1. What is your age? (*Please circle*)

   18-23 24-29 30-35 36-41 42-47 48+

2. What is your gender? (Please circle)

Female Male Other

3. What is your ethnicity? (Please circle)
   New Zealand/European Maori Asian Pacific Island Other
   If you circled "other", please specify:
4. Have you ever participated in a study using the IAPS before? (Please circle)
   Yes No

5. Have you used SQL-Tutor previously? (Please circle)
   A little Somewhat A lot

Emotion rating scale

1. I am easily **excited**:

Strongly Disagree *Neutral* *Strongly Agree*

1 2 3 4 5 6 7 8 9

2. I am easily **amused**:

Strongly Disagree *Neutral* *Strongly Agree*

1 2 3 4 5 6 7 8 9

3. I am easily **contented**:

Strongly Disagree *Neutral* *Strongly Agree*

[Faiza Tahir]

1 2 3 4 5 6 7 8 9

4. I am easily **saddened**:

Strongly Disagree *Neutral* *Strongly Agree*

1 2 3 4 5 6 7 8 9

5. I am easily **feared**:

Strongly Disagree *Neutral* *Strongly Agree*

1 2 3 4 5 6 7 8 9

6. I am easily **angered**:

Strongly Disagree *Neutral* *Strongly Agree*

1 2 3 4 5 6 7 8 9

7. I am easily **disgusted/bored**:

Strongly Disagree *Neutral* *Strongly Agree*

1 2 3 4 5 6 7 8 9

2. Self-reporting Scale

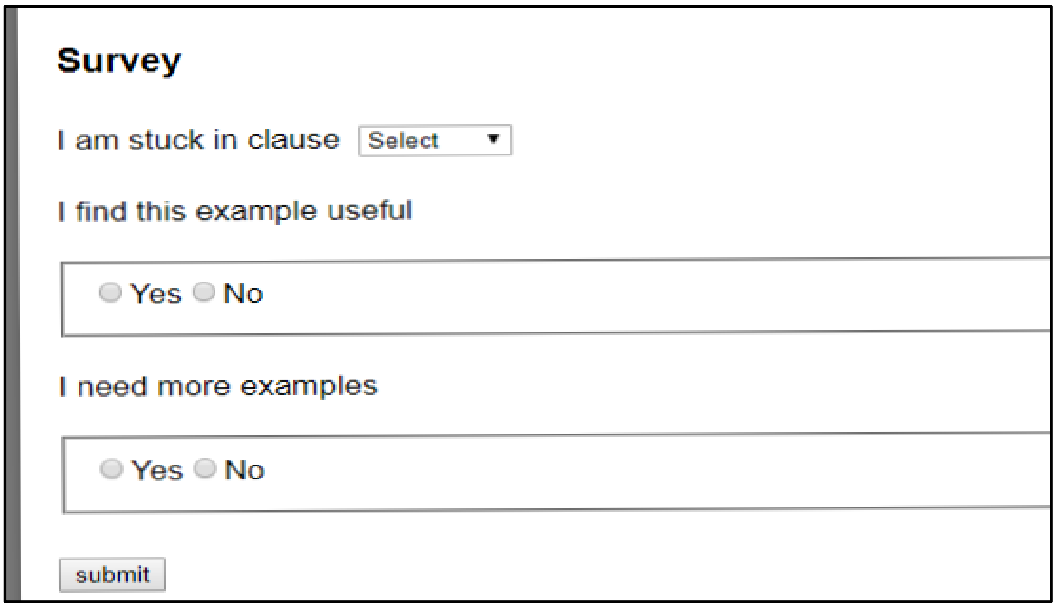
This photo made me

(Please select only one emotion)

	Very Strongle Agree	Strongly Agree	Agree	Slightly Agree	Neutral	Slightly Disagree	Disagree	Strongly Disagree	Very Strongly Disagree
angry	○	○	○	○	○	○	○	○	○
sad	○	○	○	○	○	○	○	○	○
fear	○	○	○	○	○	○	○	○	○
disgust	○	○	○	○	○	○	○	○	○
joy	○	○	○	○	○	○	○	○	○
surprised	○	○	○	○	○	○	○	○	○
contented	○	○	○	○	○	○	○	○	○

3. Example Questionnaire

[Faiza Tahir]



Questionnaire for Gamification Study (correct answers are highlighted)

1. Quiz

1. What clause of the SELECT statement allows conditions to be specified on groups of tuples?

a. SELECT

b. WHERE

c. GROUP BY

d. HAVING

2. What does DISTINCT do in an SQL query ?

   a. Sorts tuples in a specified order
   b. **Eliminates duplicate tuples**
   c. Groups tuples
   d. Eliminates tuples that do not meet a specified condition

3. Which predicate allows to check whether the value of an attribute is a member of the list of pre-specified values?

   a. NOT EXISTS

b. MEMBER
c. EXISTS
d. **IN**

4. The HAVING clause is applied to each group of tuples.
True False

5. We need to find the titles of all movies other than comedies. The following statement will achieve that.

SELECT TITLE
FROM MOVIE
WHERE TYPE = NOT('comedy')

True False

6. Which of the following should be used to fill the blank below to find the mean price?

a. MAX(price)
b. COUNT(price)
c. **AVG(price)**
d. SUM(price)

SELECT ________ FROM BOOK

7. A SELECT statement contains a nested query in the WHERE clause, comparing the value of an attribute to the values returned by the nested SELECT with an IN predicate. Which of the following predicates can be used in that statement instead of IN?

a. **Any**

b. All

c. Every

d. Each

8. Two tables are given:
STUDENT(StudNo, Name, Department)
GRADES(StudNo, *Course*, Grade)

What is the effect of the following statement:

SELECT name
FROM student
WHERE EXISTS (select * from grades
where student.studno=grade.studno AND Course LIKE 'MATH___');

a. Find students who have taken no mathematics courses.

b. Find students who have passed no courses.
c. **Find students who have taken some mathematics courses.**
d. Find students who have passed at least one course.

9. The BOOK table is defined as follows:

BOOK (Book_No, Title, Genre, Year, Price, No_of_Pages)

We need to find the mean price of books for each genre. The query below is incorrect because:

a. TITLE should be added to the GROUP BY clause
b. The GROUP BY clause is not needed
c. **TITLE should be removed from the SELECT clause**
d. PRICE should be added to the GROUP BY clause

```
SELECT genre, title, AVG(PRICE)
FROM book
GROUP BY genre;
```

2. Survey 1 & 2

1. How confident are you in writing SQL queries?
   Not at all True — Very True
   1 2 3 4 5 6 7

2. I feel confident in my ability to learn SQL querying.
   Not at all True — Very True
   1 2 3 4 5 6 7

3. I am capable of learning SQL.
   Not at all True — Very True
   1 2 3 4 5 6 7

4. I feel able to meet the challenge of performing well in SQL querying.
   Not at all True — Very True
   1 2 3 4 5 6 7

5. I think that the problems about SQL queries are very interesting
   Not at all True — Very True
   1 2 3 4 5 6 7

6. It is not important for me to know SQL
   Not at all True — Very True
   1 2 3 4 5 6 7

7. It is easy to stay focused on problems about SQL queries
   Not at all True — Very True
   1 2 3 4 5 6 7

8. I am keen to learn more about the SQL queries
   Not at all True Very True
   1 2 3 4 5 6 7

9. I think that the problems about SQL queries are uninteresting
   Not at all True Very True
   1 2 3 4 5 6 7

10. I think that other database topics are more relevant than SQL
    Not at all True Very True
    1 2 3 4 5 6 7

11. I think that during database class, more attention should be paid to SQL.
    Not at all True Very True
    1 2 3 4 5 6 7

3. Survey 3

Q1 Badges motivated me to participate more than I would have otherwise

Strongly disagree Strongly agree
1 2 3 4 5

Q2 I found being able to earn "badges" increased my enjoyment of using SQL-Tutor.

Strongly disagree Strongly agree
1 2 3 4 5

Q3 I would prefer not to see "badges" in SQL-Tutor".

Strongly disagree Strongly agree
1 2 3 4 5

Q4 The badges awarded for solving problems motivated me to solve more problems than I would have otherwise.

Strongly disagree Strongly agree
1 2 3 4 5

Q5 I found daily challenges useful and exciting.

Strongly disagree Strongly agree
1 2 3 4 5

Q6 I prefer to have daily challenge in SQL-Tutor in future.

Strongly disagree Strongly agree
1 2 3 4 5

Q7 I found quizzes useful and enjoy attempting quizzes.

Strongly disagree Strongly agree
1 2 3 4 5

Q8 I prefer to have quizzes in SQL-Tutor in future.

Strongly disagree Strongly agree
1 2 3 4 5

Questionnaires for Self-Regulated Learning Study

1. Survey 1 & 2

1. I'm certain I can understand the ideas taught in SQL.

Not at all true for me Very true for me

1 2 3 4 5

2. I set personal standards for performance in my learning.

Not at all true for me Very true for me

1 2 3 4 5

3. I ask myself questions about what I am to study before I begin to learn.

Not at all true for me Very true for me

1 2 3 4 5

4. When I am learning, I combine different sources of information (for example: people, web sites, printed material).

Not at all true for me Very true for me

1 2 3 4 5

5. I am sure I can do an excellent job on the problems and tasks assigned for this class.

Not at all true for me Very true for me

1 2 3 4 5

6. I know how well I have learned once I have finished a task.

Not at all true for me Very true for me

1 2 3 4 5

7. I try to identify others whom I can ask for help if necessary.

Not at all true for me Very true for me

1 2 3 4 5

8. I ask others for more information when I need it.

Not at all true for me Very true for me

1 2 3 4 5

9. I try to translate new information into my own words.

Not at all true for me Very true for me

1 2 3 4 5

10. I set short-term (daily or weekly) goals as well as long-term goals (for the whole course).

Not at all true for me Very true for me

1 2 3 4 5

11. I try to apply my previous experience when learning.

Not at all true for me Very true for me

1 2 3 4 5

12. I think I will receive a good grade in this class.

Not at all true for me Very true for me

1 2 3 4 5

13. I read beyond the core SQL materials to improve my understanding.

Not at all true for me Very true for me

1 2 3 4 5

14. When I am learning, I try to relate new information I find to what I already know.

Not at all true for me Very true for me

1 2 3 4 5

15. I expect to do very well in this class.

Not at all true for me Very true for me

1 2 3 4 5

16. When I study SQL, I make notes to help me organise my thoughts.

Not at all true for me Very true for me

1 2 3 4 5

17. I create my own examples to make information more meaningful.

Not at all true for me Very true for me

1 2 3 4 5

18. I know that I will be able to learn the material for this class.

Not at all true for me Very true for me

1 2 3 4 5

19. I change strategies when I do not make progress while learning.

Not at all true for me Very true for me

1 2 3 4 5

20. I ask myself if there were other ways to do things after I finish learning.

Not at all true for me Very true for me

1 2 3 4 5

21. I ask myself how what I am learning is related to what I already know.

Not at all true for me Very true for me

1 2 3 4 5

22. I set realistic deadlines for learning.

Not at all true for me Very true for me

1 2 3 4 5

23. I organise my study time to accomplish my goals to the best of my ability.

Not at all true for me Very true for me

1 2 3 4 5

24. When I do not understand something, I ask others for help.

Not at all true for me Very true for me

1 2 3 4 5

25. When planning my learning, I use and adapt strategies that have worked in the past.

Not at all true for me Very true for me

1 2 3 4 5

26. I think of alternative ways to solve a problem and choose the best one.

Not at all true for me Very true for me

1 2 3 4 5

27. Even if I am having trouble learning, I prefer to do the work on my own.

Not at all true for me Very true for me

1 2 3 4 5

28. I set goals to help me manage studying time for my learning.

Not at all true for me Very true for me

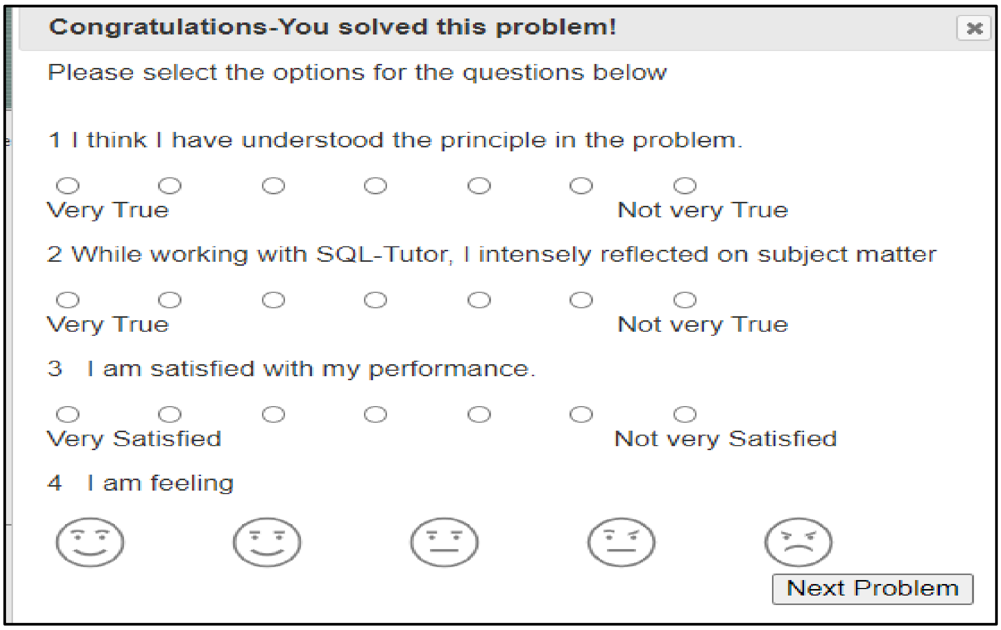
1 2 3 4 5

29. I think about what I have learned after I finish.

Not at all true for me Very true for me

1 2 3 4 5

2. Self-reflection Prompts



Eye Tracking Experiment Questionnaires

1. Demographic and emotional intensity questionnaire

UC UNIVERSITY OF CANTERBURY

Questionnaire

6. What is your age? (*Please circle*)

   18-23 24-29 30-35 36-41 42-47 48+

7. What is your gender? (Please circle)

Female Male Other

8. What is your ethnicity? (Please circle)
   New Zealand/European Maori Asian Pacific Island Other
   If you circled "other", please specify:
9. Have you used SQL-Tutor previously? (Please circle)
   A little Somewhat A lot

Emotion rating scale

8. I am easily **excited**:

Strongly Disagree *Neutral* *Strongly Agree*

1 2 3 4 5 6 7 8 9

9. I am easily **amused**:

Strongly Disagree *Neutral* *Strongly Agree*

1 2 3 4 5 6 7 8 9

10. I am easily **contented**:

Strongly Disagree *Neutral* *Strongly Agree*

1 2 3 4 5 6 7 8 9

11. I am easily **saddened**:

Strongly Disagree *Neutral* *Strongly Agree*

1 2 3 4 5 6 7 8 9

12. I am easily **feared**:

Strongly Disagree *Neutral* *Strongly Agree*

1 2 3 4 5 6 7 8 9

13. I am easily **angered**:

Strongly Disagree *Neutral* *Strongly Agree*

1 2 3 4 5 6 7 8 9

14. I am easily **disgusted/bored**:

Strongly Disagree *Neutral* *Strongly Agree*

1 2 3 4 5 6 7 8 9

2. Dashboard Questionnaire

1	**How well you understand the following elements on dashboard?** Pre-test score Current knowledge level Total time Session time Total Problems completed Session problems completed Complete solution used Highest Complexity problem Goal progress bars Class progress Strategy (Practice problem, challenge me) Suggestion Time/week (graph) Problems/week (graph) Clause wise progress (graph)	**Not at all 1**	**2**	**3**	**4**	**Completely 5**
2	**How useful is each element for learning?** Pre-test score Current knowledge level Total time Session time Total Problems completed Session problems completed	**Not at all 1**	**2**	**3**	**4**	**Very much 5**

	Complete solution used Highest Complexity problem Goal progress bars Class progress Strategy (Practice problem, challenge me) Suggestion Time/week (graph) Problems/week (graph) Clause wise progress (graph)					
3	**How much did each element motivate you to improve your performance?** Pre-test score Current knowledge level Total time Session time Total Problems completed Session problems completed Complete solution used Highest Complexity problem Goal progress bars Class progress Strategy (Practice problem, challenge me) Suggestion Time/week (graph) Problems/week (graph) Clause wise progress (graph)	**Not at all 1**	**2**	**3**	**4**	**Very much 5**
4	**Did the element help you to reflect on your progress?** Pre-test score Current knowledge level Total time Session time Total Problems completed Session problems completed Complete solution used Highest Complexity problem Goal progress bars Class progress Strategy (Practice problem, challenge me) Suggestion Time/week (graph)	**Not at all 1**	**2**	**3**	**4**	**Very much 5**

Problems/week (graph)
Clause wise progress (graph)
Pre-test score

5. Do you think the dashboard supported you in managing your time?

Strongly disagree Strongly agree

1 2 3 4 5

6. Do you think the dashboard supported your problem selection strategy?

Strongly disagree Strongly agree

1 2 3 4 5

7. Does seeing the class progress assist your learning?

Strongly disagree Strongly agree

1 2 3 4 5

8. Do you think there should be more suggestions on your learning strategy on dashboard?

Strongly disagree Strongly agree

1 2 3 4 5

9. Do you think the information on dashboard is constructed logically?

Strongly disagree Strongly agree

1 2 3 4 5

11. Do you think the information is delivered effectively?

Strongly disagree Strongly agree

1 2 3 4 5

12. Do you have any suggestions for improving the dashboard in SQL-Tutor?

Appendix C. Information Sheets

Information sheets for briefing students at the start of the study.

1. Affect Detection Study

Extending Worked Examples in Intelligent Tutoring Systems (ITSs): from Static to Adaptive

Participant Information Sheet

I am a PhD student in the Department of Computer Science and Software Engineering at the University of Canterbury. I am conducting a research project that investigates whether an adaptive strategy for providing examples during problem solving is useful for learning.

I would like to invite you to participate in my study. If you agree to participate, you can come to room 243 (Jack Erskine building) and use your UC user code to log into an Intelligent Tutoring System for problem solving in Structured Query Language (SQL) called SQL-Tutor.

