

RESEARCH

Open Access



Investigating the causal relationships between badges and learning outcomes in SQL-Tutor

Faiza Tahir^{1*} , Antonija Mitrovic¹ and Valerie Sotardi²

*Correspondence:

faiza.tahir@pg.canterbury.ac.nz

¹ Intelligent Computer Tutoring Group, Department of Computer Science and Software Engineering, University of Canterbury, Private Bag 4800, Christchurch 8041, New Zealand

Full list of author information is available at the end of the article

Abstract

The practice of adding game elements to non-gaming educational environments has gained much popularity. Gamification has been shown in some studies to enhance engagement, motivation and learning outcomes in technology-supported learning environments. Although gamification research has matured, there are some shortcomings such as inconsistency in applying gamification theories and frameworks and evaluating multiple game mechanics simultaneously. Moreover, there is little research on applying gamification to Intelligent Tutoring Systems (ITS). This paper investigates the causal effects of gamification on learning in SQL-Tutor, a mature ITS teaching students how to phrase queries in SQL. Having conducted a study under realistic conditions, we present a quantitative analysis of the performance of 77 undergraduate students enrolled in a database course. There are three main findings of our study: (1) gamification affects student learning by mediating the time-on-task; (2) students' background knowledge does not influence time-on-task unless students achieve badges; and (3) students' interest in topic (motivational construct) moderates the relationship between badges and time-on-task, but does not improve learning outcomes directly.

Keywords: Gamification, Goals, Assessment, Challenges, Badges, Learning behavior, Time-on-task, Causal analysis

Introduction

Engagement is a crucial ingredient for learning. One strategy to increase motivation in technology-supported learning environments is gamification (Deterding et al., 2011), i.e. the use of gaming elements such as leader boards, points, badges, and other virtual achievements common in games. These virtual achievements are not necessarily connected to tangible rewards; they are meant to increase user engagement and motivation to use those applications. For example, PeerWise (Denny et al., 2018) awards virtual badges to students for writing or answering questions. Leader boards are often used in applications where social activities are important, like comparing the performance of students in a course (Huang et al., 2020).

The term gamification was first used a decade ago (Deterding et al., 2011) and has gained much popularity since. Gamification was found to be effective in many projects

in maintaining user engagement by encouraging their actions and fostering the quality and productivity of those actions (Hamari, 2013). However, gamification does not always yield positive results. In a few cases, gamification may go unnoticed by users, or even have negative effects, which were completely unintended (Diefenbach & Müssig, 2019). Moreover, despite the growing number of gamified educational environments, 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 itself does not improve learning, gamification would not help. On the other hand, if an educational system is highly effective, gamification may not provide additional benefits. 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 behavior.

Intelligent tutoring systems (ITSs) have a long history of proven results in education (Anderson et al., 1995; Mitrovic, 2012; van Lehn, 2006). 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 & Alevan, 2013), and supporting affective states of learners (Tahir et al., 2019). This study aims to explore the effects of gamification on students' engagement and motivation while using SQL-Tutor (Mitrovic, 1998, 2003), a mature ITS that teaches the standard query language (SQL). The effectiveness of SQL-Tutor has been proven in multiple studies (Mitrovic, 2012; Mitrovic & Ohlsson, 1999).

We start by reviewing the literature on gamification and its effects. Section [Research questions and hypotheses](#) presents our approach to gamifying SQL-Tutor, while Sect. [Experimental procedure](#) discusses the experiment design. We then present our findings in Sect. [Results](#), and finally, the conclusion and limitations of the current work.

Related work

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 also subject to various technical and social concerns (Sanford & Madill, 2006; Susi et al., 2007). The idea of gamification is to focus on learning, not on play; this means separating the gamification from playfulness.

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 developing standalone games (Dicheva et al., 2015; Landers et al., 2017). Standalone games are resource intensive as they must deal with a lot of technical and cultural issues, various perceptions of game narratives, and so on. Gamification does not require the detailed game implementation knowledge and few man hours which makes it less cost intensive than standalone educational games. Gamification aims to increase motivation, by combining the efficiency of utilitarian systems and enjoyment of hedonic systems (Koivisto & Hamari, 2019).

The ease of applying gamification and its potential benefits are reasons for its popularity. Three meta-analysis studies (Alhammad & Moreno, 2018; Hamari et al., 2014;

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 the gamification research: (1) focus on behavioral 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 identify several major methodological problems with reported studies, which include small sample sizes, short durations with no control conditions, using several gamification mechanics, not reporting negative or neutral effects, and reliance on self-reporting instruments. Pereira et al. (2020) also point out the lack of studies focusing on the specific effects of individual gaming elements.

The theory of gamified learning (Landers, 2014; Landers et al., 2017) specifies that gamification has an effect on learning by influencing learners' behaviors or attitudes, via two theoretical paths. Gamification elements may influence a particular learning behavior, which in turn directly influences learning outcomes; thus, the learning behavior acts as a mediator. In other situations, the influence of students' behaviors or attitudes changes the effectiveness of instructional content—that is, the learning behavior moderates the relationship between the content and learning outcomes. In a study using leader boards and time-on-task as the mediating behavior, Landers and Landers (2014) found a 27% improvement in learning in the experimental group in comparison with the control group. Helmeffalk (2019) proposes another gamification framework (M-PM-O) having a similar path between game mechanics and outcomes mediated by psychological processes (flow, enjoyment, engagement, motivation, etc.). The author suggests that different moderators (such as demography, time, space, or platform) may affect the mediating relationship and highlights the importance of evaluating the effect of a single game mechanic on a particular learning outcome under the influence of a particular mediator (psychological process).

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 behavioral changes. The proposed framework (RECIPE) elaborates the features of meaningful gamification which are: it should provide a narrative as context (Exposition), allow players (students) to 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 behavioral change to interact with the environment purposefully.

Gamification has been applied to various technology-enhanced 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 to engage with the system for a longer time once they start 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 have a significant effect on psychological behaviors (competence, autonomy, and relatedness), which in turn increase 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, some cases where the use of the system is compulsory or it is a part of the course, such as learning management systems, 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 behavior. The leader board was found to be the best motivational element. Denny et al. (2018) investigate the effects of badges on learning outcomes by mediating self-testing behavior 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 on undergraduate and postgraduate students (Legaki et al., 2020) analyzed the effects of challenge-based gamification on learning performance. The gamified system targeted the playing behavior of students, in comparison with the reading-only group and the no-intervention groups. The authors reported a 34.75% improvement in students' learning performance for the gamified group. Besides the gamified system being part of the course, another major contributing factor of these successes is targeting students' learning behaviors which caused performance and learning.

However, not all studies of gamification in learning revealed positive results, with some even revealing negative effects of gamification on motivation and learning. Haaranen et al. (2014) investigated the effects of badges in a data structures and algorithm course. The badges were awarded for time management, early submission and successfully 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 negative effects include loss of performance, undesired behavior, indifference, and declining motivation (Hanus & Fox, 2015; Toda et al., 2017).

Although there are numerous papers investigating the effect of gamification in online educational systems, there is little research focusing on the gamification of ITSs. Long and Aleven (2014) investigated the effects of two gamification features in Lynette (the equation solving ITS): re-practicing of 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-practiced previously completed problems but received no rewards on their performance (Long & Aleven, 2014). 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 practicing new problems. As the result, the gamified group showed higher learning outcomes compared to the non-gamified group and improved problem-selection strategy. In another study, Abramovich et al.

(2013) studied the CS2N intelligent tutoring system, in which badges were awarded for skills mastery or for continued use of the system. They revealed that despite an increase in students' interest in the topic, and a decrease in counter-productive behaviors, badges influenced learning negatively. Moreover, the authors highlighted the interplay of motivation for students with different background knowledge and attributed badge design for poor motivation in students. González et al. (2014) extend the architecture of EMATICS, an ITS that teaches basic mathematical operations to children, by adding the functionality related to the gamification mechanics to the pedagogical module. Authors can define activities using points, badges, and leader boards. Dermeval and Bittencourt (2020) emphasize the importance of involving teachers in designing gamified ITSs. They propose a gamification domain ontology, identify combinations of game elements for achieving particular behaviors, and provide an authoring tool for teachers to customize gamification.

Most of the gamified systems explored the effects on student engagement and motivation as mentioned in studies above and also in (Hamari et al., 2014). However, motivation investigated in those studies was measured either by the number of awards a student has achieved, or the effort to achieve those rewards (number of problems attempted, number of edits, etc.), or their opinions about the future use of the system. The interplay of various motivational aspects was neglected in the research. For example, self-efficacy, mentioned in social cognition theory, is a powerful tool to influence students' motivation, achievement, and self-regulated learning (Schunk & DiBenedetto, 2020). Bandura and Schunk (1981) reported that self-efficacious individuals tend to work harder and persist longer in difficult and challenging tasks. At the start of the task, students' self-efficacy is based on their prior experience. As they are working, the attitude toward the goal, information processing, feedback from the teacher on the effort and rewards give them signals on how they are learning, which in turn help them to assess their efficacy (Schunk, 1991). Rewards are considered to be a mechanism to increase self-efficacy if they are linked with student 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, stimulate intrinsic motivation and perceived competence (Houffort et al., 2002). The theory mentioned that an increase or decrease in intrinsic motivation also depends on increase and decrease on one's perception and feelings of competence. Topic-interest is the least explored motivational aspect in the literature. This is the interest developed when an individual is introduced to a topic (Ainley et al., 1999) and influences student's affect responses related to their persistence and learning (Ainley et al., 2002). The four-phase model of interest development shows a developed interest and provided facilitation increased student's self-efficacy (Hidi & Renninger, 2006).

From this brief literature review, we can infer that these aspects of motivation are strongly related to learning, and rewards might help to strengthen the relationship between motivation and learning. These motivational aspects are linked and complement each other (Mayer, 1998). In our study, we explore the influence of these aspects in the context of gamified SQL-Tutor and determine whether one of these is sufficient or what unique contribution each of these aspects has on student motivation. Besides

motivational aspects, we identified other methodological gaps via literature review. Various gamification research inconsistently considers students' behavior and gamification frameworks. Most of the studies that reported positive results neither follow the gamified theory or a specific framework nor report design guidelines. Nearly all the studies applied multiple game mechanics to influence more than one learning behavior. Therefore, the understanding of which game mechanic is suitable for a particular learning behavior remains unclear. Lack of empirical studies and especially controlled experiments is another big reason to remain inconclusive about the effects of gamification. Our study attempts to fill these gaps.

Research questions and hypotheses

We made the following research questions and related hypotheses, based on the results from literature (e.g., Landers & Landers, 2014), and our own experience.

Research Question 1: What are the effects of gamification on learning? We expect that the experimental group students, who will receive badges, 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 learning outcomes (H2). We expect that badges will have an indirect effect on learning outcomes, by influencing time-on-task (H3).

Research Question 2: 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 low background knowledge. We expect that prior knowledge will affect the time student spend in SQL-Tutor and that badges will moderate this relationship (H4).

Research Question 3: What is the effect of gamification on student motivation? As discussed previously, some studies found that gamification increases motivation. Therefore, we expect the experimental group students would report increased levels of self-efficacy, perceived competence, and topic-interest after the study (H5), and that higher motivation will lead to higher learning outcomes (H6).

Gamifying SQL-Tutor

SQL-Tutor has been in regular use in database courses at the University of Canterbury since 1998 and has also been used by numerous students worldwide. The system teaches problem solving in SQL by providing 300+ problems from 10 databases. There are four different problem-selection strategies offered by SQL-Tutor. The student can ask for the next problem on the current database, select any problem he/she likes from the full list of problems, or from a list of problems focusing on a particular clause of the SQL *Select* statement. The student can also ask the system to select a problem adaptively, based on the student model. The problem-solving interface allows students to complete problems by specifying the query, as illustrated in Fig. 1. The interface also provides information about the current database (the bottom pane). Whenever the student submits a solution, SQL-Tutor provides feedback, which can be on six different levels. *Simple Feedback* is the lowest level, indicating whether the submitted solution is correct or not. The next level of feedback is the *Error flag*, which specifies the part of the solution that is wrong. The next level is *Hint*, which