First you will be asked to fill in a short demographic questionnaire. After that you will be asked to sit in front of the Tobii eye tracker, which will capture your eye gaze and your emotions will be recorded by iMotions software. You will then be required to complete eye calibration, where you will be asked to look at few photos on the screen in order for the computer to track your eyes accurately. Then you will be shown a series of photos on screen, arranged in 4 blocks of 12 photos. After each photo you will be asked to record your current emotional state by selecting an answer on screen. You will be asked to select only one type of emotion. After each block you will receive a short 15 second rest. This part of the experiment will last 20 minutes.

In the second part of the session, you will be solving problems in SQL-Tutor. You will need to login to SQL-Tutor with your UC user code. The “Example” button will be available to you on problem solving interface. Whenever you think an example would be useful for you to complete the current problem, you can request to see the example by clicking on the Example button. At the end of the example, you will be asked to give feedback about the usefulness of the example. The second part of the session will take another 20 minutes. Please note that your eye gaze and your facial expressions will be recorded during the session.

Participation is voluntary and you have the right to withdraw at any stage without any penalty. As a participant in this study, you will receive a $20 reboot café voucher for participating in the study.

The results of the project may be published, but you may be assured of the complete confidentiality of data gathered in this investigation: your identity will not be made public without your prior consent. To ensure anonymity and confidentiality, you will be assigned a random id number by which each participant will anonymously be identified. Faiza Tahir will have access to the data as well as Prof. Tanja Mitrovic; the data will be stored on a secured machine that requires authenticated password to access. The data will be destroyed after ten years. A thesis is a public document and will be available through the UC Library

My PhD project is supervised by Prof. Tanja Mitrovic (Tanja.mitrovic@canterbury.ac.nz) and we will be pleased to discuss any concerns you may have about participation in the project.

[Faiza Tahir]

This project has been reviewed and approved by the University of Canterbury Human Ethics Committee, and participants should address any complaints to The Chair, Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz).

If you agree to participate in the study, you are asked to complete the consent form and return to the Intelligent Computer Tutoring Group (ICTG) lab, Department of Computer Science and Software Engineering, University of Canterbury.

Faiza Tahir

faiza.tahir@pg.canterbury.ac.nz

2. Gamification Study

Information Sheet

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Department: Computer Science and Software Engineering

Telephone: +64 3369 2777

Email: faiza.tahir@pg.canterbury.ac.nz 17-June-2019

Extending Worked Examples in Intelligent Tutoring Systems (ITSs): from Static to Adaptive

Information Sheet for students

I am a PhD student in the Department of Computer Science and Software Engineering at the University of Canterbury. I am conducting a research project that investigates whether an adaptive strategy for providing examples during problem solving is useful for learning.

I would like to invite you to participate in my study. If you agree to participate, you can use your UC user code to log into SQL-Tutor, which is an Intelligent Tutoring System in which you can practice SQL queries.

First you will be asked to complete a short online pre-test and then you can solve problems with SQL-Tutor as much as you want over the period of four weeks. Please note that SQL-Tutor will record information about your sessions, such as selection of problems, time to solve problems, total attempts on problems, and hints used. During problem solving, you will be able to ask for an example similar to the current problem, using the "Example" button from the system's interface. At the end of example, you will be asked to give feedback about the usefulness of the example.

Participation is voluntary and you have the right to withdraw at any stage without any penalty, in which case all data from your sessions with SQL-Tutor will be removed from the study. At the end of the study, there will be a lucky draw including all students who completed the study. The prizes are two vouchers worth $100 each.

[Faiza Tahir]

The results of the project may be published, but you may be assured of the complete confidentiality of data gathered in this investigation: your identity will not be made public. To ensure anonymity and confidentiality, all data collected from your interactions with SQL-Tutor will be stored using a random Id number. Faiza Tahir will have access to the data as well as Prof. Tanja Mitrovic; the data will be stored on a secured machine that requires authenticated password to access. The data will be destroyed after ten years. A thesis is a public document and will be available through the UC Library

My PhD project is supervised by Prof. Tanja Mitrovic (tanja.mitrovic@canterbury.ac.nz) and we will be pleased to discuss any concerns you may have about participation in the project.

This project has been reviewed and approved by the University of Canterbury Human Ethics Committee, and participants should address any complaints to The Chair, Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz).

If you agree to participate in the study, you are asked to complete the consent form and return to the Intelligent Computer Tutoring Group (ICTG) lab, Department of Computer Science and Software Engineering, University of Canterbury.

Faiza Tahir

faiza.tahir@pg.canterbury.ac.nz

3. Self-regulated Learning Study

UC
UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Department: Computer Science and Software Engineering

Telephone: +64 3369 2777

Email: faiza.tahir@pg.canterbury.ac.nz

20-July-2020

Investigating the effects of self-regulation support in SQL-Tutor

Information Sheet for students

I, Faiza Tahir, am a PhD student in the Department of Computer Science and Software Engineering at the University of Canterbury. I am conducting a research project on providing support during interaction with SQL-Tutor. I would like to invite you to participate in my study. If you agree to participate, you can use your UC user code to log into SQL-Tutor, which is an Intelligent Tutoring System in which you can practice SQL queries.

First you will be asked to complete a short online pre-test and a questionnaire. Next, you will select a goal for every session. You will be able to select problems to solve. You will be given prompts to report your affective (emotional) state after solving problem and you can monitor your progress anytime during learning. You can solve problems with SQL-Tutor as much as you want over the period of four weeks. Please note that SQL-Tutor will record information about your sessions, such as selection of problems, time to solve problems, total attempts on problems, and hints used. At the end of the study you will be asked to complete a post-test and a questionnaire.

Participation is voluntary and you have the right to withdraw at any stage without any penalty. If you do choose to withdraw, please email the researcher at the contact details listed above and all data gathered from you will be destroyed. The outcomes of this study (and the degree to which participants engage with SQL-Tutor) will not have any influence on your grades for COSC265 or any other course.

At the end of the study, there will be a lucky draw which include those students who opted-in the study. The prizes are two vouchers worth $100 each.

The results of the project may be published, but you may be assured of the complete confidentiality of data gathered in this investigation: your identity will not be made public. To ensure anonymity and confidentiality, all data collected from your interactions with SQL-Tutor will be stored using a random Id number. Faiza Tahir will have access to the data as well as Prof. Tanja Mitrovic; the data will be stored on a secured machine that requires authenticated password to access. The data will be destroyed after ten years. A thesis is a public document and will be available through the UC Library

My PhD project is supervised by Prof. Tanja Mitrovic (tanja.mitrovic@canterbury.ac.nz) and Dr Valerie Sotardie (valerie.sotardi@canterbury.ac.nz). We will be pleased to discuss any concerns you may have about participation in the project. This project has been reviewed and approved by the University of Canterbury Human Ethics Committee, and participants should address any complaints to The Chair, Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz).

If you agree to participate in the study, you are asked to complete the consent form and return to the Department of Computer Science and Software Engineering, University of Canterbury.

Faiza Tahir, faiza.tahir@pg.canterbury.ac.nz

4. Eye tracking and Affect Detection Experiment

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Department: Computer Science and Software Engineering

Telephone: +64 3369 2777

Email: faiza.tahir@pg.canterbury.ac.nz

25-Sep-2020

Investigating the effects of self-regulation support in SQL-Tutor

Information Sheet for students

[Faiza Tahir]

I, Faiza Tahir, am a PhD student in the Department of Computer Science and Software Engineering at the University of Canterbury. I am conducting a research project on providing support during interaction with SQL-Tutor. I would like to invite you to participate in my study. If you agree to participate, you can use your UC user code to log into SQL-Tutor, which is an Intelligent Tutoring System in which you can practice SQL queries.

First you will be asked to fill in a short demographic and emotion rating questionnaire. After that you will be asked to sit in front of the Tobii eye tracker, which will capture your eye gaze and your emotions will be recorded by iMotions software. You will then be required to complete eye calibration. Once the calibration is successful you will need to login to SQL-Tutor with your UC user code and start solving problems. The session will take up to 40 minutes. Please note that your eye gaze and facial expressions will be recorded during the session by iMotions software.

Participation is voluntary and you have the right to withdraw at any stage without any penalty. The outcomes of this will not have any influence on your grades for COSC265 or any other course.

As a participant in this study, you will receive a $20 voucher for CAFE101 for participating in the study.

The results of the project may be published, but you may be assured of the complete confidentiality of data gathered in this investigation: your identity will not be made public. To ensure anonymity and confidentiality, all data collected from your interactions with SQL-Tutor will be stored using a random Id number. Faiza Tahir will have access to the data as well as Prof. Tanja Mitrovic; the data will be stored on a secured machine that requires authenticated password to access. The data will be destroyed after ten years. A thesis is a public document and will be available through the UC Library

My PhD project is supervised by Prof. Tanja Mitrovic (tanja.mitrovic@canterbury.ac.nz) and Dr Valerie Sotardi (valerie.sotardi@canterbury.ac.nz). They will be pleased to discuss any concerns you may have about participation in the project. This project has been reviewed and approved by the University of Canterbury Human Ethics Committee, and participants should address any complaints to The Chair, Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz).

If you agree to participate in the study, you are asked to complete the consent form and return to the Department of Computer Science and Software Engineering, University of Canterbury.

Thank you

Faiza Tahir

Appendix D. Consent forms

1. Affect Detection Study

Extending Worked Examples in ITSs: from Static to Adaptive

Consent Form

I have been given a full explanation of this project and have had the opportunity to ask questions.

I understand what is required of me if I agree to take part in the research.

I understand that participation is voluntary and I may withdraw at any time without penalty. Withdrawal of participation will also include the withdrawal of any information I have provided should this remain practically achievable.

I understand that any information or opinions I provide will be kept confidential to the researchers involved in this study (Faiza Tahir and her supervisors) and that any published or reported results will not identify the participants. I understand that a thesis is a public document and will be available through the UC Libraries.

I understand that all data collected for the study will be kept on a password-protected UC computer within the ICTG lab (Jack Erskine building) and will be destroyed after 10 years.

I understand the study is considered low risk because it does not raise any issues of deception, threat, invasion of privacy, mental, physical or cultural.

I understand that I am able to receive a summary of the findings of the study by providing the contact details at the end of this form.

I understand that I can contact the researcher *Faiza Tahir (faiza.tahir@pg.canterbury.ac.nz)* or her supervisor Professor Tanja Mitrovic *(tanja.mitrovic@canterbury.ac.nz)* for further information. If I have any complaints, I can contact the Chair of the University of Canterbury Human Ethics Committee, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz).

I hereby consent to take part in this study.

Participant's Name:

Signature: **Date:**

☑ I would like to receive a summary of the findings of the study through my E-Mail address:

E-Mail Address:

2. Gamification Study

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Department: Computer Science and Software Engineering

Telephone: +64 3369 2777

Email: faiza.tahir@pg.canterbury.ac.nz

Extending Worked Examples in Intelligent Tutoring Systems (ITSs): from Static to Adaptive

Consent Form for Students

I have been given a full explanation of this project and have had the opportunity to ask questions.

I understand what is required of me if I agree to take part in the research.

I understand that participation is voluntary and I may withdraw at any time without penalty. Withdrawal of participation will also include the withdrawal of any information I have provided should this remain practically achievable.

I understand that any information or opinions I provide will be kept confidential to the researchers involved in this study (Faiza Tahir and her supervisors), and that any published or reported results will not identify the participants. I understand that results of this study may be published in PHD thesis of researcher and will be available through the UC Libraries.

I understand that all data collected for the study will be kept on a password-protected UC computer within the Intelligent Computer Tutoring Group (ICTG) lab, Department of Computer Science and Software Engineering (Jack Erskine building) and will be destroyed after 10 years.

I understand the study is considered low risk because it does not raise any issues of deception, threat, invasion of privacy, mental, physical or cultural.

I understand that I am able to receive a summary of the findings of the study by providing the contact details at the end of this form.

I understand that I can contact the researcher *Faiza Tahir (faiza.tahir@pg.canterbury.ac.nz)* or her supervisor Professor Tanja Mitrovic *(tanja.mitrovic@canterbury.ac.nz)* for further information. If I have any complaints, I can contact the Chair of the University of Canterbury Human Ethics Committee, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz).

I hereby consent to take part in this study.

Participant's Name:

Signature: **Date:**

☑ I would like to receive a summary of the findings of the study through my E-Mail address:

E-Mail Address:

[Faiza Tahir]

3. Self-Regulated Learning Study

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Department: Computer Science and Software Engineering

Telephone: +64 3369 2777

Email: faiza.tahir@pg.canterbury.ac.nz

Investigating the effects of self-regulation support in SQL-Tutor

Consent Form for students

Include a statement regarding each of the following:

☐ I have been given a full explanation of this project and have had the opportunity to ask questions.

☐ I understand what is required of me if I agree to take part in the research.

☐ I understand that participation is voluntary and I may withdraw at any time without penalty. Withdrawal of participation will also include the withdrawal of any information I have provided should this remain practically achievable.

☐ I understand that any information or opinions I provide will be kept confidential to the researchers involved in this study (Faiza Tahir and her supervisors), and that any published or reported results will not identify the participants. I understand that results of this study may be published in the PhD thesis of the researcher and will be available through the UC Libraries.

☐ I understand that all data collected for the study will be kept on a password-protected UC computer within the Intelligent Computer Tutoring Group (ICTG) lab, Department of Computer Science and Software Engineering (Jack Erskine building) and will be destroyed after 10 years.

☐ I understand the study is considered low risk because it does not raise any issues of deception, threat, invasion of privacy, mental, physical or cultural.

☐ I understand that I am able to receive a summary of the findings of the study by providing the contact details at the end of this form.

☐ I understand that I can contact the researcher Faiza Tahir (faiza.tahir@pg.canterbury.ac.nz) or her supervisor

Professor Tanja Mitrovic (tanja.mitrovic@canterbury.ac.nz) for further information. If I have any complaints, I can contact the Chair of the University of Canterbury Human Ethics Committee, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz).

☐ I would like to receive the summary of findings of the study through my E-Mail address.

☐ By signing below, I agree to participate in this research project.

Name:______________________________Signed:____________________________Date:________________

Email address (for report of findings):____

4. Eye tracking and Affect Detection Experiment

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Department: Computer Science and Software Engineering

Telephone: +64 3369 2777

Email: faiza.tahir@pg.canterbury.ac.nz

Investigating the effects of self-regulation support in SQL-Tutor

Consent Form for students

Include a statement regarding each of the following:

☐ I have been given a full explanation of this project and have had the opportunity to ask questions.

☐ I understand what is required of me if I agree to take part in the research.

☐ I understand that participation is voluntary and I may withdraw at any time without penalty. Withdrawal of participation will also include the withdrawal of any information I have provided should this remain practically achievable.

- ☐ I understand that any information or opinions I provide will be kept confidential to the researchers involved in this study (Faiza Tahir and her supervisors), and that any published or reported results will not identify the participants. I understand that results of this study may be published in the PhD thesis of the researcher and will be available through the UC Libraries.

- ☐ I understand that all data collected for the study will be kept on a password-protected UC computer within the Intelligent Computer Tutoring Group (ICTG) lab, Department of Computer Science and Software Engineering (Jack Erskine building) and will be destroyed after 10 years.

- ☐ I understand the study is considered low risk because it does not raise any issues of deception, threat, invasion of privacy, mental, physical or cultural.

- ☐ I understand that I am able to receive a summary of the findings of the study by providing the contact details at the end of this form.

- ☐ I understand that I can contact the researcher Faiza Tahir (faiza.tahir@pg.canterbury.ac.nz) or her supervisor Professor Tanja Mitrovic (tanja.mitrovic@canterbury.ac.nz) for further information. If I have any complaints, I can contact the Chair of the University of Canterbury Human Ethics Committee, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz).

- ☐ I would like to receive the summary of findings of the study through my E-Mail address.

- ☐ By signing below, I agree to participate in this research project.

Name:______________________________Signed:____________________________Date:________________

Email address (for report of findings):____

[Faiza Tahir]

Appendix E. Human Ethics Approval Letters

1. Affect Detection Study

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

HUMAN ETHICS COMMITTEE

Secretary, Rebecca Robinson
Telephone: +64 03 369 4588, Extn 94588
Email: human-ethics@canterbury.ac.nz

Ref: HEC 2020/86

21 August 2020

Faiza Tahir
Computer Science and Software Engineering
UNIVERSITY OF CANTERBURY

Dear Faiza

The Human Ethics Committee advises that your research proposal "Investigating the Effects of Self-Regulation Support in an Intelligent Tutoring System" has been considered and approved.

Please note that this approval is subject to the incorporation of the amendments you have provided in your email of 18th August 2020.

Best wishes for your project.

Yours sincerely

Dr Dean Sutherland
Chair
University of Canterbury Human Ethics Committee

2. Gamification Study

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

HUMAN ETHICS COMMITTEE

Secretary, Rebecca Robinson
Telephone: +64 03 369 4588, Extn 94588
Email: human-ethics@canterbury.ac.nz

Ref: HEC 2019/24/LR-PS

31 July 2019

Faiza Tahir
Computer Science and Software Engineering
UNIVERSITY OF CANTERBURY

Dear Faiza

Thank you for submitting your low risk application to the Human Ethics Committee for the research proposal titled "Extending Worked Examples in Intelligent Tutoring Systems (ITSs): from Static to Adaptive".

I am pleased to advise that this application has been reviewed and approved.

Please note that this approval is subject to the incorporation of the amendments you have provided in your email of 24th July 2019.

With best wishes for your project.

Yours sincerely

Dr Dean Sutherland
Chair, Human Ethics Committee

3. Self-Regulated Learning Study

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

HUMAN ETHICS COMMITTEE

Secretary, Rebecca Robinson
Telephone: +64 03 369 4588, Extn 94588
Email: human-ethics@canterbury.ac.nz

Ref: HEC 2020/86

21 August 2020

Faiza Tahir
Computer Science and Software Engineering
UNIVERSITY OF CANTERBURY

Dear Faiza

The Human Ethics Committee advises that your research proposal "Investigating the Effects of Self-Regulation Support in an Intelligent Tutoring System" has been considered and approved.