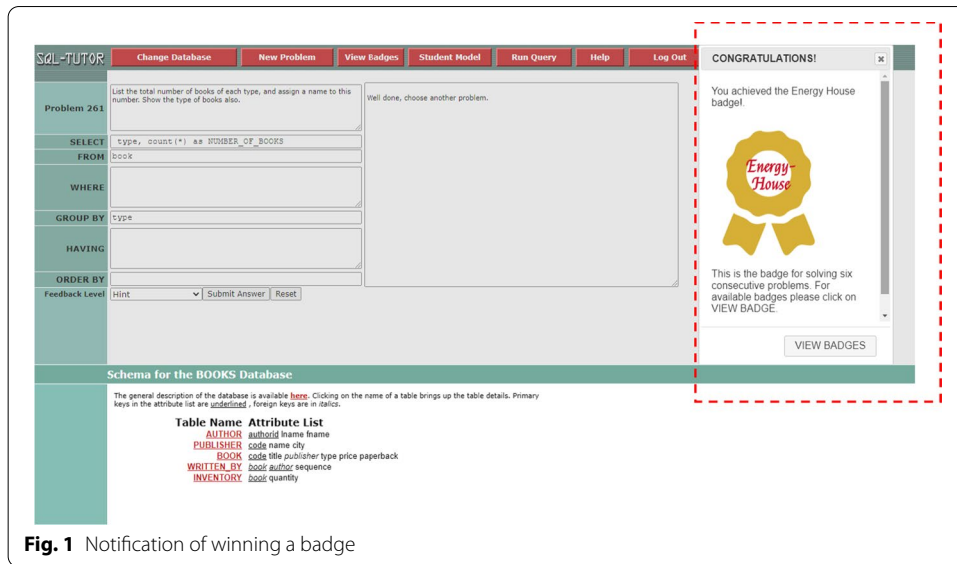


Fig. 1 Notification of winning a badge

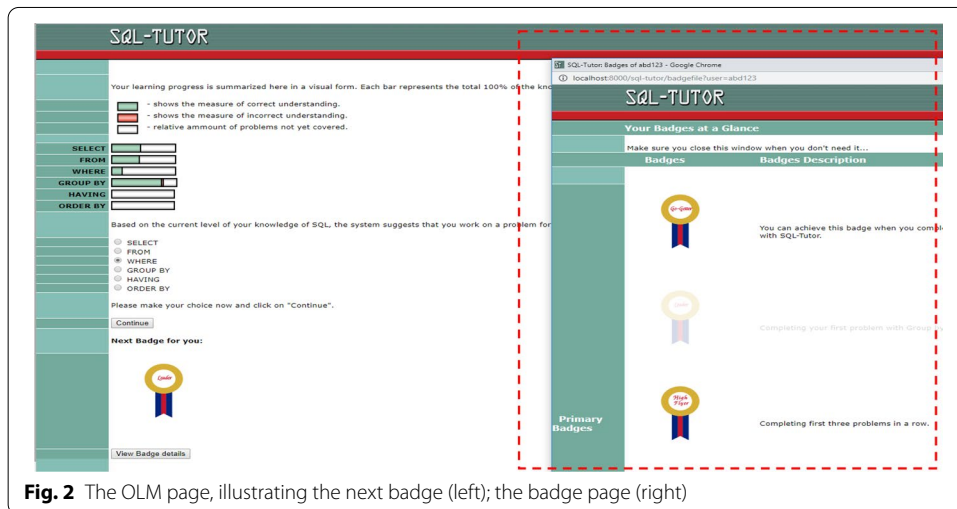


Fig. 2 The OLM page, illustrating the next badge (left); the badge page (right)

provides help on one error in the solution; it specifies the nature of the error and provides information about the domain principle that is violated by the student's solution. These three initial levels of feedback are provided by the system automatically on subsequent submissions. The system also provides three higher levels of feedback, which are only available on request. The *Partial solution* provides the correct version of the erroneous part of the solution identified by the Error Flag feedback. The *List all errors* message provides hints on all identified errors. The last feedback level is *Complete Solution*, which provides the correct solution for the problem. The student can select any feedback level anytime during problem solving. The system tracks the student's progress in the form of a student model that indicates the estimate of correctly learnt knowledge as well as an estimate of what is yet to learn. A visualization of the student model is available to the student on request as seen in Fig. 2 (left). SQL-Tutor

provides an Open Learner Model (OLM), in the form of skill meters for each clause of the Select statement.

Our approach to gamifying SQL-Tutor (Mitrovic, 2003) is based on the theory of gamified learning (Landers et al., 2017). Landers identified nine most often used categories of game elements: action language, assessment, conflict/challenge, control, environment, game fiction, human interaction, immersion, and rules/goals. We selected three of these categories: *goals*, *challenges*, and *assessment*. Goals and challenges belong to the goal-setting theory, and assessment provides the testing effects. There are no hard and fast rules for selecting game elements, and the suitability of a particular game element in a specific environment is still under research. In our study, *action language*, *environment*, and *immersion* might not be suitable because SQL-Tutor already provides an established problem-solving environment based on specific domain knowledge. Additionally, SQL-Tutor provides complete freedom to students on problem solving; therefore, the *control* game element is not appropriate. Similarly, human interaction and game fiction cannot be selected because they are more enticing for young students than undergraduate students and contrasting the nature of the ITS.

In this study, the selection of the game elements was based on (a) its suitability to the system in context (SQL-Tutor), and (b) influence on targeted learning behavior. As mentioned above, SQL-Tutor teaches problem solving in SQL by providing numerous practice problems which are arranged according to their complexity. Learners need to practice those problems regularly (i.e., time-on-task) to improve their expertise in SQL. To ensure this regular practice, the *rules/goals* category was selected under the light of goal-setting theory (Locke & Latham, 1990, 2013). According to this theory, learners must set SMART (specific, measurable, achievable, realistic, and time-bound) goals for themselves and track their progress to achieve those goals. We 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 goal-setting theory (Locke & Latham, 2019) also suggests that difficult goals lead to higher performance. As mentioned in a review (Collins et al., 2004), meeting a goal is not enough; one should struggle for excellence. The previous study (Tahir et al., 2019) on SQL-Tutor revealed that students attempted fewer high complexity problems and many low complexity problems. This indicates that learners not only need to practice problem solving regularly, but they should attempt and solve more complex problems. This provides the basis to select our next game element—*challenges*. Challenges grow competition in students either in the form of standing in the class or achievement of the skill. Munshi et al., (2018a, 2018b) show that students 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 a challenge; the difference between challenges and goals lies in the complex and hard to achieve nature of challenges. Goals do not consider specific problems or students' current knowledge. However, challenges consist of problems that are higher in complexity than the ones student has solved.

As suggested by the goal-setting theory, setting challenging goals improves students' learning outcomes. However, only setting goals does not lead to higher learning

outcomes: learners must strive toward achieving those goals. Self-assessment (also known as self-testing) is one of the key features in evaluating and tracking the progress toward achieving goals. SQL-Tutor provides self-assessment in the form of pre-/posttests administered before and after the study. These self-assessment tests provide an opportunity for students to reflect on their progress and motivate them to strive for better performance. Based on the vital role of assessments in problem solving, the next game element selected is assessment. Assessment in this study is implemented in the form of a quiz that has the same nature and structure as pre-/posttests, and administered at the end of the second week. The reason to select the quiz is that the students have already attempted the pretest at the beginning of the study and are familiar with the notations. The quiz provides them another opportunity to assess their knowledge and spend more time on problem solving.

We implemented goals, assessment, and challenges in SQL-Tutor via different types of badges, presented in Table 1. The goal-setting behavior is supported by selecting daily and weekly goals stated as winning criteria for badges. The self-testing behavior is addressed by providing a quiz. Challenges are implemented via several badges, and as daily challenges, which consist of complex unsolved problems. We assume that all these game elements would influence time-on-task, which has been shown in many studies to influence learning outcomes (Landers & Landers, 2014; Denny et al., 2018).

There are three groups of badges: 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 discourages the use of 'complete solution.' This 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 five consecutive days, and solving daily challenges. 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. In this category, badges are awarded when the student completes five problems every day for ten

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
	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

days or solves five daily challenges in two weeks. The last badge is awarded to those extraordinary students who completed five problems every day, for 20 consecutive days.

When the student fulfills the condition for a badge, he/she receives the notification about that badge immediately, as shown in Fig. 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 (Fig. 2). For the study, we modified the OLM page to show the next badge the student could achieve, as shown in Fig. 2. Daily challenges are presented to students once they achieve all primary badges, as shown in Fig. 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 when the student completes five daily challenges over 2 weeks respectively.

We also developed a quiz, consisting of seven multiple-choice questions and two true/false questions of the same type of questions used in the pre-/posttest. The *Genius* badge is awarded for attempting the quiz, independently of the score achieved. When the student completes a quiz, the scores are 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.

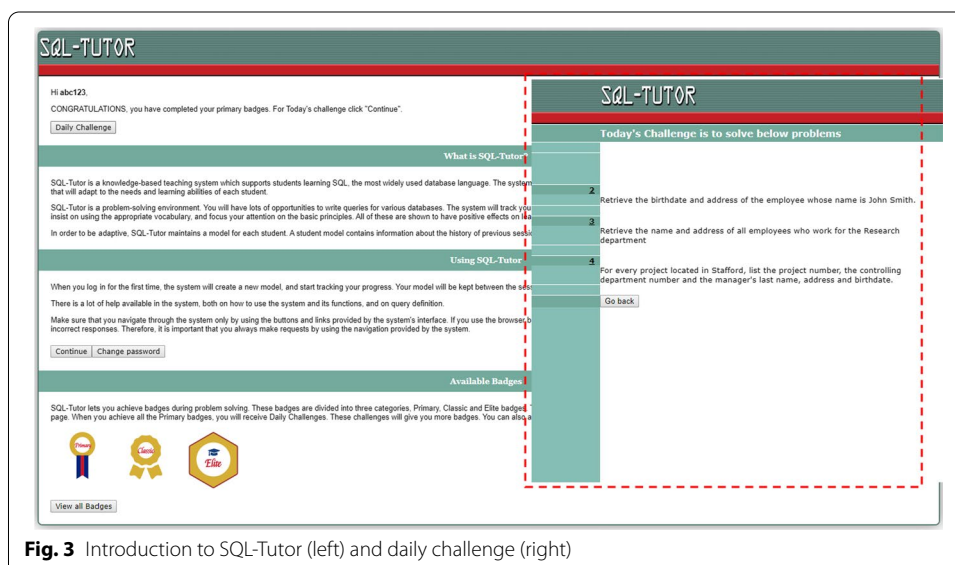


Fig. 3 Introduction to SQL-Tutor (left) and daily challenge (right)

Experimental 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 laboratory 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 laboratory session. The use of SQL-Tutor was voluntary; the students did not receive any course credit for solving problems in SQL-Tutor. 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 pretest and Survey 1. The survey contained questions on their previous experiences with gamification, as well as questions on self-efficacy and perceived competence related to SQL adapted 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'). Perceived confidence was measured by three items: '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.' We 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 our data (Cronbach alpha = 0.88). Survey 1 also contained seven items on topic-interest, adapted from van Harsel et al. (2019), in which we referred to 'SQL Querying' instead of the original context. The reliability of these items was good, with Cronbach alpha = 0.83.

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-/posttest and the quiz were of similar complexity; each contained seven multiple-choice questions and two true/false questions (worth one mark each).

The posttest and Survey 2 (same as Survey 1) were administered at the end of the fourth week. A major piece of the course assessment was the laboratory test focusing on SQL, worth 20% of the final grade, administered two days after the posttest. After the laboratory test, the students were invited to complete Survey 3. There were two versions of this survey. The version of Survey 3 for the experimental group contained 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 1 ('strongly disagree') to 5 ('strongly agree').

Results

Table 2 presents the summary statistics of the study. The average score on the pretest was 58.73%. 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,

Table 2 Summary statistics of SQL-Tutor usage

	Mean	SD
Pretest %	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
Posttest %, $n = 28$	69.05	25.9
Laboratory test%	60.83	17.07

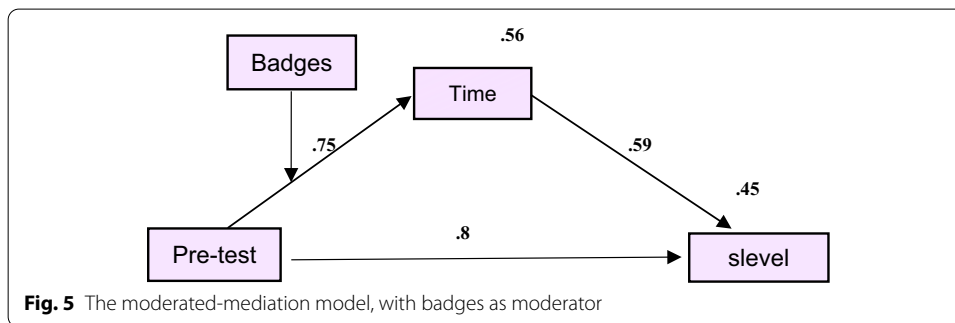
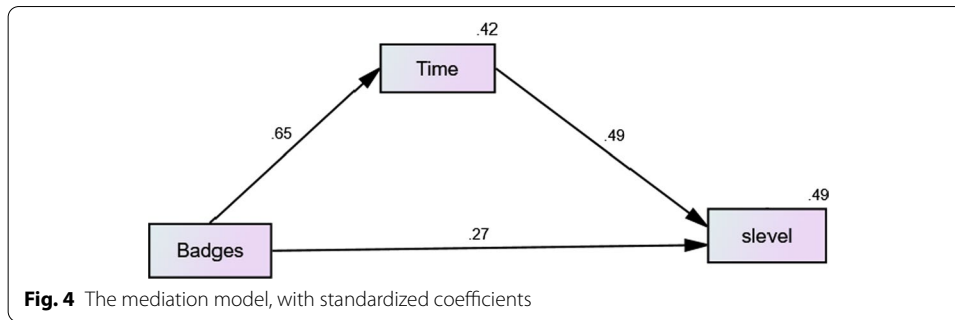
Table 3 Summary statistics of SQL-Tutor usage for experimental and control groups: mean (Sd)