Please note that this approval is subject to the incorporation of the amendments you have provided in your email of 18th August 2020.

Best wishes for your project.

Yours sincerely

Dr Dean Sutherland
Chair
University of Canterbury Human Ethics Committee

4. Eye Tracking and Affect detection Experiment

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

HUMAN ETHICS COMMITTEE

Secretary, Rebecca Robinson
Telephone: +64 03 369 4588, Extn 94588
Email: human-ethics@canterbury.ac.nz

Ref: HEC 2020/86 Amendment 1

1 October 2020

Faiza Tahir
Computer Science and Software Engineering
UNIVERSITY OF CANTERBURY

Dear Faiza

Thank you for your request for an amendment to your research proposal "Investigating the Effects of Self-Regulation Support in an Intelligent Tutoring System" as outlined in your emails dated 22nd and 25th September 2020.

I am pleased to advise that this request has been considered and approved by the Human Ethics Committee.

Yours sincerely

Dr Dean Sutherland
Chair, Human Ethics Committee

Appendix F. List of Publications

Journal Articles

1. Tahir, F., Mitrovic, A., & Sotardi, V. Investigating the Causal Relationships between Badges and Learning Outcomes in SQL-Tutor. (Under revision, not included in the thesis)

Conference Proceedings

2. Tahir, F., Mitrovic, A., & Sotardi, V. (2021). Investigating effects of selecting challenging goals. International Conference on Artificial Intelligence in Education AIED, 2021

3. Tahir, F., Mitrovic, A., & Sotardi, V. (2021). Do gender, experience and prior knowledge matters when learning with gamified ITS. International Conference of Learning analytics ICALT 2021.

4. Tahir, F., Mitrovic, A., & Sotardi, V. (2020). Investigating the Effects of Gamifying SQL-Tutor. International Conference on Computers in Education ICCE 2020.

5. Tahir, F., Mitrovic, A., & Sotardi, V. (2019). Towards adaptive provision of examples during problem solving. Investigating the Effects of Gamifying SQL-Tutor. International Conference on Computers in Education ICCE 2019.

Posters and Extended summaries

6. Tahir, F., Mitrovic, A., & Sotardi, V. (2021). Visual Attention Patterns on Dashboard during Learning with SQL-Tutor. Companion Proceedings 10th International Conference on learning analytics & Knowledge (LAK21).

7. Tahir, F., Mitrovic, A., & Sotardi, V. (2019). Using Gamification to Effect Learning Behaviors in Intelligent Tutoring System (2019). Proceedings of the 27th International Conference on Computers in Education. Taiwan: Asia-Pacific Society for Computers in Education (775-778).

Towards Adaptive Provision of Examples During Problem Solving

Faiza TAHIR[a*], Antonija MITROVIC[a] & Valerie SOTARDI[b]
[a]*Intelligent Computer Tutoring Group, Department of Computer Science and Software Engineering, University of Canterbury, New Zealand*
[b]*School of Educational Studies and Leadership, University of Canterbury, New Zealand*
*faiza.tahir@pg.canterbury.ac.nz

Abstract: Intelligent Tutoring Systems (ITSs) are effective in supporting learning, as shown in numerous studies. The goal of our project is to develop an adaptive strategy that would be capable of identifying situations during problem solving in which the student would benefit from worked examples. As a first step towards developing such a strategy, we conducted a pilot study in the context of SQL-Tutor, a mature ITS that teaches database querying. The participant could ask for a worked example whenever he/she wanted during problems solving. After each example, the participant specified whether the example was useful, and whether additional examples were needed. Participants' facial expressions and eye gaze were recorded. The findings show that the participants generally found examples useful, although in some cases they stated additional examples would be beneficial. The analysis of the eye gaze shows that students compared provided examples to their own solutions. Affect analysis shows that negative emotions reduced while engagement increased when participants viewed examples, and immediately after examples.

Keywords: intelligent tutoring system, problem solving, worked examples, eye tracking, affective modeling

1. Introduction

Intelligent Tutoring Systems (ITSs) have been proven to be very effective in supporting learning (Kulik & Fletcher, 2016; Mitrovic, 2012; VanLehn, 2011). The main activity in ITSs is problem solving, where the student receives help from the ITS adaptively, based on his/her actions and knowledge. On the other hand, there is a long tradition of research on learning from worked examples (WEs), starting from 1950s (Atkinson et al., 2000). A worked example contains the problem statement and a step-by-step solution with accompanying explanations. Atkinson et al. (2000) suggested the importance of worked examples in early stages of skill acquisition. Learning can also be increased when WEs are combined with self-explanation (Große & Renkl, 2007), problem solving (Cooper & Sweller, 1987), faded examples, (Renkl & Atkinson, 2003), or erroneous examples (Große & Renkl, 2007).

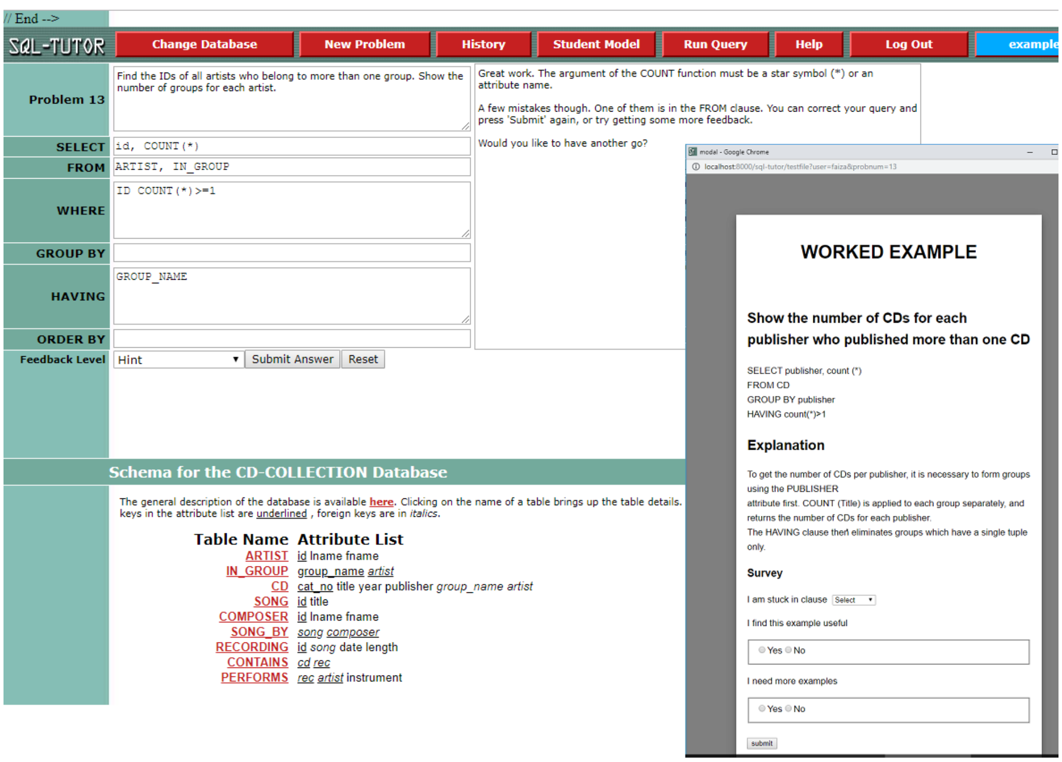
Examples have also been found beneficial when incorporated into ITSs. ELM-PE is one of the first ITSs to incorporate examples and their explanations (Burow & Weber, 1996). SE-Coach (Conati, Larkin, & VanLehn, 1997) guided students to self-explain examples; on the basis of student explanations and student model, it estimated the student understanding of a particular example. EA-Coach (Muldner & Conati, 2007) provided examples adaptively, based on learners' characteristics and example utility. Another study revealed the positive effects of providing WEs adaptively by fading their steps in a cognitive tutor (Salden et al., 2009). A study with SQL-Tutor compared learning from problems only, WEs only, or alternatively provided examples and problems to learners, found that a mixture of WE and problem solving resulted in best learning outcomes (Najar & Mitrovic, 2014). In follow-up studies, Najar and colleagues (2015) showed that adaptive selection of learning activities resulted in highest learning gains. Later on, erroneous examples were introduced in SQL-Tutor and proved to be helpful for advanced learners (Chen, Mitrovic & Mathews, 2019). Another study used a concept-based similarity approach to select most similar examples for the learner, when the learner fails to complete a Java program (Hosseini & Brusilovsky, 2017).

Most of these studies focused either on adaptive strategies for presenting WEs and/or problems, or on adaptive provision of example steps. However, there is a lack of research on adaptive strategies for providing examples to students when they need help during problems solving. In order to fill this gap, we designed and conducted a pilot study with SQL-Tutor (Mitrovic, 2003), the goal of which was to observe when and how students use worked examples during problem solving.

We start by presenting the worked example version of SQL-Tutor used in the pilot study, and then describe the procedure in Section 3. Section 4 presents the findings, while Section 5 presents conclusions.

2. Experimental Setup

The version of SQL-Tutor used in the pilot study contained ten problems. The screenshot in Figure 1 shows the problem-solving environment of SQL-Tutor. At the top of the page, there are several buttons allowing the student to change the database, select a problem, look at the history of the session or his/her student model, run the query, ask for help or exit the system. For each problem, there was one WE that was isomorphic to the problem, using the same database and same domain principles. Figure 1 shows a WE, which includes the problem statement, the solution accompanied with an explanation. After the explanation, the student was required to specify which clause of the Select statement he/she had difficulty with, and then to specify whether the example was useful, and whether additional examples were needed. The three questions were mandatory. In the study, we used the Tobii eye tracker to record the participant's eye gaze, and iMotions to record facial expressions.



3. Procedure

We recruited seven undergraduate and three postgraduate students (two females, eight males), who were all studying Computer Science. Six participants were domestic students, while the remaining four were international (three Asian, and one Latin American). Five participants were aged 18-23, three 24-29 and two 30-35. All participants were familiar with SQL, and some of them have worked with SQL-Tutor before the study. Each student had an individual session (40 minutes long), and was awarded a \$20 voucher for their participation.

At the beginning of the study, the participants provided informed consent, and filled a short questionnaire, in order to collect basic information about participants and their level of familiarity with SQL-Tutor. The participant sat in front of the Tobii screen, and the standard Tobii calibration was completed. The calibration test took 6 seconds, and the experiment started only if results were excellent. Otherwise, the position of participant was readjusted and recalibration took place. The experimenter sat to the other side, and monitored the participant's face and eye gaze captured by both iMotions and Tobii. This monitoring ensured that during experiment full face of participant was captured so iMotions could record the facial features properly. The participants were instructed to solve at least five problems in SQL-Tutor, and to ask for examples as needed.

iMotions recorded participants' facial expressions, which needed to be post-processed first, and later converted into action units and emotions by using the Affectiva AFFDEX facial expression analysis engine. After post processing, only those recordings with the AFFDEX sampling rate quality higher than 80% were included in the analyses (i.e. in 80% of samples it was possible to identify facial features). Affectiva AFFDEX generated probabilistic estimates for the seven emotions (anger, disgust, surprise, sadness, joy, fear and contempt) based on macro-expressions (lasting 0.5-4 seconds) of each participant. We selected the amplitude-based thresholding technique to focus on the strongest emotion.

4. How Much have Participants used Examples?

Five participants have not used SQL-Tutor prior to the study. Two participants used SQL-Tutor a lot, while the remaining three had limited experience with the system. Table 1 shows how many participants attempted and completed each problem, asked for examples, and how much time was spent on average on the problem/example. The *Example* column specifies the number of participants who viewed examples. The participants mostly solved the problems in the provided order, from the easiest to the hardest. On average, participants attempted 6 problems (sd = 1.89). The four most difficult problems were attempted much less often, and no one completed problems 9 and 10. The participants completed 62% of the problems they attempted, and viewed examples in 59% of the cases. . For problems 1-5, as the problem complexity grows, the example use increases. In problems 2, 4, 5 and 10, participants viewed the examples more than once. When they viewed examples for the first time, they spent on average a minute viewing them. Upon the second and third viewing, this time decreased to 10-20 seconds only. The average time per example is proportional to the average time on problem.

Table 1

Problem, Example and Feedback Use; Time in minutes for problems, and in seconds for examples

Problem	Attempted by	Completed by	Time/ problem	Example	Time/ example	Feedback
1	9	9	1.48 (0.95)	2	40 (7.07)	1
2	8	7	4.18 (2.02)	7	41 (9.72)	3
3	9	7	1.6 (1.34)	1	25 (0)	2
4	10	6	5.58 (3.18)	7	72 (30)	4
5	9	4	6.5 (3.79)	6	46 (54)	5
6	5	3	3.46 (2.62)	3	38 (7.63)	2
7	2	1	2.1 (0.14)	2	16.5 (12)	0
8	3	1	2.05 (0.07)	2	26 (20)	1
9	3	0	2.2 (0.52)	3	30 (16)	0
10	3	0	7.3 (5.23)	3	43 (15.2)	0

The *Feedback* column of Table 1 shows the number of participants who have explicitly required specific levels of feedback (such as hint, partial/complete solution) while solving problems. More participants have used feedback for the easier problems (1-5) than for the rest of problems. This trend is opposite to how participants used examples.

Table 2

Participants' Opinions on Examples

Example	Viewed by	Useful	More examples
1	2	1	2

2	7	7	1
3	1	1	1
4	7	5	2
5	6	4	2
6	3	2	0
7	2	1	1
8	2	2	0
9	3	3	0
10	3	2	1

Table 2 shows participants' responses to the three questions given with WEs. In 78% of the cases participants found them useful, and in 36% of these cases, they wanted more WEs. For complex problems (problems 7-10), when completion rate was low (below 20%), the participants found examples very useful (80% of the cases), even when they have not completed those problems. This shows that regardless of success in problem solving, the participants found the examples useful. Please note that our study was voluntary, and therefore there was no need for students to complete all problems.

5. Eye Gaze Analysis

We analyzed the eye tracking data to determine how the participants read worked examples. Such analysis allows us to understand whether the participants use WEs appropriately. Each WE is isomorphic to the problem, and we expected students to compare the solution provided in the WE to their solution. The area of interest (AOI) was defined to cover the whole example (i.e. title, solution and explanation). The metrics included in eye tracking analysis are: (1) *Time in AOI*, i.e. the total time spent looking at the AOI; (2) *Visits*, i.e. the number of times the participant's eye gaze returns to the AOI; (3) *Fixation count*, showing the total number of fixations within the AOI; (4) Duration of the first fixation on the AOI; and (5) the average fixation duration in AOI. Table 3 shows the metrics for the ten examples (including multiple viewings) averaged over all participants who viewed those examples.

The *Time in AOI* column provides the average time spent by participants while examining a WE. The average number of visits to the AOI seems to increase as problems become more complex. As the number of example steps grows in later examples, the participants looked more often towards the problem solving area and schema.

Table 3

Averages (Standard Deviations) for Eye Tracking Metrics. Times are Reported in Seconds

Example	Time in AOI	Visits	Fixation count	First fixation duration (s)	Fixation duration
1	40 (2.8)	13 (1.4)	43 (6.3)	.2 (.07)	.19 (0)
2	33 (18.5)	9 (9.06)	38 (30)	.14 (.05)	.21 (.030)
3	24 (0)	5 (0)	7 (0)	.24 (0)	.23 (0)
4	59 (30)	27 (25)	132 (127)	.25 (.073)	.24 (.03)
5	59 (61.9)	18 (18.5)	111 (116.9)	.2 (.094)	.24 (.05)
6	51 (20.5)	23 (2.12)	82 (21)	.23 (.063)	.2 (.04)
7	23 (0)	13 (0)	51 (0)	.22 (0)	.25 (0)
8	49 (24.9)	32 (14)	107 (94)	.46 (.37)	.24 (.05)
9	34 (4.7)	10 (6.5)	55 (32)	.19 (.053)	.22 (.04)
10	47 (16.3)	20 (9.5)	107 (52)	.31 (.31)	.22 (.02)

The average fixation count shown in Table 3 is highest in examples 4, 5, 8 and 10. The highest fixation count shows that these participants did not just glance over those examples, but studied them thoroughly, not only the first time but also for the second or third viewing. The high fixation count and average duration of the first fixation on more complicated examples strengthens the above findings that as the number of example steps grows, more fixations were recorded.

6. Affect Analysis

Affectiva AFFDEX analyzes facial expressions and reports the values of seven emotions: anger, sadness, surprise, disgust, joy, contempt and fear, based on Ekman's (1999) categorization of emotions. However, these are general emotions, not the emotions specific to learning (Baker et al., 2010; Craig et al., 2004). Woolf and colleagues (2009) suggested mappings between Ekman's basic emotions to learning-related emotions: joy mapped to excitement, anger mapped to frustration, surprise mapped to boredom, and fear mapped to anxiety.

In line with the above mentioned research, we considered anger, joy, fear and surprise. We additionally included engagement, which is also crucial for learning (Craig et al., 2004; D'Mello, Picard, & Graesser, 2007). We observed some general trends. At the start of each problem, the dominant emotion was surprise, and once the problem was solved, the dominant emotion was joy. When participants received feedback from SQL-Tutor (upon submitting their solutions), the level of surprise was higher. In those situations when participants were able to solve the problem after receiving feedback, again the level of joy was increased. However, if they were not able to solve the problem, we noticed higher levels of anger, showing the participants' frustration.

Another event of interest is when students asked for examples. We analyzed the emotions for three different time intervals: (1) one minute before example use, (2) during example use, and (3) one minute after example use. Firstly, during one minute before participants asked for examples, the dominant emotions were anger and surprise, which seem to suggest that participants asked for examples when they were frustrated. Engagement increased and surprise decreased during or after working with examples. Fear was the least detected emotion; it decreased while and after working with examples and increased slightly when they were working on examples. Joy increased when they were working with examples, and immediately after, when the participants were able to complete problems after viewing examples. On the other hand, if the example did not help the participant solve the problem, we observed increased values for anger and surprise. In some of those cases, the participants asked for the example for the second time, and after that abandoned the problems. This is consistent with findings reported in the literature showing that frustration may lead to boredom, in which case learners loose interest in learning activities.