	Experimental (42)	Control (35)
Pretest %	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)
Posttest %	$n = 17, 67.97 (26.32)$	$n = 11, 70.71 (26.42)$
Laboratory test %	60.43 (16.49)	61.31 (17.97)

max = 1441, 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 posttest; we believe the reason for the low completion rate was that the posttest was not mandatory. In addition, the posttest was given to the students only two days before the laboratory test. The average score on the posttest was 69.05% and for the laboratory test, it was 60.83%. In addition to defining queries, which students practiced in SQL-Tutor, the laboratory test covered other SQL topics, and therefore, the laboratory test cannot be considered as the direct learning outcome. For those reasons, we use the student level (*slevel*) 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 *slevel* achieved = 1, max *slevel* achieved = 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.

RQ 1: What is the effect of gamification on student learning?

Table 3 presents statistics for the two groups. We found (using the Shapiro–Wilk test) that data was not normally distributed. Hence, the Mann–Whitney U was used to compare the means of two groups. There was no significant difference in the pretest 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 with the control group, although the differences are not significant. Therefore, our hypothesis H1 is not supported. There was also no significant difference between the groups on the number of active days, student levels, the posttest, and laboratory test scores.

To evaluate H2, we regressed the student level on time-on-task. The time-on-task strongly predicts the student level ($\beta=0.536$), and was statistically significant ($t=5.5$, $p<0.001$). Variance in student level explained by time-on-task was 28.7%. Therefore, hypothesis H2 was supported.

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 4 shows the standardized regression coefficients for the mediation model. The direct effect of badges on the student level is not significant ($p=0.08$), 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 ($p<0.001$), as is the direct effect of time on the student level ($p<0.005$). 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 [0.165, 0.501] does not include zero; therefore, the null hypothesis is rejected. 52.26% of the total effect is mediated. Therefore, hypothesis H3 is confirmed.

RQ 2: Do students with different levels of prior knowledge react differently to gamification?

We also investigated the relationships between students' prior knowledge (using the pretest score), the number of badges achieved as the moderating variable,

time-on-task as the mediating variable, and student level as the outcome variable (Fig. 5). The direct effect of the pretest score on the time-on-task was not significant ($p = 0.48$). However, the interaction variable (pretest \times badges) significantly ($p < 0.01$) and positively affects ($\beta = 46.4$) time-on-task, which shows that badges moderate the effects of pretest over time-on-task. R^2 change due to moderation effect is 0.098, indicating that the interaction effect accounted for 9.8% added variation in time-on-task. Moreover, time-on-task significantly affects the student level ($\beta = 0.0044$, $p < 0.0001$), 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 (pretest) and outcome (slevel) variables; however, the index of moderated mediation tested against bootstrap sample and 95% confidence interval confirmed the moderated mediation effect [0.0032, 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 pretest and student level is conditional on the levels of badges. The more badges students get, the more time they will spend on the task, regardless of their pretest score.

As the moderation effect of badges was significant, it is important to investigate different levels of the conditional effects. Figure 6 shows the moderation effects at +1sd (0.86), mean (0), and -1sd (-0.86) of badges. The significant moderation effects ($\beta = 46.4$, $p < 0.005$) were found only on higher badges (+1sd). It means that students who achieved more badges invested significantly higher time-on-task, particularly the ones 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

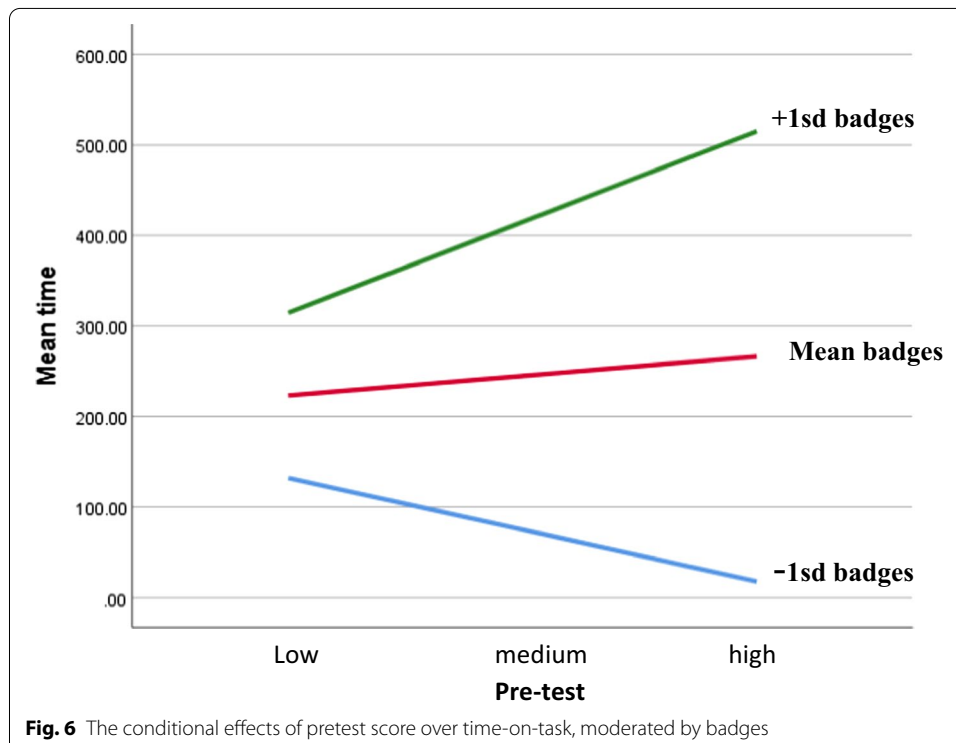


Table 4 Self-efficacy, perceived competence, and topic-interest statistics: mean (sd)

	Self-efficacy		Perceived competence		Topic-interest	
	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
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)

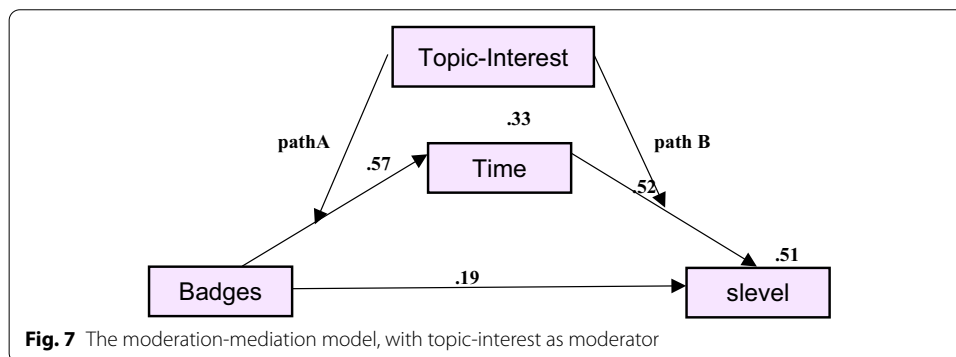


Fig. 7 The moderation-mediation model, with topic-interest as moderator

Fig. 6) regardless of their prior knowledge scores. In fact, the higher prior knowledge group experienced the worst case as their mean time approached zero.

RQ 3: 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. We analyzed 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. We analyzed the scores for each group separately, in order to comprehend the independent effects on student motivation.

Self-efficacy increased after using SQL-Tutor as shown in Table 4, although the differences are not 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 pretest. 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 the differences were not significant, Hypothesis H5 is not supported.

To evaluate Hypothesis H6, we took topic-interest scores of each student in the experimental group from Survey 1 and tested its effects on our mediation model. The model is shown in Fig. 7, where path A tests the effects of topic-interest as a moderator in the relationship between badges and time-on-task, and path B tests the effects of topic-interest as a moderator between time-on-task and student level. We selected model 58 in the Process macro to evaluate the two paths (Hayes, 2017).

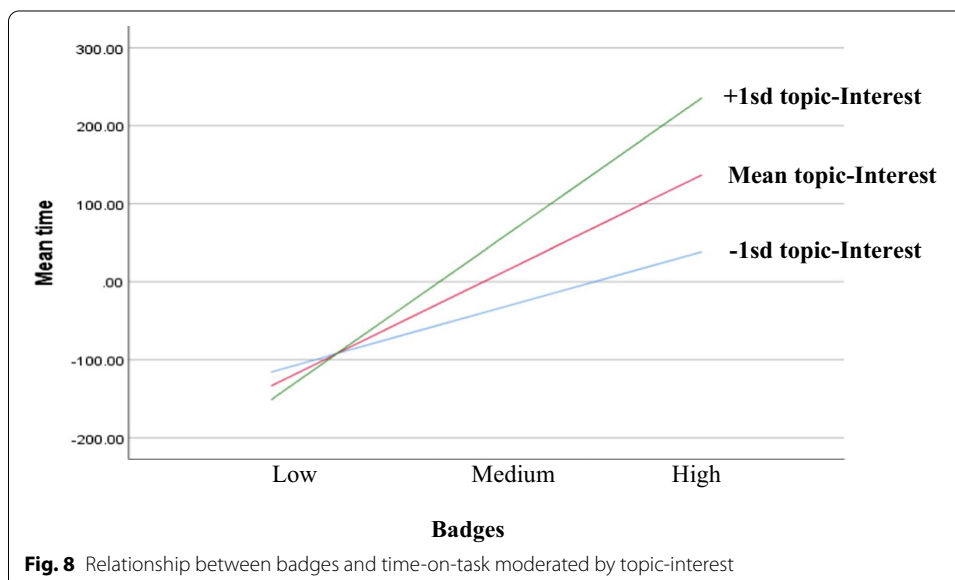
The results of path A revealed a significant positive relation between badges and time-on-task ($\beta = 122, p < 0.0005$) and no significant relation between topic-interest and time-on-task ($p = 0.2$). However, the interaction variable (badges x topic-interest)

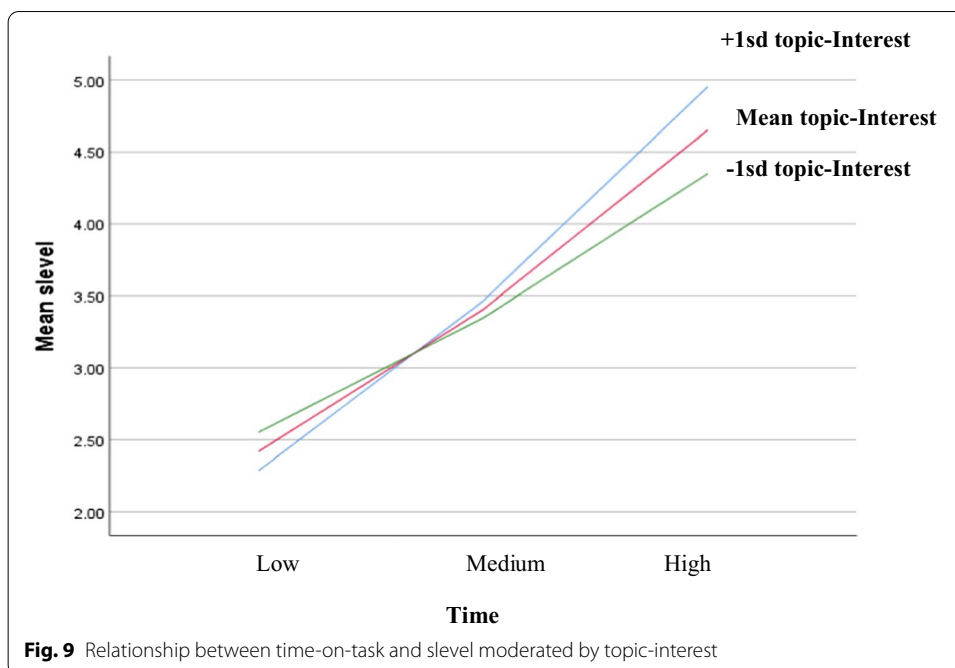
has a significant positive effect ($\beta = 60.86$, $p = 0.05$) over time-on-task. This infers that topic-interest moderates the effects of badges on time-on-task. R^2 change due to moderation effect is 0.0674, indicating that the interaction effect accounted for 6.74% added variation in time-on-task. Since the moderation is symmetric, we can also interpret our results as the badges moderate the effects of student's interest in the topic and the time they spent on SQL-tutor.