In summary, we found that participants asked for examples when the levels of anger (i.e. frustration) and surprise (i.e. anxiety) were elevated. Working with examples reduced such negative emotions and increased joy. After viewing examples, when participants turned again to problem solving, the intensity of negative emotions was low, but gradually increased if they were unable to solve the problem. The level of engagement increased for all participants during and after viewing examples. Therefore, examples have positive impact on participants' affective states, which will be helpful in learning with SQL-Tutor.

Conclusions

This paper presented the pilot study the goal of which was to analyze how participants use, study and feel about worked examples in their problem solving journey with SQL-Tutor. The results show that participants used examples extensively, particularly when the complexity of problems increased. Most participants agreed on the usefulness of examples and a few required more examples. This indicates the demand for examples during problem solving, regardless of success in problem solving. The eye gaze analysis revealed that participants tried to understand example structure by comparing examples with their solutions. Lastly, the positive impact of examples on participants' emotions is as examples reduced participant's negative emotions, and increased engagement and up to some extent joy.

The presented findings illustrate the need for and effectiveness of WEs, supported by cognitive and affective states of participants. These findings provide a starting point for developing an adaptive strategy for providing WEs adaptively, during problem solving. A limitation of our study is the small sample size. We plan to collect more data about how and when students use example in the forthcoming study in a large database course, which will enable us to develop and evaluate the adaptive strategy in follow-up studies.

Acknowledgements

We thank Jay Holland for helping with the study, as well as our participants.

References

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from Examples: Instructional Principles from the Worked Examples Research. *Review of educational research, 70*(2), 181-214.

Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *Human-Computer Studies, 68*(4), 223-241.

Burow, R., & Weber, G. (1996). *Example explanation in learning environments.* International Conference on Intelligent Tutoring Systems, pp. 457-465.

Chen, X., Mitrovic, A., & Matthews, M. (2019). Learning from Worked Examples, Erroneous Examples and Problem Solving: Towards Adaptive Selection of Learning Activities. *IEEE Transactions on Learning Technologies*. 10.1109/TLT.2019.2896080

Conati, C., Larkin, J., & VanLehn, K. (1997). A computer framework to support self-explanation. In *Proc. Int. Conf. Artificial Intelligence in Education* (Vol. 39, pp. 279-276).

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology*, *79*(4), 347.

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

D'Mello, S., Picard, R. W., & Graesser, A. (2007). Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems, 22*(4), 53-61.

Ekman, P. (1999). Facial expressions. *Handbook of cognition and emotion, 16*(301), e320.

Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction*, *17*(6), 612-634.

Hosseini, R., & Brusilovsky, P. (2017). A study of concept-based similarity approaches for recommending program examples. *New Review of Hypermedia and Multimedia*, *23*(3), 161-188.

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: a meta-analytic review. *Review of Educational Research*, *86*(1), 42-78.

Mitrovic, A. (2003). An intelligent SQL tutor on the web. *International Journal of Artificial Intelligence in Education*, *13*(2-4), 173-197.

Mitrovic, A. (2012). Fifteen years of constraint-based tutors: what we have achieved and where we are going. *User modeling and user-adapted interaction*, *22*(1-2), 39-72.

Muldner, K., & Conati, C. (2007). Evaluating a Decision-Theoretic Approach to Tailored Example Selection. Proc. IJCAI, pp. 483-488.

Najar, A. S., & Mitrovic, A. (2014). Examples and tutored problems: is alternating examples and problems the best instructional strategy? *Research & Practice in Technology Enhanced Learning, 9*(3), 439-459.

Najar, A.S., Mitrovic, A., & Neshatian, K. (2015). Eye tracking and studying examples: how novices and advanced learners study SQL examples. *Journal of computing and information technology, 23*(2), 171-190.

Najar, A. S., Mitrovic, A., McLaren, B. (2015) Examples and tutored problems: adaptive support using assistance scores. Proc. 24th Int. Joint conference on Artificial Intelligence, pp. 4317-4323.

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational psychologist, 38*(1), 15-22.

Salden, R. J., Aleven, V. A., Renkl, A., & Schwonke, R. (2009). Worked examples and tutored problem solving: redundant or synergistic forms of support? *Topics in Cognitive Science, 1*(1), 203-213.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*(4), 197-221.

Woolf, B., Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: recognising and responding to student affect. *Int. Journal of Learning Technology, 4*(3-4), 129-164.

Using Gamification to Effect Learning Behaviors in Intelligent Tutoring System

Faiza TAHIR[a*], Antonija MITROVIC[a] & Valerie SOTARDI[b]
[a]*Intelligent Computer Tutoring Group, Department of Computer Science and Software Engineering, University of Canterbury, New Zealand*
[b]*School of Educational Studies and Leadership, University of Canterbury, New Zealand*
*faiza.tahir@pg.canterbury.ac.nz

Abstract: Engagement and motivation is always a challenge in online learning environments. The benefits of learning environments have proven its history for many years, but effectively engaging users with these environments and motivating them is an active and important research problem. In this work, I will investigate the potential of gamification on motivation and user engagement in an intelligent tutoring system SQL-Tutor. This work is inspired by the growing trend of gamification and its positive effects in various domains.

Keywords: intelligent tutoring system, problem solving, gamification, self-regulation

1. Introduction and Literature Review

The goal of Intelligent Tutoring Systems (ITSs) is to provide individual support to learners according to their needs and abilities. Research shows that learning can be affected by boredom and frustration which lead to low performance for both low and high achievers (Munshi et al., 2018), and abandoned learning activities (Baker et al., 2010). Engagement and motivation are crucial for learning with ITSs (D`Mello et al., 2007). Gamification was introduced as a term with the definition "the use of game design elements in non-game contexts" (Deterding et al., 2011). It is considered as less expensive in contrast to standalone games (Landers et al., 2017). As games are originally intended for enjoyment, gamification is also defined as motivational information systems which combine the efficiency of utilitarian systems and enjoyment of hedonic systems (Koivisto et al., 2019). Adoption of gamification is reported in many fields, particularly in education, health science and crowdsourcing. Hamari and colleagues conducted surveys on gamification research in 2014 and 2019. These analysis revealed that education is an area where gamification is applied mostly and accompanied with positive results. Detailed analysis of these studies showed that they are focusing on behavioural change of learners through the use of gamification and focused only on psychological changes which are engagement, enjoyment and motivation. According to these surveys, most popular gamification elements are points, badges and leader boards and there is a huge lack of empirical evidence in gamification studies.

The theory of gamified learning presented by Landers (2014) specified the two mechanisms of introducing gamification in learning process. One is mediator, which adds game elements to affect learner's behaviour, which in turn increases/decreases learning outcomes. The other is moderator, where game intervention affects learner behaviour (psychologically) which affects the relation between instructional content and learning outcome. In subsequent work of Lander (2017), he provided the mapping of different game elements to psychological theories in order to emphasise that gamification will improve learning outcome if carefully applied with the help of psychological behaviours. On the basis of this theory, Landers and colleagues (2014) conducted an experiment by selecting time-on-task as mediating psychological behaviour and leader board as game intervention and the results were significantly improved learning.

Gamification has been applied mostly to web-based learning environments such as Code academy, Khan academy and Stack Overflow (Marder, 2015; van Roy et al., 2018) but its application in ITSs in not much explored. Few studies conducted are worth mentioning here. Denny and colleagues (2018) conducted a study on Peerwise, a system for peer learning, with points and badges added as gamification intervention. The targeted behaviours were engagement, motivation and self-testing. The results showed positive effects of gamification, particularly for badges. In another study, Long and colleagues (2014) explored the effects of adding badges to an ITS and mediate the process with re-practising behaviour. The results show partial success, and they figured out that re-practising is not an optimal mediating behaviour to improve learning. In the subsequent study, Long and colleagues (2015) explored the effects of gamification with self-regulating strategies in the ITS. The results showed the positive attitude of students towards achieving badges and other game elements (Long, Aman, & Aleven, 2015). Starting from the mixed findings about the effectiveness of gamification in ITSs, this research project will move forward the debate by empirically investigating the impact of gamification in SQL-Tutor, a mature ITS for teaching Structured Query language (Mitrovic, 2003). Many versions of SQL-Tutor have been released, providing new features and more support such as probabilistic student model (Mayo & Mitrovic, 2000), various problem selection strategies (Mayo, 2001; Mitrovic & Martin, 2004; Mathews, 2012), an animated pedagogical agent (Mitrovic & Suraweera, 2000), positive feedback (Barrow, Mitrovic, Ohlsson, & Grimley, 2008), worked examples (Shareghi Najar, Mitrovic, & McLaren, 2014) and erroneous examples support (Chen, Mitrovic, & Mathews, 2016).

This PhD project will make a number of contributions. First, we will analyse the literature to identify the psychological behaviours that work as mediators or moderators and have effect on learning performance of learners when working with ITS. On the basis of the identified learning behaviours, a gamified learning intervention will be designed and implemented in SQL-Tutor. Second, we will examine the effects of gamification on students' learning outcomes. The last contribution of this research will be to explore the relationship between enjoyment, engagement and motivation of learners with the gamified ITS.

2. Proposed Work

From the brief literature review it is evident that gamification can be promising if applied within correct context. In this project I will focus on *the impact of gamified SQL-Tutor on students' learning behaviours and subsequently their learning outcomes.* Figure 1 shows the overview of this project's theoretical framework.

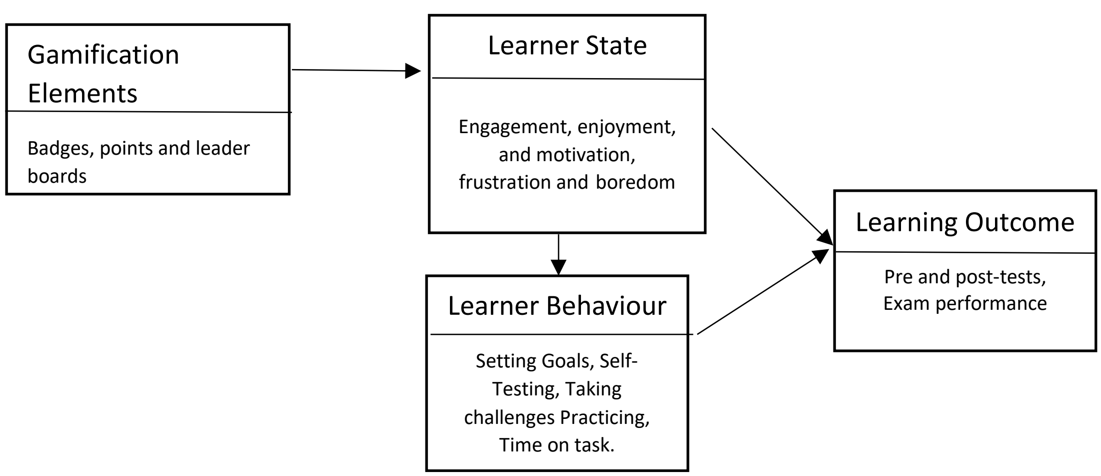


Figure 1. Theoretical framework of the project

This project will be conducted in three phases. The initial phase includes a pilot study which was conducted and I collected data about the students' affective states, their engagement and attention levels while working with SQL-Tutor and its impact on students' performance. The results showed that affect states (enjoyment, frustration, boredom) and psychological behaviours (motivation, engagement and attention) are affecting student's performance while learning from SQL-Tutor. It also revealed that students became frustrated and their engagement reduced while attempting to solve complex problems and in the absence of motivation intervention they left the problem solving. Along with identification, I will further analyse the results to find the required levels of these behaviours for achieving high performance. The learning behaviours are selected on the basis of these learner's psychological states and will be gamified in SQL-Tutor in next phase. The research question addressed in the pilot study are following:

RQ 1: To what extent do engagement, enjoyment, frustration and boredom affect students' performance while working with SQL-Tutor?

In phase two, I will design and implement badges as a gamification intervention, select learning behaviours to target through gamification and measure the learning outcomes. Badges will act as motivational affordance to increase motivation, engagement, excitement and help learners to keep going in case of negative affective states. On the basis of these psychological behaviours, I will focus on five learning behaviours: goal-setting behaviour, self-testing behaviour, taking conflict/challenge behaviour, practising and time-on-task behaviour. These behaviours will implement in SQL-Tutor with gamification techniques. For example, the goal-setting behaviour will be supported by implementing a badge that will be given when the learner completes five problems in a day, or completes five problems every day for five consecutive days. Self-testing behaviour will be supported by providing badges when students attempt optional quizzes. The conflict/challenge behaviour will be supported by providing badges for solving complex problems and daily challenges. Badges will also be provided on completing problems daily, for completing more difficult problems (e.g., problems requiring the Group by clause) or completing a specific number of problems in one session. These learning behaviours act as mediators between gamification and learning outcomes as suggested by the Landers theory of gamified learning. The research questions addressed in this phase are following:

RQ 2: Do badges influence students to complete more problems, and remain motivated, enjoyed and engaged for longer?

RQ 3: Which learning behaviours act as optimal mediators to increase students' performance in the presence of badge interventions? (This RQ will be investigated in five subquestions, each focusing on a specific learning behaviour)?

On the basis of the finding from the study, I plan to introduce other gamification interventions in the last phase of my project. Leader boards and points are other two popular gaming intervention as mentioned by Hamari, (2019). I will conduct another study in order to compare the effects of various types of gaming interventions along with their gaming attributes and learning behaviours. The research questions addressed in this phase are:

RQ 4: Is a combination of points, badges and leader boards more effective than when those elements are used individually?

RQ 5: Which learning behaviours best combine with points, badges and leader board and yields optimal result in terms of student performance?

3. Research Methodology

To answer the research questions above, I will conduct exploratory research to determine the main aspects of gamification in SQL-Tutor, and develop techniques and procedures to apply gamification effectively. This project will consist of three experiments based on three phases mentioned above and system in context is SQL-Tutor. Phase one consist of pilot study focused on RQ1 started with the demographic questionnaire and then identify and analyse student's affective states and psychological behaviours with the help of iMotions (https://imotions.com) software package. Phase 2 and 3 focusing on RQ2-RQ5, a classroom study will be conducted in each phase which will provide the modified version of SQL-Tutor to one group of randomly selected students (experimental) and standard version of SQL-Tutor to other group of students (control). The data will be quantitative in nature in all three phases and collected through the student logs of SQL-Tutor and exported from iMotions. The effects or learning outcome/performances of students will be measured via pre-/post-tests. This data will be analysed with the help of correlation and regression analysis. At the end of phase 2 and 3 a questionnaire will be provided to get the opinion and future intentions of students on using gamification interventions and learning behaviours. Motivation will analyse both by student interaction with the SQL-Tutor and through the questionnaire provided at the end of each experiment.

4. References

Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *Human-Computer Studies*, *68*(4), 223-241.

Barrow, D., Mitrovic, A., Ohlsson, S., & Grimley, M. (2008). Assessing the impact of positive feedback in constraint-based tutors. In International Conference on Intelligent Tutoring Systems (pp. 250-259). Springer, Berlin, Heidelberg.

Chen, X., Mitrovic, A., & Mathews, M. (2016). Do Erroneous Examples Improve Learning in Addition to Problem Solving and Worked Examples?. In International Conference on Intelligent Tutoring Systems (pp. 13-22). Springer, Cham.

D'Mello, S., Picard, R. W., & Graesser, A. (2007). Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems*, *22*(4), 53-61.

Denny, P., McDonald, F., Empson, R., Kelly, P., & Petersen, A. (2018). Empirical Support for a Causal Relationship Between Gamification and Learning Outcomes. Proc. CHI Conference on Human Factors in Computing Systems, Montreal, Canada (p. 311). ACM.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining "gamification". Proc. 15th Int. Academic MindTrek Conference: Envisioning Future Media Environments, Tampere, Finland (pp. 9-15). ACM.

Hamari, J., Koivisto, J., & Sarsa, H. (2014). *Does gamification work?--a literature review of empirical studies on gamification.* Proc. 47th Hawaii Int. Conf. on System Sciences, 14, pp. 3025-3034.

Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. *International Journal of Information Management, 45*, 191-210.

Landers, R. N. (2014). Developing a theory of gamified learning: Linking serious games and gamification of learning. *Simulation & Gaming*, *45*(6), 752-768.

Landers, R. N., Armstrong, M. B., & Collmus, A. B. (2017). How to use game elements to enhance learning: Applications of the theory of gamified learning. *Serious Games and Edutainment Applications* (pp. 457-483): Springer.

Landers, R. N., & Landers, A. K. (2014). An empirical test of the theory of gamified learning: The effect of leaderboards on time-on-task and academic performance. *Simulation & Gaming, 45*(6), 769-785.

Long, Y., & Aleven, V. (2014). Gamification of joint student/system control over problem selection in a linear equation tutor. Proc. Int. Conf. Intelligent Tutoring Systems (pp. 378-387). Springer, Cham.

Long, Y., Aman, Z., & Aleven, V. (2015). Motivational design in an intelligent tutoring system that helps students make good task selection decisions. Proc. Int. Conf. Artificial Intelligence in Education (pp. 226-236). Springer, Cham..

Marder, A. (2015). Stack overflow badges and user behavior: an econometric approach. Proc. IEEE/ACM 12th Working Conference on Mining Software Repositories (pp. 450-453). IEEE.

Mathews, M. M. (2012). "A Framework for Multiple Adaptable Pedagogical Strategies in Intelligent Tutoring Systems."

Mayo, M., & Mitrovic, A. (2000). Using a probabilistic student model to control problem difficulty. In International Conference on Intelligent Tutoring Systems (pp. 524-533). Springer, Berlin, Heidelberg.

Mayo, M., & Mitrovic, A. (2001) Optimising ITS Behaviour with Bayesian Networks and Decision Theory. *Artificial Intelligence in Education*, 12(2), 124-153.

Mitrovic, A., & Suraweera, P. (2000). Evaluating an animated pedagogical agent. In International Conference on Intelligent Tutoring Systems (pp. 73-82). Springer, Berlin, Heidelberg.

Mitrovic, A. (2003). An intelligent SQL tutor on the web. International Journal of Artificial Intelligence in Education, 13(2-4), 173-197.

Mitrovic, A., & Martin, B. (2004). Evaluating adaptive problem selection. In International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (pp. 185-194). Springer, Berlin, Heidelberg.