As the interaction term in path A was found significant, we want to probe the interaction to better comprehend the moderated relationship between badges and time-on-task, as shown in Fig. 8. At +1 sd on the topic-interest which indicates the higher topic-interest, the relationship was positive and significant ($\beta = 175$, $p < 0.0005$). Similarly, at the mean (0) which represents the medium topic-interest, the relationship was positive and significant ($\beta = 122$, $p < 0.0005$). Finally, at -1sd of topic-interest which represents the low topic-interest, the relationship was negative and insignificant ($\beta = 69.8$, $p > 0.05$). Figure 8 shows that students who have greater topic-interest earned more badges that motivated them to spend more time-on-task.

The analysis of path B (Fig. 7) revealed no significant relationship between badges and student level ($p > 0.1$) but a significant positive relation between time-on-task and student level ($\beta = 0.005$, $p < 0.0001$). However, the interaction effect between time-on-task and student's topic-interest is negative ($\beta = -0.001$) at $p = 0.09$. R^2 change was 0.0398, indicating that 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 9 shows the moderation effects at high topic-interest (+1SD = 0.87), medium topic-interest (mean = 0), and low topic-interest (-1SD = -0.86). It is evident that students who have the lowest interest in the topic but spent more time-on-task, significantly ($\beta = 0.004$, $p < 0.001$) improved their student level. On the other hand, students with the higher interest





in the topic achieved the highest student level also by spending more time-on-task. These results partially support Hypothesis H6.

From the results of H5 and H6, we 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 who have a higher interest in the topic. In order to motivate those students who are less motivated or have less interest in the topic, we need interventions that can raise their interest in the topic and increase their motivation along with gamification.

Further investigation of the experimental group

Overall, the experimental group students achieved from 4 to 7 badges, with a mean of 5.43 (SD = 0.86). 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 = 0.76). 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 5 presents the differences found between the two subgroups.

Due to the small sample size, we conducted the Mann–Whitney *U* test with the Bonferroni corrections for multiple comparisons with statistical significance accepted at

Table 5 Comparing experimental group students who visited the badge page or not: mean (sd)

	Seen badge page (23)	Not seen (19)	Significant
Pretest %	54.59 (25.05)	65.49 (21.88)	$p = .22$
Time-on-task (min)	365.30 (272.27)	195.32 (316.96)	$U = 348.5, p < 0.001$
Solved problems	47.48 (36.86)	29.47 (44.49)	$U = 326.5, p < 0.01$
Constraints	287.74 (60.98)	247.84 (75.82)	$U = 299.5, p < 0.05$
Badges	5.74 (.81)	5.05 (.78)	$U = 317, p < 0.01$
Student level	3.70 (1.72)	2.84 (1.39)	$p = 0.07$
Posttest %	$n = 13; 4.38 (2.93)$	$n = 8; 5.88 (3.72)$	$p = 0.34$

Table 6 Student level

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

the $p < 0.05$. The results show no significant difference between the two subgroups on the pretest scores. The students who visited the badge page have interacted with SQL-Tutor significantly more, measured either as the total time ($U = 348.5, p < .001$), and solved more problems ($U = 326.5, p < 0.01$) than their peers, and also achieved significantly more badges ($U = 317, p < 0.01$). The students who have seen more badges have used significantly more constraints ($U = 299.5, p < 0.05$) 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 with 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 ($p = 0.07$) or the posttest scores ($p = 0.34$).

Self-testing behavior

As mentioned in Sect. 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 (Table 6). 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 ($F = 3.07, p < 0.05$, partial $\eta^2 = 0.11$) 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 ($p = 0.01$, partial $\eta^2 = 0.09$). Students who attempted the quiz achieved a significantly higher student level.

Table 7 Comparing students who attempted/did not attempt the quiz: mean (sd)

	Not attempted (51)	Attempted (26)	Significant
Pretest %	56.65 (25.75)	62.82 (26.66)	$p = 0.33$
Time-on-task (min)	189.73 (153.89)	397.88 (321.47)	$t = 3.85, p < 0.001$
Solved problems	25.98 (19.09)	60.00 (46.28)	$t = 4.56, p < 0.001$
Max problem complexity	6.37 (1.93)	7.77 (1.42)	$t = 3.26, p < 0.005$
Constraints	244.24 (62.44)	317.23 (63.09)	$t = 4.83, p < 0.001$
Student level	3.22 (1.43)	4.23 (1.88)	$t = 2.64, p < 0.05$
Posttest %	$n = 13; 4.38 (2.93)$	$n = 8; 5.88 (3.72)$	$p = 0.08$

Table 8 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%

Table 7 presents the statistics for students who attempted or did not attempt the quiz. There was no significant difference on the pre-/posttest scores and the laboratory test scores. The students who attempted the quiz interacted with SQL-Tutor significantly more, measured in terms of time and solved problems. They used more constraints, and solved more complex problems, thus achieving higher student levels. 6.5 Survey 3 responses. We received 21 responses from the experimental group and 22 responses from the control group students. Table 8 summarizes the responses to the four questions on badges from the experimental group students. The Cronbach alpha for those questions is 0.88.

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%).

Discussion and 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 that 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-tasks (solving problems 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 increase student learning in ITs (measured as the student level in SQL-Tutor), and that this relation can be mediated by the 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. We determined that time-on-task correlated and predicted 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 were already highly motivated, and used SQL-Tutor to prepare for the laboratory test. However, we found evidence that goal-setting, challenges, and self-testing behaviors implemented as badges indirectly and significantly affected learning outcomes through the time-on-task as the mediator.

The second finding of the study is the prior knowledge did not directly affect time-on-task, however, when combined with badges it yielded significant effects. 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 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 our study.