Munshi, A., Rajendran, R., Ocumpaugh, J., Biswas, G., Baker, R. S., & Paquette, L. (2018). Modeling Learners' Cognitive and Affective States to Scaffold SRL in Open-Ended Learning Environments. Proc. 26th Conf. User Modeling, Adaptation and Personalization, Singapore, Singapore (pp. 131-138). ACM.

Shareghi Najar, A., Mitrovic, A., & McLaren, B. M. (2014). Adaptive Support versus Alternating Worked Examples and Tutored Problems: Which Leads to Better Learning?. In International Conference on User Modeling, Adaptation, and Personalization (pp. 171-182). Springer, Cham.

van Roy, R., Deterding, S., & Zaman, B. (2018). Uses and Gratifications of Initiating Use of Gamified Learning Platforms. Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (p. LBW565). ACM.

Investigating the Effects of Gamifying SQL-Tutor

Faiza TAHIR[a*], Antonija MITROVIC[a] & Valerie SOTARDI[b]
[a]*Intelligent Computer Tutoring Group, Department of Computer Science and Software Engineering, University of Canterbury, New Zealand*
[b]*School of Educational Studies and Leadership, University of Canterbury, New Zealand*
*faiza.tahir@pg.canterbury.ac.nz

Abstract: The practice of adding game elements to non-gaming educational environments has gained much popularity. Gamification has been found in some studies to increase learner engagement, motivation, and academic performance. However, there is a lack of empirical evidence to prove the effects of gamification in advanced learning technologies like Intelligent Tutoring Systems (ITS). This paper reports the results of an empirical study that included three categories of game elements (goals, assessment and challenges) implemented as badges in the context of SQL-Tutor. The study was conducted in a class under realistic conditions. SQL-Tutor was used voluntarily by 77 undergraduate students enrolled in a second-year database course. Although there were no differences between the experimental and control groups in terms of their interaction with SQL-Tutor and learning outcomes, we found a significant mediating effect of time-on-task on the direct relation between badges and achievement in the gamified condition. We also found evidence that not all students were interested in badges.

Keywords: Gamification, goals, assessment, challenges, badges, learning behavior, time-on-task, mediation effect.

1. Introduction

Engagement and motivation are crucial for effective learning. The amount of user interaction with an educational system is an important indicator of learning outcomes. In online learning, engagement refers to the student's involvement with the system and motivation refers to his/her determination to achieve a goal. One strategy to increase motivation is gamification, i.e. the use of gaming elements such as leaderboards, points, badges and other virtual achievements common in games. These virtual achievements are not always connected to a tangible reward; they are meant to increase user involvement and their motivation to use those applications. For example, the TripAdvisor website (tripadvisor.com) rewards its users' points which do not have any monetary value. Badges are commonly used in educational environments. For example, PeerWise (Denny et al., 2018) awards virtual badges to students for writing or answering questions. Leaderboards are often used in applications where social activities are important, like comparing the performance of users in a course.

The term gamification was first used almost a decade ago (Deterding et al., 2011) and has gained much popularity. Gamification was found to be effective in many projects in maintaining user engagement by encouraging their actions and fostering quality and productivity of those actions (Hamari, 2013). However, the application of gamification in non-gaming environments does not always yield positive results. In a few cases, gamification may go unnoticed by users, and in other cases, it had negative effects on users which were completely unintended (Diefenbach & Müssig, 2019). Moreover, despite the growing number of educational environments incorporating gamification, there is a lack of empirical evidence proving its efficiency in a particular context/environment. Gamification might help in increasing engagement, enjoyment and motivation. However, if the learning environment is not proved to improve learning, gamification would not help. On the other hand, if an educational system is highly effective, gamification may not provide an additional benefit. Therefore, the process of applying gamification in a particular system should consider both the system's effectiveness and the impact of gamification on the learner's behaviour.

Intelligent Tutoring Systems (ITSs) have a long history of proven results in education. There are many strategies used to address engagement and motivation in ITSs, such as supporting metacognitive strategies, e.g. self-regulation and self-assessment (Long & Aleven, 2013) and supporting affective states of learners. This study aims to explore the effects of gamification in SQL-Tutor (Mitrovic, 1998; Mitrovic, 2003), a mature ITS that teaches the Standard Query Language (SQL). The effectiveness of SQL-Tutor has been proven in multiple studies (Mitrovic & Ohlsson, 1999; Mitrovic, 2012). We start by providing a brief literature review of gamification and its effects. Section 3 presents our approach to gamifying SQL-Tutor, while Section 4 discusses the experiment design. We then present our findings in Section 5, and finally, the conclusion and limitations of the current work.

2. Related Work

Gamification is defined as "the use of game design elements in non-game contexts" (Deterding et al., 2011). It is considered to be less expensive in contrast to standalone games (Dicheva et al., 2015; Landers et al., 2017). As games are originally intended for enjoyment, gamification is also defined as motivational information systems which combine the efficiency of utilitarian systems and enjoyment of hedonic systems (Koivisto et al., 2019). Adoption of gamification is reported in many fields, particularly in education, health science and crowdsourcing. Several systematic literature reviews (Hamari, Koivisto, & Sarsa, 2014; Koivisto & Hamari, 2019) report that the most used game elements are points, badges and leaderboards, and the largest positive effects are on motivation and engagement, and less so on learning outcomes. However, not all study report positive results, with some even reporting negative effects of gamification on students' motivation and learning. Detailed analysis of these studies showed that they were focusing on behavioural change of learners through the use of gamification and focused primarily on engagement, enjoyment and motivation. These reviews also point out methodological problems with the evaluations studies, which include small sample sizes, lack of control conditions, evaluating several gamification elements simultaneously and short duration of studies.

The theory of gamified learning proposed by Landers (Landers 2014; Landers et al., 2017) specifies that gamification has an effect on learning by influencing the learner's behaviours or attitudes, via two theoretical paths. Some gamification elements influence learning behaviours/attitudes, which in turn directly influence learning outcomes; thus, gamification acts as a mediator. In other situations, the influence of students' behaviours or attitudes change the effectiveness of instructional content – that is gamification moderates the relationship between the content and learning outcomes. In a study using leader boards and the time-on-task as the mediating behaviour, Landers and Landers (2014) found a significant improvement in learning.

Gamification has been applied to many web-based learning environments such as Code academy, Khan Academy and Stack Overflow (Marder, 2015; van Roy et al., 2018), and with mixed effects on student learning. Denny and colleagues (2018) conducted a study on Peerwise, a system for peer learning, with points and badges added as the gamification intervention and proved their effectiveness by targeting the engagement, motivation and self-testing behaviour. In another similar study, gamification was examined on university students and computer games development course was gamified (O'Donovan, Gain, & Marais, 2013). The gamification elements were experience points, badges, leader boards, storyline and theme, presented with the help of gamified visuals. The study reported significant improvements in terms of student engagement and motivation, and the leader board was considered the biggest motivational element. The behaviours influenced most were attendance and attempting quizzes.

In another study, Haaranen and colleagues (2014) investigated the effects of badges in an online learning environment for a data structures and algorithm course. The badges were awarded for time management, early submissions and successfully completing exercises. The results showed that students were mostly indifferent about badges, and also the badges did not have significant effects on student behaviours and learning outcomes. The authors reported that students stopped working once they achieved enough scores for passing the course. However, no negative effects of badges were observed, and the authors suggested that the effects of gamification were highly context-dependent.

There is very little research focusing on gamification of ITSs. Long and colleagues (2014) explored the effects of two gamification features in Lynette (ITS), which is re-practising of previously completed problems and rewards over the student performance on each completed problem. The results showed that gamifying ITS does not result in increased learning or enjoyment of students. However, the highest learning gains were reported for those students who re-practised previously completed problems but received no rewards on their performance (Long & Aleven, 2014). In the subsequent study, Long rewarded (star) students when they selected unmastered problems and showed perseverance on practising new problems and found encouraging results of these strategies on student's metacognitive skills (Long & Aleven, 2016).

This brief overview of literature acknowledges that three methodological gaps exist: 1) the effects of gamification are highly context-dependent and may be overlooked in research designs, 2) research on gamification inconsistently considers students' behaviours or attitudes and 3) insufficient design guidelines are available due to a lack of empirical studies. Our study attempts to fill all these gaps.

3. Gamifying SQL-Tutor

We selected three categories of game elements from the nine categories discussed in the Theory of gamified learning (Landers et al., 2017): goals, assessment and challenges. Challenges grow the competition in students either in the form of standing in the class or achievement of the skill. Research (Munshi et al., 2018) shows that student become bored/frustrated if they are not challenged enough. Therefore, complex problems in the form of challenges can be helpful to retain their interest. Goals are also considered as a form of challenge; however goal-setting theory states that goals can motivate students if they are SMART (specific, measurable, achievable, realistic and time-bound) (Locke & Latham, 1990). The goals selected in this study are according to these lines: they have only one condition (specific), can be measured through completed problems (measurable), achievable, realistic, and can be achieved within the 4-weeks study period (time-bound). The difference between challenges and goals lies in the complex and hard to achieve challenges. SQL-Tutor provides assessments in the form of pre/post-tests at the start/end of study.

We implemented goals, assessment and challenges in SQL-Tutor via different types of badges (Table 1). The goal-setting behavior is supported by fixing daily and weekly goals stated as wining criteria for badges. The self-testing behavior is addressed by providing a quiz. Challenges are implemented via several badges, and also as daily challenges, which consist of complex unsolved problems. We hypothesize that all these game elements influence time-on-task, which has been shown in many studies to influence learning outcomes (Landers et al., 2014; Denny et al., 2018).

Table 1. *Definitions of badges and the relevant learning behaviors*

Group	Badge	Criterion	Behavior	Earned By
Primary	Go getter	Completing the first problem	Goal-setting	100%
	High flyer	3 problems in one session	Goal-setting	100%
	Achiever	5 problems in a day	Goal-setting	100%
	Activist	5 problems without complete solution	Challenge	16.66%
	Leader	problem with the "Group by" clause	Challenge	16.66%
Classic	Energy house	6 problems in a row	Goal-setting	100%
	Scholar	5 problems/day for 5 consecutive days	Goal-setting	2.38%
	Fireball	10 problems in one day	Goal-setting	92.80%
	Champion	First daily challenge	Challenge	7%
Elite	Genius	Attempting the quiz	Self-testing	38.09%
	Human dynamo	5 problems/day for 10 days	Goal-setting	0%
	Einstein	5 daily challenges over 2 weeks	Challenge	0%
	Live-Wire	5 problems per day for 20 days	Goal-setting	0%

The thirteen badges are divided into three groups: primary, classic and elite. The purpose of primary badges is to grab the student's attention at the early stage of using SQL-Tutor, such as awarding a badge for solving the first problem, or for solving a problem using a difficult clause (group by). This category also includes the *Activist* badge which discouraged the use of "complete solution". Please note that when the student submits a solution to SQL-Tutor, he/she can also specify the level of feedback. The complete solution is the highest level of feedback in SQL-Tutor, which provides the full solution to the problem. Therefore, the Activist badge checks that the student solved the problem on his/her own, rather than copying the full solution provided by the system.

The classic group contains four badges, which emphasize practicing regularly, for example completing five problems for consecutive 5 days and solving complex problems of the daily challenge. The last group, elite badges, consists of four badges and their main purpose is to keep engaging the student with SQL-Tutor over a longer period of time. In this category, badges are awarded when the student completes five problems every day for ten days, or solves five daily challenges in two weeks. The last badge awarded to those extraordinary students who completed five problems every day, for 20 consecutive days.

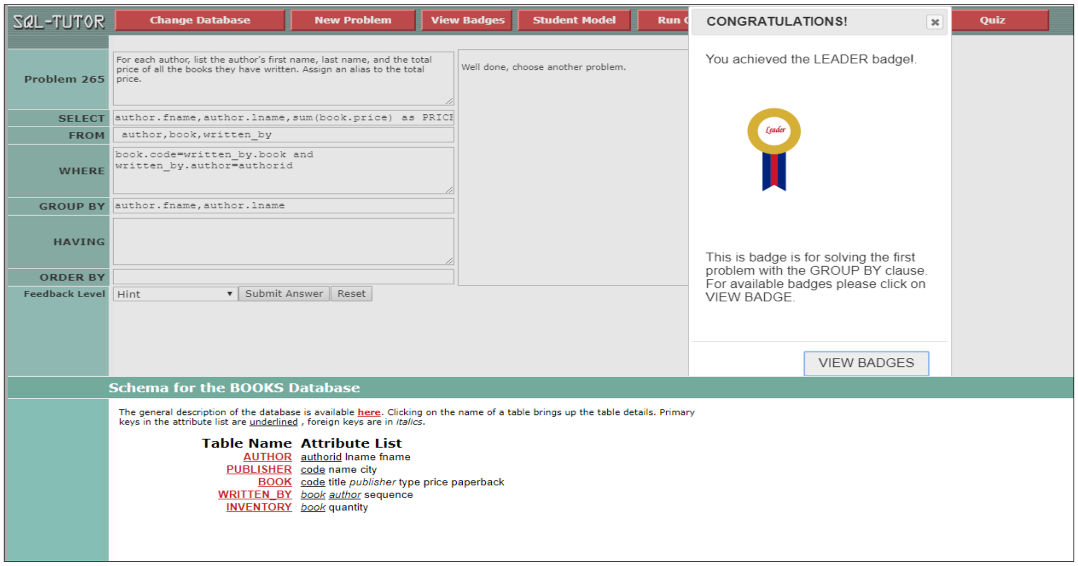


Figure 1. Notification of winning a badge

When the student fulfills the condition for a badge, he/she receives the notification about that badge immediately, as shown in Figure 1. Students can view all the badges awarded to them on the badge page, which also showed the badges which have not been achieved yet. SQL-Tutor also provides an Open Learner Model (OLM), in the form of skill meters. For the study, we modified the OLM page to show the next badge the student could achieve, as shown in Figure 2.

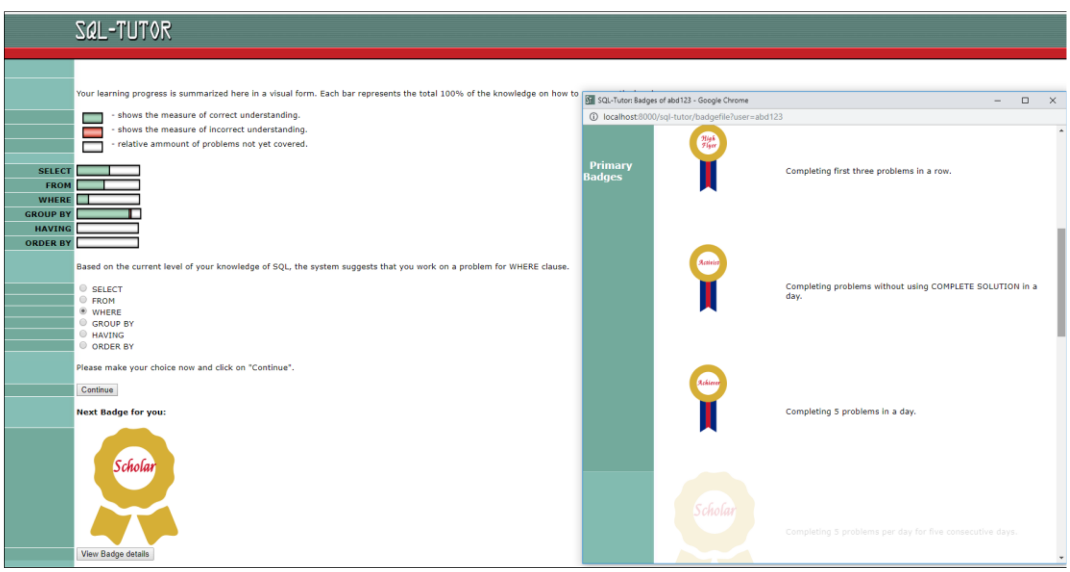


Figure 2. The OLM page, illustrating the next badge (left); the badge page (right)

Daily challenges are presented to students once they achieve all primary badges. A daily challenge consists of three problems, selected adaptively based on the student model. The problems selected for a daily challenge need to be challenging for the student. SQL-Tutor summarizes the student's learning progress using the student level, which ranges from 1 to 9. Problems in SQL-Tutor also have a complexity level (defined by the teacher) ranging over the same scale. Therefore the problems selected for the daily challenge are previously unsolved problems, which satisfy two conditions: 1) their level of complexity is equal to the current student level or one level higher, and (2) these problems require the clauses of the SELECT statement which the student needs to practice (as per the student model). Each day, the daily challenge is presented to the student upon logging in, and is also available on the problem-selection page. Two badges (*Champion* and *Einstein*) are awarded when the student completes the first daily challenge, or when the student completes five daily challenges over two weeks respectively.

We also developed a quiz, consisting of seven multiple choice questions and two true/false questions. The *Genius* badge is awarded for attempting the quiz, independently on the score achieved. When the student completes a quiz, the scores is shown immediately, so that the student can reflect on his/her knowledge. Awarding badge on attempting the quiz maximizes the effects of students' self-testing abilities.

4. Experimental Procedure and Hypotheses

The participants were recruited from the 198 students enrolled in the second-year course on relational database systems at the University of Canterbury in 2019. Before the study, the students have learnt about the relational data model and SQL in lectures and had two labs sessions, in which they created tables and performed basic SQL queries in Oracle. The students were introduced to SQL-Tutor in a lab session. The use of SQL-Tutor was voluntary; the students did not receive any course credit for solving problems in SQL-Tutor. All enrolled students were randomly allocated to the control group (using the standard version of SQL-Tutor) or the experimental group, who used the gamified version. We obtained informed consent from 77 students (25% female, 62% male, 13% not specified); 42 in the experimental group and 35 in the control group.

The study lasted for four weeks. When students logged into SQL-Tutor for the first time, they received the pre-test, a short demographic questionnaire and a question about their previous experience of using gamification. The students could use SQL-Tutor whenever they wanted. The quiz was given at the end of the second week of the study to both control and experimental groups. The pre/post-test and the quiz were of similar complexity; each contained seven multiple choice questions and two true/false questions (worth one mark each).

The post-test was administered online at the end of the fourth week. A major piece of the course assessment was the lab test focusing on SQL, worth 20% of the final grade. The lab test was given two days after the post-test. After the lab test, the students were invited to complete a survey. There were two versions of the survey. For the experimental group, there were four questions related to their opinion of the badges, and two questions related to daily challenges. Both groups received two questions about the quiz. The responses to these questions were recorded on the 5-point Likert scale, from 'strongly disagree' (1) to 'strongly agree' (5).

We made the following hypotheses, based on the results from literature (e.g. Landers & Landers, 2014), and from our own experience:

H1: The time-on-task is positively correlated with learning outcomes.
H2: The experimental group participants will spend more time solving problems in SQL-Tutor in comparison to the control group.
H3: Badges will have a mediating effect on learning outcomes, by influencing the time-on-task.

5. Results

The average score on the pre-test was 58.73% (sd = 26.05). The students interacted with SQL-Tutor on 3.39 days (referred to as *Active Days*) over four weeks (sd = 2.69, min = 1, max = 12), spending 260 min (min = 41, max = 1,441, sd = 243) in the system. During that time, the students solved an average of 37.47 problems (sd = 34.74, min = 3, max = 204). Only 28 students completed the post-test; we believe the reason for the low completion rate was that the post-test was not mandatory. In addition, the post-test was given to the students only two days before the lab test. The average score on the post-test was 69.05% (ds = 25.90). For the lab test, the average score was 60.83% (sd = 17.07). In addition to defining queries, which students practiced in SQL-Tutor, the lab test covered other SQL topics, and therefore the lab test cannot be considered as the direct learning outcome. For those reasons, we use the student level at the end of the interaction with SQL-Tutor as a measure of students' learning. The average student level was 3.56 (sd = 1.66, min = 1, max = 8). In the experimental group, 66% of students reported having used some form of gamification, compared to 57% of the control group participants.

Evaluating the Hypotheses

To evaluate H1, we regressed the student level on time-on-task. The time-on-task strongly predicts the student level ($\beta = .536$), and was statistically significant ($t = 5.5$, $p < .001$). Variance in student level explained by time-on-task was 28.7%. Therefore, hypothesis H1 was supported.

Table 2 presents statistics for the two groups. There was no significant difference on the pre-test scores of the two groups, showing that the students had comparable levels of pre-existing knowledge. The experimental group students spent more time on task, had more sessions, attempted and solved more problems, and attempted more complex problems in SQL-Tutor in comparison to the control group, although none of the differences are significant. Therefore, our hypothesis H2 is not supported. There was also no significant difference between the groups on the number of active days, student levels, the post-test and lab test scores.

Table 2. *Summary statistics of SQL-Tutor usage: mean (sd)*

	Experimental (42)	Control (35)
Pre-test %	59.52 (24.02)	57.78 (28.62)
Time-on-task (min)	288.40 (302.02)	225.94 (143.44)
Sessions	7.29 (7.84)	6.11 (4.49)
Active Days	3.33 (3.09)	3.46 (2.13)
Attempted problems	42.26 (42.75)	37.34 (26.94)
Solved Problems	39.33 (40.99)	35.23 (25.72)
Max Problem Complexity	6.95 (1.78)	6.71 (2.02)
Student level	3.31 (1.62)	3.86 (1.68)
Post-test %	n = 17, 67.97 (26.32)	n = 11, 70.71 (26.42)
Lab test %	60.43 (16.49)	61.31 (17.97)

To evaluate H3, we used the data for the experimental group only. We analyzed the mediation effect using the Process macro, version 3.5 software for SPSS (Hayes, 2017), with the student level as the dependent variable. Figure 3 shows the standardized regression coefficients for the mediation model. The direct effect of badges on the student level is not significant ($p = .08$), but the significant relationship in this first step is not a requirement for mediation (Shrout & Bolger, 2002). The direct effect of badges on time is significant ($p < .001$), as is the direct effect of time on the student level ($p < .005$). The indirect and total effects in the model are tested using bootstrap samples and 95% confidence

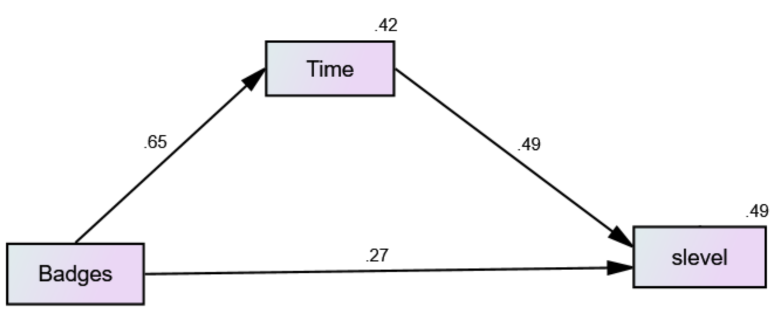


Figure 2. The mediation model, with standardized coefficients

intervals. Results show that the standardized, indirect effect of badges on the student level is β = 0.32. The confidence interval for the estimate of the indirect effect [.165, .501] does not include zero; therefore the null hypothesis is rejected. 52.26% of the total effect is mediated. The Sobel test of significance of mediation gives 2.62 ($p < .01$), indicating that time on task mediates the direct relationship between the number of badges and the student level. Therefore, hypothesis H3 is confirmed.

Further Investigation of the Experimental Group

Overall, the experimental group students achieved from 4 to 7 badges, with a mean of 5.43 (sd = .86). The percentage of students from the experimental group who earned various badges is shown in the last column of Table 1. On the very first day of interacting with SQL-Tutor, the students achieved an average of 4.60 badges (sd = .76). Only seven students achieved all primary badges; therefore they were the only ones who were given daily challenges. For that reason, it is not possible to make any conclusions about the daily challenges.

The literature review shows that in some cases, students are not interested in badges when they are not directly related to course credit. To investigate whether there is a difference in how much the experimental group students were interested in badges, we divided the experimental group students into two subgroups: those who visited the badge page at least once (23 students), and those who have never visited that page (19 students). Table 3 presents the differences found between the two subgroups.

Table 3. *Comparing experimental group students who visited the badge page or not: mean (sd)*

	Seen badge page (23)	Not seen (19)	Significant
Pre-test %	54.59 (25.05)	65.49 (21.88)	$p = .22$
Time-on-task (min)	365.30 (272.27)	195.32 (316.96)	$U = 348.5, p < .001$
Sessions	9.48 (7.69)	4.63 (7.37)	$U = 334.5, p < .005$
Active Days	4.13 (3.22)	2.37 (2.71)	$U = 312.5, p < .05$
Attempted problems	51.91 (39.51)	30.58 (44.62)	$U = 332, p < .005$
Solved Problems	47.48 (36.86)	29.47 (44.49)	$U = 326.5, p < .01$
Constraints	287.74 (60.98)	247.84 (75.82)	$U = 299.5, p < .05$
Badges	5.74 (.81)	5.05 (.78)	$U = 317, p < .01$
Student level	3.70 (1.72)	2.84 (1.39)	$p = .07$
Post-test %	n = 13; 4.38 (2.93)	n = 8; 5.88 (3.72)	$p = .34$
Lab test %	59.74 (13.90)	61.26 (19.55)	$p = .81$

There was no significant difference between the two subgroups on the pre-test scores. The students who visited the badge page have interacted with SQL-Tutor significantly more, measured either as the total time ($p < .001$)), the number of sessions ($p < .005$), or the number of active days ($p < .05$). Those students attempted/solved more problems ($p < .005$ and $p < .01$ respectively) than their peers, and also achieved significantly more badges ($p < .01$). The students who have seen more badges have used significantly more constraints than their peers. In SQL-Tutor, domain knowledge is represented in terms of more than 700 constraints. Therefore, the students who visited the badge page covered a higher proportion of the domain in comparison to their peers. Therefore, there is evidence that visiting the badge page is correlated with more time-on-task and engagement. However, there was no significant difference between the two subgroups in terms of learning, measured either by the student level achieved ($p = .07$), post-test scores ($p = .34$) or the lab test score ($p = .81$).

Self-testing Behavior

As mentioned in Section 4, the quiz was completely optional and provided to both experimental and control groups. To analyze students' self-testing behavior, we investigated whether there is a difference in the student level achieved based on whether the students took the quiz and the group they were in. We introduced a dummy QuizTaken variable, with values of 0 (quiz not taken) or 1 (quiz taken). In the control group, 12 students attempted the quiz while 23 did not. For the experimental group, 14 out of 42 students attempted the quiz. A two-way ANOVA (F = 3.07, $p < .05$, partial $\eta^2 = .11$) revealed neither a significant interaction between group and QuizTaken, nor the main effect of group, but there was a significant effect of the self-testing behavior ($p = .01$, partial $\eta^2 = .09$) Students who attempted the quiz achieved a significantly higher student level.

Table 4. *Student level*

Group	QuizTaken	Students	Student Level
Control	0	23	3.48 (1.38)
	1	12	4.58 (2.02)

Table 5 presents the statistics for students who attempted or did not attempt the quiz. Ther was no significant difference on the pre-/post-test scores and the lab test scores. The students who attempted the quiz interacted with SQl_Tutor sitnificantly more, measured in terms of tme, sessions, active days and attempted/solved problems. They used more constraints and solved more complex problems, thus achieveing higher student levels.

Table 5. *Comparing students who attempted/not attempted the quiz: mean (sd)*

	Not attempted (51)	Attempted (26)	Significant
Pre-test %	56.65 (25.75)	62.82 (26.66)	p = .33
Time-on-task (min)	189.73 (153.89)	397.88 (321.47)	t = 3.85, p < .001
Sessions	5.20 (5.43)	9.81 (7.46)	t = 3.09, p < .005
Active Days	2.39 (1.86)	5.35 (3.01)	t = 5.32, p < .001
Attempted problems	28.27 (21.37)	63.08 (47.47)	t = 4.44, p < .001
Solved Problems	25.98 (19.09)	60.00 (46.28)	t = 4.56, p < .001
Max Problem Complexity	6.37 (1.93)	7.77 (1.42)	t = 3.26, p < .005
Constraints	244.24 (62.44)	317.23 (63.09)	t = 4.83, p < .001
Student level	3.22 (1.43)	4.23 (1.88)	t = 2.64, p < .05
Post-test %	n = 13; 4.38 (2.93)	n = 8; 5.88 (3.72)	p = .08
Lab test %	59.74 (13.90)	61.26 (19.55)	p = .10

5.3 Survey Responses

We received 21 survey responses from the experimental group and 22 responses from the control group students. Table 6 summarizes the responses to the four questions on badges from the experimental group students. The Cronbach alpha for those questions is 0.88.

Table 6. *Responses from the experimental group (1 - strongly disagree to 5 - strongly agree)*

Question	1	2	3	4	5
Badges motivated me to participate more than I would have otherwise.	22%	26%	39%	4%	9%
I found being able to earn badges increased my enjoyment of using SQL-Tutor	9%	35%	26%	26%	4%
I would prefer not to see badges in SQL-Tutor.	0%	39%	35%	17%	9%
The badges awarded for solving problems motivated me to solve more problems than I would have otherwise.	17%	31%	39%	13%	0%

The responses of the experimental group indicate that students did not find badges very motivating. Students were indifferent in their responses about the enjoyment when they received badges. However, 39% of students stated they wanted to see the badges. We do not discuss the questions on daily challenges, as only seven students received them during the study. Almost 62% of students wanted to see the daily challenges in SQL-Tutor; this figure reveals that students were interested in daily challenges in principle. The students from both groups enjoyed attempting quiz (control = 68%, experimental = 62%) and prefer to see them in SQL-Tutor (control= 86%, experimental = 62%).

6. Conclusions

This paper presents a classroom study in which we analyzed the effect of gamification in the context of SQL-Tutor. Our findings highlight the effects of gamification in the context of an ITS, under realistic conditions, in a study that lasted four weeks.

Starting from Lander's theory of gamified learning (2014), we designed badges which supported goal setting, assessment and challenges—three common categories of game elements. We hypothesized that the badges would motivate students to spend more time on task (i.e. problem solving in SQL-Tutor). The goal-setting behavior is supported by setting SMART goals/criteria for achieving each badge. Challenges motivate students to perform more complex tasks, and the quiz allowed students to test their knowledge.

Our study provides initial evidence that badges can positively increase student achievement (measured as the student level achieved in SQL-Tutor), and that this relation can be mediated by the amount of time participants spend on the task. The results show the impact of gamification on learning through behavioral change, supporting the theory of gamified learning with the time-on-task as a valid behavior target for gamification. From the statistical analysis, we first determined that time-on-task correlates and predicts learning outcomes. We did not find a difference between gamified and non-gamified groups in terms of time spent in SQL-Tutor, problems completed, and learning outcomes. A possible explanation for this finding is that the students are already highly motivated, and used SQL-Tutor to prepare for the lab test. However, we found evidence that goal-setting, challenges and self-testing behaviors implemented as badges indirectly and significantly affect learning outcomes through the time-on-task as the mediator.

There are two major limitations of our study, the first being the small sample size. The second limitation was the design of the badges, which could be designed in a more visually attractive manner. As discussed, almost 46% of students in the experimental group did not access the badge page despite receiving badge notifications. This shows that the design of badges was not attractive enough to entice some learners and motivate them to achieve.

7. Acknowledgements

We thank Jay Holland for helping with the study, as well as our participants.

8. References

Denny, P., McDonald, F., Empson, R., Kelly, P., & Petersen, A. (2018). Empirical Support for a Causal Relationship Between Gamification and Learning Outcomes. *Proc. CHI Conference on Human Factors in Computing Systems*, Montreal, Canada (p. 311). ACM.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining "gamification". *Proc. 15th Int. Academic MindTrek Conference: Envisioning Future Media Environments*, Tampere, Finland (pp. 9-15). ACM

Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in education: A systematic mapping study. *Educational Technology & Society*, 18(3), 75-88.

Diefenbach, S., & Müssig, A. (2019). Counterproductive effects of gamification: An analysis on the example of the gamified task manager Habitica. *Human-Computer Studies*, 127: p. 190-210.

Hamari, J. (2013). Transforming homo economicus into homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service. *Electronic commerce research and applications, 12*(4), 236-245.

Hamari, J., Koivisto, J., & Sarsa, H. (2014). *Does gamification work?--a literature review of empirical studies on gamification.* Paper presented at the 2014 47th Hawaii international conference on system sciences.

Haaranen, L., Ihantola, P., Hakulinen, L., & Korhonen, A. (2014). How (not) to introduce badges to online exercises. Proc. 45th ACM technical symposium on Computer science education (pp. 33-38).

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.

Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. International Journal of Information Management, 45, 191-210.

Landers, R. (2014). Developing a theory of gamified learning: Linking serious games and gamification of learning. *Simulation & gaming*, *45*(6), 752-768.

Landers, R., & Landers, A. (2014). An empirical test of the theory of gamified learning: The effect of leaderboards on time-on-task and academic performance. *Simulation & Gaming*. 45(6): p. 769-785.

Landers, R., Armstrong, M., & Collmus, A. (2017). How to use game elements to enhance learning: Applications of the theory of gamified learning. Serious games and edutainment applications, 457-483.Hamari, J. (2013). Transforming homo economicus into homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service. Electronic commerce research and applications, 12(4), 236-245.

Locke, E. A., & Latham, G. P. (1990). A theory of goal setting & task performance: Prentice-Hall, Inc.

Long, Y., & Aleven, V. (2013). Supporting students' self-regulated learning with an open learner model in a linear equation tutor. Paper presented at the International conference on artificial intelligence in education. (pp. 219-228). Springer, Berlin, Heidelberg.

Long, Y., & Aleven, V. (2014). Gamification of joint student/system control over problem selection in a linear equation tutor. Paper presented at the International Conference on Intelligent Tutoring Systems. (pp. 378-387). Springer, Cham.

Long, Y., & Aleven, V. (2016). Mastery-oriented shared student/system control over problem selection in a linear equation tutor. Paper presented at the International conference on intelligent tutoring systems. (pp. 90-100) Springer, Cham.

Marder, A. (2015). Stack overflow badges and user behavior: an econometric approach. *Proc. IEEE/ACM 12th Conf. Mining Software Repositories* (pp. 450-453). IEEE.

Mitrovic, A. (1998). Experiences in implementing constraint-based modeling in SQL-Tutor. In B. Goettl, H. Halff, C. Redfield, V. Shute (Eds.) Proc. *Int. Conf. Intelligent Tutoring Systems* (pp. 414-423). Springer, Berlin, Heidelberg.

Mitrovic, A. & Ohlsson, S. (1999). Evaluation of a Constraint-Based Tutor for a Database Language. *Artificial Intelligence in Education*, 10(3-4), 238-256.

Mitrovic, A. (2003). An intelligent SQL tutor on the web. *Artificial Intelligence in Education*, 13(2-4), 173-197.

Mitrovic, A. (2012). Fifteen years of constraint-based tutors: what we have achieved and where we are going. *User modeling and user-adapted interaction*, 22(1-2), 39-72.

Munshi, A., Rajendran, R., Ocumpaugh, J., Biswas, G., Baker, R. S., & Paquette, L. (2018). *Modeling learners' cognitive and affective states to scaffold SRL in open-ended learning environments.* Paper presented at the Proceedings of the 26th conference on user modeling, adaptation and personalization

O'Donovan, S., Gain, J., & Marais, P. (2013). A case study in the gamification of a university-level games development course. *Proc. South African Institute for Computer Scientists and Information Technologists Conference* (pp. 242-251). ACM.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: new procedures and recommendations. *Psychological methods*, *7*(4), 422-445.

Van Roy, R., Deterding, S., & Zaman, B. (2018). Uses and Gratifications of Initiating Use of Gamified Learning Platforms. Extended Abstracts, *CHI Conference on Human Factors in Computing Systems* (pp. 1-6). ACM.

Visual Attention Patterns on Dashboard during Learning with SQL-Tutor

Faiza Tahir, Antonija Mitrovic, Valerie Sotardi
University of Canterbury
faiza.tahir@pg.canterbury.ac.nz, {tanja.mitrovic, valerie.sotardi}@canterbury.ac.nz

ABSTRACT: In this study, we investigate how students use a dashboard in SQL-Tutor, an intelligent tutoring system that teaches the SQL query language. The dashboard is shown each time the student solves a problem, illustrating the student's progress both in graphical and a text-based form. The analyses of students' eye-tracking data show that students give much attention to the dashboard, especially when the dashboard is shown for the first time. In subsequent situations, students tend to focus on goal progress and the visualization of the student model. These results will help us to refine the dashboard in SQL-Tutor with important visualizations.

Keywords: Learning analytics, dashboard, intelligent tutoring system, eye tracking.

INTRODUCTION AND BACKGROUND

Learning analytics includes effective ways of measuring, analyzing and reporting learning outcomes students achieve in various learning environments. Learning analytics dashboards are tools which effectively visualize learning information to students and teachers (Verbert, Duval, Klerkx, Govaerts, & Santos, 2013). The purpose behind presenting this information is to make students aware of their learning progress, increase their self-monitoring and reflecting skills, and regulate learning strategies (Bodily et al., 2018), with the goal of improving learning. Many studies explore effects of visualizations used in dashboards through learner's eye-tracking data analysis. Eye tracking studies mainly focus on differences between novices and experts. For example, Barral et al. (2020) investigate how various types of students process information represented by charts and graphs, and provide evidence that adaptive guidance provided in the form of narrative-based charts visualization can benefit both novices and advanced students. Another eye tracking study compared nine different notational visualization systems, and reported that students preferred simple, easy and straightforward notations (Vatrapu, Reimann, Bull, & Johnson, 2013). Goldberg and Helfman (2011) revealed that students retrieved information easily from linear graphs organised either vertically or horizontally, rather than radial graphs. Another project combined eye-tracking data and students' interactions (log data) to analyze the perceptual speed, visual working memory, spatial memory, and visual scanning of students on different visualizations (Conati, Lallé, Rahman, & Toker, 2020). The authors suggested the importance of gaze data in user modelling and as a predictor of user interactions. All these studies investigated students' perceptions and preferences of different visualizations, and also the potential for predicting future performance. However, there is no research on how frequently dashboards should be shown to students to influence their skills and student interaction with dashboards in my opinion. In this contribution, we provide the initial results of an eye tracking study conducted to fill these gaps. The context of this study is SQL-Tutor, a mature intelligent tutoring system (ITS) for teaching problem solving in SQL (Mitrovic, 2003).

STUDY DESIGN AND PROCEDURE

We extended SQL-Tutor to support the three phases of Zimmerman's (1990) theory pf self-regulated learning, which targets goal setting, self-reflection, and monitoring skills of students. To support self-regulation, we added a dashboard to SQL-Tutor. The dashboard is presented to the student upon completion of a problem (Figure 1). The top section of the dashboard provides the overall information about the student's history, such as his/her pre-test score, current knowledge level, total time spent with SQL-Tutor, total problems solved with SQL-Tutor, the highest problem complexity, and the percentage of attempts on which the student required to see the complete solution. The second section of the dashboard visualizes the student's progress and the average class progress on each goal in the form of skill meters. If the student has achieved the current goal, the dashboard shows an appreciation message along with the next goal selection option; otherwise, it shows two strategies to select the next problem. The bottom section of the dashboard presents two graphs, which track the problems completed and time spent with SQL-Tutor per week. The last component is the open student model, i.e. the visualization of the student's knowledge in terms of six clauses of the SQL Select statement (select, from, where, order by, having, and group by).

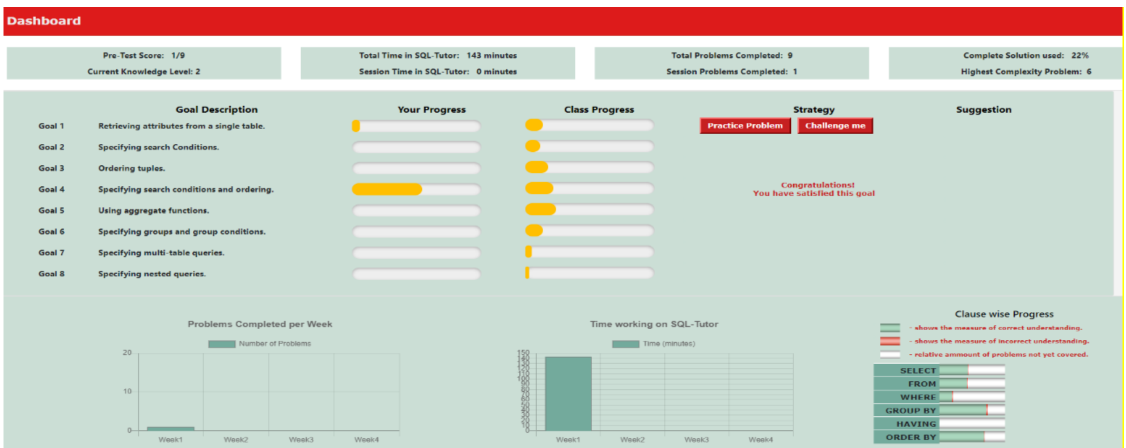


Figure 1. Dashboard of SQL-Tutor

The participants recruited for the study were undergraduate or postgraduate Computer Science students who had previous experience of problem solving in SQL-Tutor. At the beginning of the session, the participant was asked to sit in front of the Tobii eye tracker and calibration test was completed. After calibration, the participant worked with SQL-Tutor while their gaze data were recorded. The students were not required to solve a specific number of problems, but were required to work for 30 minutes.

PRELIMINARY INSIGHTS AND FUTURE WORK

To examine the visualization patterns and student attention on the dashboard, we identified parts of sessions where the dashboard was shown for the first time (phase 1) or last time (phase 3), as well as from the middle of the session (phase 2), when the student has completed five problems. Even though the students are still being recruited and data analysis is not completed, we can already deliver some preliminary insights. Students spent an average of 15s (sd = 16) looking at the dashboard in Phase 1, when it was first presented to them. However, this time declines on subsequent viewings (phases 2 and 3). The gaze patterns revealed that students looked at all three sections of the dashboard in phase 1. However, in phases 2 and 3, they only focused on the current goal progress, learning strategies, and the open learner model, as illustrated in Figure 2. Students looked at the graphical presentation of completed problems instead of text-based in their subsequent viewings, which shows their preferences. An interesting finding is that students did not pay attention to the class progress until they achieved a goal. Once they achieved a goal, they not only spent more time on the dashboard and looked at class progress but they also focused on their open learner model to assess their knowledge. The highest problem complexity measure on the dashboard did not receive attention. The possible explanation of this could be that the study was voluntary and students were not motivated to solve complex problems. These initial findings give us insights into student preferences for various elements of the dashboard and some indications on how frequently they want to see the dashboard. Further analysis of these results will help reveal the reasons for such behaviors and refine the dashboard.

Figure. 2 Aggregate eye gaze pattern after solving five problems

REFERENCES

Bodily, R., Kay, J., Aleven, V., Jivet, I., Davis, D., Xhakaj, F., & Verbert, K. (2018). *Open learner models and learning analytics dashboards: a systematic review.* Proceedings of the 8th International Conference on learning analytics and knowledge, pp. 41-50.

Barral, O., Lallé, S., & Conati, C. (2020). *Understanding the effectiveness of adaptive guidance for narrative visualization: a gaze-based analysis.* Proceedings of the 25th International Conference on Intelligent User Interfaces, pp. 1-9.

Conati, C., Lallé, S., Rahman, M. A., & Toker, D. (2020). Comparing and Combining Interaction Data and Eye-tracking Data for the Real-time Prediction of User Cognitive Abilities in Visualization Tasks. *ACM Transactions on Interactive Intelligent Systems, 10*(2), 1-41.

Goldberg, J., & Helfman, J. (2011). Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. *Information visualization, 10*(3), 182-195.

Mitrovic, A. (2003). An intelligent SQL tutor on the web. *International Journal of Artificial Intelligence in Education, 13*(2-4), 173-197.

Vatrapu, R., Reimann, P., Bull, S., & Johnson, M. (2013). *An eye-tracking study of notational, informational, and emotional aspects of learning analytics representations.* Proceedings of the 3rd International Conference on Learning Analytics and Knowledge, pp. 125-134.

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist, 57*(10), 1500-1509.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational psychologist, 25*(1), 3-17.

Do Gaming Experience And Prior Knowledge Matter When Learning With a Gamified ITS?

Faiza Tahir
Department of Computer science and Software engineering
University of Canterbury,
New Zealand
faiza.tahir@pg.canterbury.ac.nz

Antonija Mitrovic
Department of Computer science and Software engineering
University of Canterbury
New Zealand
tanja.mitrovic@canterbury.ac.nz

Valerie Sotardie
School of Edicational Studies and Leadership
University of Canterbury
New Zealand
valerie.sotardi@canterbury.ac.nz

***Abstract* — Gamification has gained much popularity, due to its positive effects on learner engagement and motivation in online learning environments. However, there is still insufficient understanding of factors, including personal traits, which affect learning, as well as studies focusing on learning behaviors which can be targeted by gamification. This paper investigates the causal effects of gamification on student learning outcomes, and the role of the students' background knowledge and prior gamification experience in the relationship. The context of our study is SQL-Tutor, an intelligent tutoring system. Although we found no evidence of improvement in learning outcomes of the gamified group, the low prior knowledge students who received badges had higher time-on-task, made more attempts on problems and received more hints during interaction with the system. We also found that students who had previous gamification experience spent more time on problem solving as compared to those who had no prior gamification experience.**

Keywords—gamification, badges, intelligent tutoring system, goals, previous gaming experience, learning outcomes

Introduction and Background

Gamification is defined as "the use of game design elements in non-gaming environments" [1]. It combines the enjoyment of games with the utility of a system to elevate user motivation [2]. The theory of gamified learning, proposed by Landers [3], identifies two theoretical paths through which gamification affects learning. In the first path, game elements influence learning behaviors which influence learning outcomes. In the second path, the influenced learning behavior moderates the relationship between learning content and learning outcomes.

There is very little research on gamification in ITSs. Abramovich and colleagues [4] awarded badges for participation and skill mastery in C2SN and reported that gamification increased topic interest but negatively influenced learning. Long and Aleven [5] awarded badges and stars while solving problems in Lynette, and reported partial positive effects on learning outcomes. However, both systems tutor middle school students, who usually are eager for awards and badges and none of them explored the influence of student's prior experience in gamification. This study presents the first empirical evidence of the gamification effects on university students in the context of SQL-Tutor, a mature ITS for teaching problem solving in Structured Query Language (SQL) [6]. Our motivation for this study is to explore the benefits of gamification on learning outcomes directly or by influencing some learning behaviours and the influence of prior knowledge and gamification experience on learning outcomes.

Study Design and Procedure

SQL-Tutor has been used in database courses at the University of Canterbury since 1998, as well as by students worldwide. The system supports problem solving in SQL by providing over 300 problems defined on thirteen databases. Students can select problems by themselves or select the most appropriate problem adaptively. While solving a problem, the student has to fill in the clauses of the *select* statement as required by the problem statement. The system provides six types of feedback upon submission, ranging from *simple feedback* to *complete solution*. The system tracks the student's actions and maintains the model of the student's knowledge.

We used the Landers theory of gamified learning as the framework for our study. We selected goals, assessment, and challenges as game elements, and implement them in the form of 13 badges, with each having a specific wining condition. Goals are implements as the wining condition of badges. Assessment is implemented as an optional quiz, and challenges are introduced by providing three complex problems per day as a daily challenge. Students received badges on either accomplishing a goal, solving daily challenges or by attempting the quiz. We divided the badges into three different levels: primary, classic, and elite (Table 1). Primary badges are given to students to capture their attention when they first interact with the system. For example, the *High flyer* badge is given for solving three consecutive problems. Classic badges emphasize practicing with the system regularly. The badges in this category are *Scholar*, for solving 5 problems for five consecutive days etc. The last level consists of Elite badges, which have the purpose of engaging learners with the system for a long period. The badges in this category are for example *Human dynamo* for solving at least five problems for ten days in a row and so on. Students can view the achieved and unaccomplished badges on the badge page which can be accessed via "view badges" button available on the problem-solving interface, the student model page, the introduction page of the system, and on the badge wining notification.

DEFINITION OF BADGES

Group	Badge	Condition
Primary	Go getter	Completing the first problem
	High flyer	3 problems in one session
	Achiever	5 problems in a day
	Activist	5 problems without complete solution
	Leader	problem with the "Group by" clause
Classic	Energy house	6 problems in a row
	Scholar	5 problems/day for 5 consecutive days
	Fireball	10 problems in one day
	Champion	First daily challenge
Elite	Genius	Attempting the quiz
	Human dynamo	5 problems/day for 10 days
	Einstein	5 daily challenges over 2 weeks
	Live-Wire	5 problems per day for 20 days

Out of 198 students enrolled in the course, 77 consented to participate in the study (25% female, 62% male, 13% others). At the start of the study, the students completed a short questionnaire, asking about their previous experience on gamification. They were then randomly allocated to the experimental condition where they interacted with the gamified version of the ITS (experimental = 42), or the non-gamified version of system (control = 35). When students logged into the ITS for the first time, they received a pre-test to estimate their prior knowledge. The pre-test consists of nine questions (1 mark for each question), which were the combination of true-false (2) and MCQs (7). Students could use the ITS whenever they wanted over the period of four weeks. At the end of the study, they received a post-test (similar to the pre-test) to assess their learning.

We made the following research hypotheses based on the results from literature and our own experience. We expect that the experimental group will be motivated by badges and learn more than the control group (H1). We expected that low prior knowledge students would engage more with the system in the gamification condition (H2). We are also interested to investigate the effects of previous gamification experience and expected that previous gamification experience would moderate the effects of badges in the study (H3).

RESULTS

The average scores on the pre/post-test were 58.73% (sd = 26.05) and 69.05% (sd = 25.9) respectively. Only 28 students completed the post-test, as it was not mandatory and administered two days before the major course test. There is another measure in the system called slevel, which is the current level of the student based on the number of completed problems. Slevel ranges from 1 to 9. The average slevel for all students were 3.56 (sd=1.66) at the end of their interaction with the ITS. The average days (Active days) students interacted with the system was 3.39 (sd=2.69) during four weeks, and the average time spent in the system was 260 minutes (sd = 243mins). During this time, students solved on average 37.47 problems (sd= 34.74). 66% of experimental and 57% of the control group students reported prior experience of gamification.

Table 2 presents the statistics for the two groups. There is no significant difference on the pre-test scores, showing both groups have similar levels of pre-existing knowledge. There were also no significant differences on the time and the number of attempted/solved problems. For calculating the normalised learning gain, we consider only those students who completed both tests (Exp = 17, Control = 11). There was no significant improvement between pre/post-test scores for the control (mean = .023, sd = 1.15) and experimental group (mean = -0.068, sd = 2.29) group. Therefore, H1 was not supported.

SUMMARY STATISTICS: MEAN (SD)

	Experimental (42)	Control (35)
Pre-test %	59.52 (24.02)	57.78 (28.62)
Time-on-task (min)	288.40 (302.02)	225.94 (143.44)
Active Days	3.33 (3.09)	3.46 (2.13)
Attempted problems	42.26 (42.75)	37.34 (26.94)
Solved Problems	39.33(40.99)	35.23 (25.72)
Student level	3.31 (1.62)	3.86 (1.68)
Post-test %	n = 17, 67.97 (26.32)	n = 11, 70.71 (26.42)

To evaluate H2, we divided students based on their pre-test scores. Those who scored more than the median (5) were labelled as High Prior Knowledge (HPK) and others as Low Prior Knowledge (LPK). We conducted Mann-Whitney U test to compare the groups. No differences were found on HPK students from the control and experimental groups.

STATISTICS FOR LOW PRIOR KNOWLEDGE GROUP: MEAN (SD)

Low Prior Knowledge	*Control (13)*	*Experimental (13)*
Pre-test %	26.4% (16.11)	30.7% (14.44)
*Time-on-task (min)	178.31 (98.25)	349.77 (230.48)
Active days	2.92 (1.66)	4.15 (3.39)
Problems Solved	23.38 (16.62)	40.92 (33.58)
*Attempts	103.77 (61.92)	203.85 (138.50)
*Hints	188.38 (109.16)	391 (261.05)

*p<0.05

Table 3 presents the results of LPK students in both groups. We found no significant differences on the pre/post-test scores, but LPK students from the experimental group spent significantly more time with system ($U = 134$, $p < .05$), seen more hints ($U = 127$, $p < .05$), and had more attempts on problems ($U = 124.5$, $p < .05$) than LPK students from the control group. These results indicate that gamification may influenced behaviours of LPK students, and motivated them to interact more with the system. This confirms our Hypothesis 2.

To investigate the relationship of students' previous experience in gamification (GE) with the badges and its subsequent effects on time-on-task and slevel, we developed the model shown in Fig. 1. We take badges as an independent variable, time-on-task as a mediator and slevel as the dependent variable in the model. We added GE (yes=1, No=0) as dichotomous moderating variable. To evaluate this model, first we regressed slevel on both time-on-task and badges. Results show that time-on-task is a significant predictor of slevel (β = .50, p = .01), but the number of badges is not a significant predictor (p = .06). However, the number of badges is a significant predictor of time-on-task (β = .64, p < .001).

To examine the effects of GE in the established mediating model, we investigate the moderating effects of GE on badges and time-on-task. The result shows the interaction term between badges and GE significantly and positively influences the time on task (t = 2.33, p = .02). That relationship is significant only when GE = 1 (t = 5.59, p < .001), but not when GE = 0. This indicate that badges may helped only those students who had prior gamification experience. As evident from Fig 2, those who had no prior gamification experience spent maximum 200 minutes (mean time-on-task) with the system. On the other hand, those with previous gamification experience spent on average 370 minutes with the system.

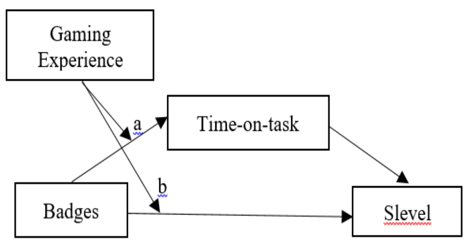


Figure 1. The moderated-mediation model, with gamification experience as a moderator

The total effect model shows the indirect effect of time-on-task is significant between badges and slevel [.0272, .4108]. This means that although GE increases time-on-task when combined with badges, in the absence of GE badges still have effect on student time-on-task, which influences their slevel. The index of moderated mediation, confirmed the moderated-mediation effect [.0864, .7859]. This supports our hypothesis 3.

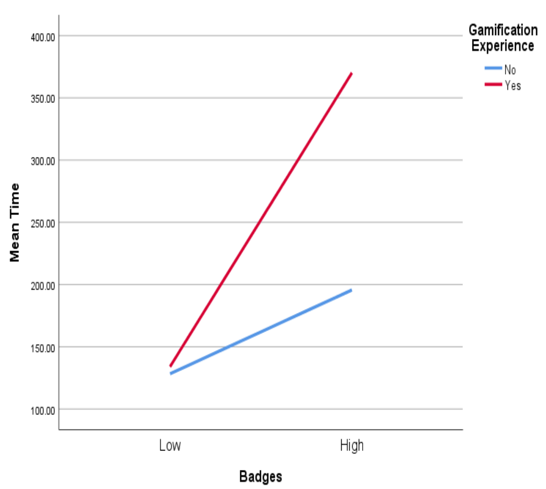


Figure 2. Relationship between badges and time-on-task when student had GE or no GE

CONCLUSIONS

In this study, we investigated the effects of gamification on learning and influence of demographic factors on gamification in context of an intelligent tutoring system. We introduced three types of badges in the system, each with different achieving criteria. The results show no difference between the experimental and control groups on student learning outcomes. However, the low prior knowledge students in the experimental group interacted with the ITSs for significantly higher time, made significantly more attempts on problems, and received significantly more hints than their counterparts from the control group. We also found that students who had previous experience with gamification interacted with the ITS for significantly longer time when they received badges, as compared to those who had no previous experience of gamification. Furthermore, time-on-task significantly mediates the relationship between badges and slevel. There are two major limitations of this study. First is the small sample size as not many students participated in the study. The possible reason could be because the use of system was completely voluntary. The second limitation is the students' access to the badge page. The possible explanation for this could be the design of badges, which failed to catch the attention of learners.

REFERENCES

Deterding, S., The lens of intrinsic skill atoms: A method for gameful design. *Human–Computer Interaction*, 2015. 30(3-4): p. 294-335.

Koivisto, J. and J. Hamari, The rise of motivational information systems: A review of gamification research. *International Journal of Information Management*, 2019. 45: p. 191-210.

Landers, R.N., Developing a theory of gamified learning: Linking serious games and gamification of learning. *Simulation & gaming, 2014*. 45(6): p. 752-768.

Abramovich, S., Schunn, C., & Higashi, R. M. (2013). Are badges useful in education? It depends upon the type of badge and expertise of learner. *Educational Technology Research and Development,* 61(2), 217-232.

Long, Y., & Aleven, V. (2014, June). Gamification of joint student/system control over problem selection in a linear equation tutor. In *International conference on intelligent tutoring systems* (pp. 378-387). Springer, Cham

Mitrovic, A. (2003). An intelligent SQL tutor on the web. *International Journal of Artificial Intelligence in Education*, *13*(2-4), 173-197.

Investigating effects of selecting challenging goals

Faiza Tahir, Antonija Mitrović[0000-0003-0936-0806] and Valerie Sotardi

University of Canterbury, Christchurch, New Zealand
faiza.tahir@pg.canterbury.ac.nz

Abstract. Goal setting is a vital component of self-regulated learning. Numerous studies show that selecting challenging goals has strong positive effects on performance. We investigate the effect of support for goal setting in SQL-Tutor. The experimental group had support for selecting challenging goals, while the control group students could select goals freely. The experimental group achieved the same learning outcomes as the control group, but by attempting and solving significantly fewer, but more complex problems. Causal modelling revealed that the experimental group students who selected more challenging goals were superior in problem solving. We also found a significant improvement in self-reported goal setting skills of the experimental group.

Keywords: Self-regulated learning, goal setting, intelligent tutoring system.

Introduction

Self-Regulated Learning (SRL) is defined as an "active, constructive process whereby learners set goals for their learning and then attempt to monitor, regulate, and control their cognition, motivation and behavior guided and constrained by their goals and the contextual features in the environment" (Zimmerman, 2011). The goal-setting theory illustrates that setting difficult goals lead to higher performance (Locke & Latham, 1990, 2019). Many studies show the benefits of goal-setting activities (Latham & Yukl, 1975; Locke & Latham, 2002), the power of self-set goals (Locke, 2001), influence of various strategies in goal attainment (Seijts at al., 2005; Masuda, 2015), and the effects of goal commitment (Landers, 2017). Zimmerman (2002) reported that students who set precise and actionable goals often reported higher self-awareness and had higher achievements. As mentioned in a meta-review of achievement (Collins, 2004), meeting a standard or goal is not enough; one should struggle for excellence. The goal-setting theory discussed the greater effects of task-specific over non-task related goals (Latham et al., 2012) and effects of selecting challenging goals (Latham et al., 2017).

Goal setting has been studied in various learning environments (Melis & Siekmann, 2004; Davis et al., 2016; Cicchinelli et al., 2018). In the context of AIED, relevant research connects students' goal-setting behavior with their motivation (Bernacki et al., 2013; Carr et al., 2013; Duffy et al,. 2015). Crystal Island (Rowe et al., 2011) asks students to solve a mystery by accomplishing eleven goals. Their results reveal that students who achieved more goals significantly improved their learning performance. In Meta-Tutor (Harley et al., 2017), four pedagogical agents support SRL via dialogs with the student. The agents determine the student's previous knowledge, and assist the student in selecting goals. Evaluation of Meta-Tutor revealed that students who collaborated more with agents learnt more. This paper discusses the effects of selecting challenging goals on learning in the context of SQL-Tutor (Mitrovic, 2003).

Study Design and Procedure

We enhanced SQL-Tutor by adding support for all three phases of the Zimmerman's model (2003), but in this paper we focus on the forethought phase only. SQL-Tutor contains over 300 problems, classified using 38 different problem templates (Mathews, 2006). A problem template covers a set of problems, which require the same problem-solving strategy. The 38 problem templates are grouped into eight high-level goals. The student is required to select a goal at the start of each session, and also after achieving a goal. The system always suggests challenging goals. The student is free to select one of the suggested goals, or any other goal.

We use a simple heuristic strategy to select a challenging goal for the student. At the start, students complete a pre-test, with scores ranging from 0 to 9. The initial goal is determined based on the student's pre-test score, while for the subsequent ones the system considers the student's current level (*slevel*). The student level ranges from 1 to 9, and it is determined dynamically, based on the student's success during problem solving (Mitrovic, 2003). For example, if the student scored 6 or more on the pre-test (i.e. the median score or higher), the challenging goal should be 8. The goal-setting page shows the number of problems per goal, and the number of problems the student has solved. The previously achieved goals are highlighted. If the student with a low pre-test score selects a very challenging goal, the system would suggest a less challenging one. To achieve a goal, the student needs to complete at least half of the relevant problems, or solve the five most complex problems.

The SRL instrument used in the study was adopted from (Kizilcec et al., 2017). Out of 24 questions, in this paper we only discuss the goal-setting subscale (4 questions). We also added five self-efficacy (SE) questions from the Motivated Strategies for Learning Questionnaire (Pintrich and De Groot, 1990). The survey used a five-point Likert scale, ranging from "Not at all true for me" (1) to "Very true for me" (5). The SRL and SE questions were included in Survey 1.

The participants, volunteers from the second-year database course at the University of Canterbury in 2020, were randomly allocated to the experimental (57) and control (42) groups. After providing informed consent, the participants completed the pre-test and Survey 1. The experimental group received support during goal setting, while the control group participants selected goals freely. After selecting a goal, students could choose any problem. The study lasted for four weeks. At the end of the study, students completed the post-test of similar structure and complexity as the pre-test, and completed Survey 2 (which was identical to Survey 1).

We hypothesized that the experimental group would achieve higher learning outcomes (H1). We formed a hypothesis for the experimental group: that selecting challenging goals would affect students' learning positively (H2). Finally, we expected that the support for goal setting would improve students' goal-setting skills (H3).

Results

We compared the pre/post-test scores of participants who completed both tests (Table 1). There is no significant difference on pre-test scores of the control (59.88%, sd = 28.82) and experimental groups (55.56%, sd = 29.18). The experimental group improved significantly from pre- to post-test (W = 298, p = .03), but the control group students did not (p = .74). Comparing normalized gains revealed no significant difference. These results partially support hypothesis H1. The control group attempted/completed significantly more problems (Table 1). The experimental group completed significantly more complex problems. These findings show that experimental group achieved higher learning gains by completing fewer but more complex problems.

- **Table 1.** Summary of major statistics: mean (sd)

	Control (42)	Experimental (57)	Significance
Attempted Problems	92.98 (61.86)	57.46 (41.33)	U = 783, p = .003
Completed Problems	91.86 (61.33)	56.44 (41.09)	U = 783, p = .003
Problem Complexity	2.92 (0.96)	3.32 (1.08)	U = 1465.5, p =.057
Time (min)	360.19 (335.33)	296.71 (233.22)	p = .58

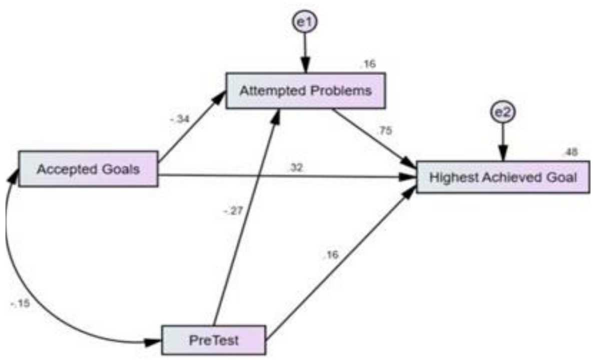
We divided the experimental group post-hoc into three subgroups (Table 2). Fourteen students always accepted the suggested goals (SG), 18 students worked on the goals in the sequential order (SEQ), while the remaining 25 students used a mixed strategy (Mix). We found no significant differences between the subgroups on the pre-/post-test scores and time, but there were statistically significant differences on the

Table 2. Summary statistics for the three subgroups: mean (sd)

	SEQ (18)	Mix (25)	SG (14)
Pre-test %	62.97 (27.75)	64.46 (24.01)	61.11 (33.98)
Post-Test %	n=9, 64.21 (31.81)	n=12, 77.78 (28.55)	n=8, 69.45 (21.23)
Time (min)	346.17 (290.49)	283.36 (163.41)	257.0 (263.09)
Attempted goals	6.39 (2.62)	7.04 (1.14)	5.00 (2.18)
Achieved goals	4.72 (2.54)	3.60 (2.43)	1.64 (2.34)

number of attempted goals (H = 8.12, p = .017), achieved goals (H = 10.13, p = .006), the number of attempted/solved problems (H = 13.88, p = .001 and H = 14.41, p = .001 respectively), and problem complexity (H = 12.20, p = .002). The post-hoc analyses revealed no significant differences between the SEQ and Mix groups. The SG subgroup attempted significantly fewer goals in comparison to the SEQ (U = 55, p = .006) and Mix groups (U = 94, p = .016), and achieved significantly fewer goals in comparison to the Mix group (U = 77, p = .003). The SG group also attempted/solved significantly fewer problems in comparison to the SEQ (U = 44, p = .002 in both cases) and Mix groups (U = 74.5, p = .003 and U = 71, p = .002 respectively). However, the average problem complexity of solved problems for the SG group was significantly higher in comparison to the SEQ (U = 40.5, p = .001) and Mix (U = 77.5, p = .004) groups.

We analyzed the data using the structural equation model (Figure 1). We hypothesized that the pre-test score and the number of attempted problems will have a positive effect on learning. The variable labelled “Accepted goals” shows how many times students accepted the suggested goals. Because not all students completed the post-test, we use a different measure of learning: the highest achieved goal (HAG). All path coefficients are significant at p < .05 except PreTest -> HAG, and the covariance between Accepted Goals and PreTest. There is a significant negative effect of Accepted goals on Attempted problems. These findings suggest that (1) students who accepted system goals tended to achieve higher goals (the confidence interval [.1345, .7074] does not include zero), and (2) students who accepted suggested goals, despite of attempting fewer problems, achieved higher goals (the confidence interval [-.5903, -.1133]). Students with lower pre-test scores achieved higher goals when they accepted system suggestion. These findings support H2.



To test hypothesis H3, we compared the scores from the two surveys (Table 3). No differences exist at the time of Survey 1 on goal setting and self-efficacy (SE). The goal-setting scores of the experimental group improved significantly (z = -1.93, p = .05), but not in the control group. There is a significant difference (z= -2.97, p < .005) on the goal-setting scores on Survey 2. The SE scores differed both as a function of group and time. At Survey 1, the experimental group had lower SE, but they increased at Survey 2 (z = -1.57, p = .1) whereas the SE scores decreased for the control group (z = -1.86, p = .06). These findings suggest that (a) students who complete the tasks in the absence of the intervention reported lower SE over time; and, (b) the goal-setting intervention may lead to considerable gains in SE, *especially* for students who started with less confidence. Although it is important to investigate further these trends in future research, these findings confirm our Hypothesis 3.

Table 3. Goal setting and self-efficacy scores: mean (sd)

	Goal Setting		**Self-Efficacy**	
	Exper. (21)	**Control (14)**	**Exper. (21)**	**Control (14)**
Survey 1	3.56 (0.63)	3.39 (0.64)	3.38 (0.65)	3.5 (0.66)
Survey 2	3.95 (0.65)	3.28 (0.65)	3.74 (0.65)	2.98 (0.67)

Our findings highlight the effects of setting challenging goals under realistic conditions, in a study that lasted four weeks. The limitations of our study are the small sample size and the low completion rates for Survey 2 and post-test. The results are in line with the goal-setting theory. In future work, we will investigate the effects of the intervention on the monitoring and self-reflection SRL phases.

References

Bernacki, M. L., Nokes-Malach, T. J., & Aleven, V. (2013). Fine-grained assessment of motivation over long periods of learning with an intelligent tutoring system: Methodology, advantages, and preliminary results. In *International handbook of metacognition and learning technologies* (pp. 629-644): Springer.

Carr, A., Luckin, R., Yuill, N., & Avramides, K. (2013). How mastery and performance goals influence learners’ metacognitive help-seeking behaviours when using Ecolab II. In *International handbook of metacognition and learning technologies* (pp. 659-668): Springer.

Cicchinelli, A., Veas, E., Pardo, A., Pammer-Schindler, V., Fessl, A., Barreiros, C., & Lindstädt, S. (2018). *Finding traces of self-regulated learning in activity streams.* Proc. 8th Int. Conf. Learning Analytics and Knowledge (pp. 191-200).

Davis, D., Chen, G., Jivet, I., Hauff, C., & Houben, G.-J. (2016). Encouraging Metacogniion & Self-Regulation in MOOCs through Increased Learner Feedback. Learning Analytics and Knowledge conference (pp. 17-22).

Duffy, M. C., & Azevedo, R. (2015). Motivation matters: Interactions between achievement goals and agent scaffolding for self-regulated learning within an intelligent tutoring system. *Computers in Human Behavior, 52*, 338-348.

Harley, J. M., Taub, M., Azevedo, R., & Bouchet, F. (2017). Let's Set Up Some Subgoals: Understanding Human-Pedagogical Agent Collaborations and Their Implications for Learning and Prompt and Feedback Compliance. *IEEE Transactions on Learning Technologies*, 11(1), 54-66.

Hayes, A. F. (2017). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. Guilford publications.

Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers & Education, 104*, 18-33.

Landers, R. N., Bauer, K. N., & Callan, R. C. (2017). Gamification of task performance with leaderboards: A goal setting experiment. *Computers in Human Behavior*, 71, 508-515.

Latham, G. P., & Piccolo, R. F. (2012). The effect of context-specific versus nonspecific subconscious goals on employee performance. *Human Resource Management, 51*(4), 511-523.

Latham, G. P., & Yukl, G. A. (1975). Assigned versus participative goal setting with educated and uneducated woods workers. *Journal of applied psychology, 60*(3), 299.

Locke, E. A. (2001). *Self-set goals and self-efficacy as mediators of incentives and personality.* In M. Erez, U. Kleinbeck, & H. Thierry (Eds.), *Work motivation in the context of a globalizing economy* (p. 13–26). Lawrence Erlbaum Associates Publishers.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting & task performance*: Prentice-Hall, Inc.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American psychologist, 57*(9), 705.

Locke, E. A., & Latham, G. P. (2019). The development of goal setting theory: A half century retrospective. *Motivation Science, 5*(2), 93.

Masuda, A. D., Locke, E. A., & Williams, K. J. (2015). The effects of simultaneous learning and performance goals on performance: An inductive exploration. *Journal of Cognitive Psychology, 27*(1), 37-52.

Mathews, M. (2006). Investigating the effectiveness of problem templates on learning in intelligent tutoring systems. Honours report, Department of Computer Science, University of Canterbury.

Melis, E., & Siekmann, J. (2004). Activemath: An intelligent tutoring system for mathematics. In *Int. Conf. Artificial Intelligence and Soft Computing* (pp. 91-101). Springer, Berlin, Heidelberg.

Mitrovic, A. (2003). An intelligent SQL tutor on the web. *Artificial Intelligence in Education, 13*(2-4), 173-197.

Pintrich, P. R., De Groot, E. V. (1990) Motivational and self-regulated learning components of classroom academic performance. *Journal of educational psychology*, *82*(1), 33.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In *Handbook of self-regulation* (pp. 451-502): Elsevier.

Rowe, J. P., Shores, L. R., Mott, B. W., & Lester, J. C. (2011). Integrating learning, problem solving, and engagement in narrative-centered learning environments. *Artificial Intelligence in Education, 21*(1-2), 115-133.

Seijts, G. H., & Latham, G. P. (2005). Learning versus performance goals: When should each be used?. *Academy of Management Perspectives, 19(1), 124-131.*

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into practice, 41*(2), 64-70.

Zimmerman, B.J., & Campillo, M. (2004) *Motivating self-regulated problem solvers.* The psychology of problem solving, 233-262.

Zimmerman, B. J., & Schunk, D. H. (2011). *Handbook of self-regulation of learning and performance*: Routledge/Taylor & Francis Group.