As mentioned in the literature review, badges do not only engage students but also affect their motivation. In this study, we evaluated student motivation by measuring their self-efficacy, perceived competence, and topic-interest. We found no differences in these three motivational constructs between the two groups. We found that badges enticed students to spend more time-on-task; for that reason, we 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. Lower interest in SQL when combined with achieved badges did motivate students to spend more time but not as much as higher interest did. Similarly, the students' interest in SQL slightly influences their time-on-task and learning outcome (student level) relationship.

In the literature review, we pointed out a few methodological gaps in the educational gamification research. In this study, we tried to fulfill those gaps by following the gamified theory of learning, analyzing the effects of a particular game mechanic (badges) on specific student behavior (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, we can conclude that gamification influences the learning behavior of students, which in turn affects their learning outcomes. It affects both higher prior knowledge and low prior knowledge students; in fact, the more badges they achieve, more the time they spend interacting with SQL-Tutor. Finally, the students' 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.

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.

Abbreviations

ITS: Intelligent tutoring system; MOOC: Massive open online course; GBL: Game-based learning; SQL: Standard query language.

Acknowledgements

We would like to take this opportunity to thank our participants and Jay Holland who helped us to administer the study.

Authors' contributions

The study presented in this paper is a part of the PhD project conducted by Faiza Tahir. Prof Mitrovic is her senior supervisor, and Dr. Valerie Sotardi is the associate supervisor. Prof Mitrovic has been working closely with Faiza on the design of the experiment, data analysis, and paper writing. Dr. Valerie has contributed to design and data analyses. All authors read and approved the final manuscript.

Funding

The first author was granted a Ph.D. Scholarship from the College of Engineering, University of Canterbury, Christchurch, New Zealand.

Availability of data and materials

Data will not be available, due to the constraints posed by the Human Ethics Committee of our University. Only the project team is allowed access to the data.

Declarations

Competing interests

We declare that we have no competing interests.

Author details

¹Intelligent Computer Tutoring Group, Department of Computer Science and Software Engineering, University of Canterbury, Private Bag 4800, Christchurch 8041, New Zealand. ²School of Educational Studies and Leadership, University of Canterbury, Christchurch, New Zealand.

Received: 2 January 2021 Accepted: 25 November 2021

Published online: 17 February 2022

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. (1999). Situational and individual interest in cognitive and affective aspects of learning. Paper presented at the American educational research association meetings, Montreal, Quebec, Canada.
- Ainley, M., Hidi, S., & Berndorff, D. (2002). Interest, learning, and the psychological processes that mediate their relationship. *Journal of Educational Psychology*, *94*(3), 545.
- 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., Corbett, A., Koedinger, K., & Pelletier, R. (1995). Cognitive Tutors: Lessons Learned. *Journal of the Learning Sciences*, 4(2), 167–207.
- 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.
- Collins, C. J., Hanges, P. J., & Locke, E. A. (2004). The relationship of achievement motivation to entrepreneurial behavior: A meta-analysis. *Human performance*, 17(1), 95–117.
- Corti, K. (2006). Games-based Learning; a serious business application. *Informe De PixelLearning*, 34(6), 1–20.
- 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*. Proc. CHI Conference on Human Factors in Computing Systems, Montreal, Canada (p. 311). ACM.
- Dermeval, D., & Bittencourt, I. I. (2020). Co-designing Gamified Intelligent Tutoring Systems with Teachers. *Revista Brasileira De Informática Na Educação*, 28, 73–91.
- 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, 190–210.
- Ellis, H., Heppell, S., Kirriemuir, J., Krotoski, A., & McFarlane, A. (2006). *Unlimited learning: Computer games in the learning landscape*. Entertainment and Leisure Software Publishers Association.
- González, C., Mora, A., & Toledo, P. (2014). Gamification in intelligent tutoring systems. In *Proceedings of the second international conference on technological ecosystems for enhancing multiculturalism* (pp. 221–225).
- 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).
- Hamari, J., Koivisto, J., & Sarsa, H. (2014). *Does Gamification Work?—A Literature Review of Empirical Studies on Gamification*. In *2014 47th Hawaii international conference on system sciences* (pp. 3025–3034). IEEE.
- 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.
- 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.
- Helmeffalk, 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.
- Houffort, N., Koestner, R., Joussemet, M., Nantel-Vivier, A., & Leves, N. (2002). The impact of performance-contingent rewards on perceived autonomy and competence. *Motivation and Emotion*, 26(4), 279–295.
- Huang, R., Ritzhaupt, A. D., Sommer, M., Zhu, J., Stephen, A., Valle, N., Hampton, J., & Li, J. (2020). The impact of gamification in educational settings on student learning outcomes: A meta-analysis. *Educational Technology Research and Development*, 68(4), 1875–1901.
- 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.
- 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., Armstrong, M., & Collmus, A. (2017). How to use game elements to enhance learning: Applications of the theory of gamified learning. In *Serious games and edutainment applications* (pp.457–483).
- 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), 769–785.
- 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.
- Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Prentice-Hall, Inc.
- Locke, E. A., & Latham, G. P. (1994). Goal setting theory. *Motivation: Theory and Research*, 13, 29.
- Locke, E. A., & Latham, G. P. (2019). The development of goal setting theory: A half century retrospective. *Motivation Science*, 5(2), 93.
- Long, Y., & Aleven, V. (2013). Supporting students' self-regulated learning with an open learner model in a linear equation tutor. In *proceedings of the international conference on artificial intelligence in education* (pp. 219–228). Springer.
- 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.
- 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.
- Marder, A. (2015). Stack overflow badges and user behavior: An econometric approach. In *Proceedings of the IEEE/ACM 12th conference on mining software repositories* (pp. 450–453). IEEE.
- Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science*, 26(1–2), 49–63.

- Mitchell, A., & Savill-Smith, C. (2004). The use of computer and video games for learning. *A review of the literature*. LSDA
- Mitrovic, A. (1998). Experiences in implementing constraint-based modeling in SQL-Tutor. In B. Goettl, H. Half, C. Redfield, V. Shute (Eds.) *Proceedings of the international conference on intelligent tutoring systems* (pp. 414–423). Berlin: Springer.
- 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.
- Mitrovic, A., & Ohlsson, S. (1999). Evaluation of a constraint-based tutor for a database language. *Artificial Intelligence in Education*, 10(3–4), 238–256.
- 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 international conference on user modeling, adaptation and personalization* (pp. 131–138). ACM.
- 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.
- Nicholson, S. (2015). A recipe for meaningful gamification. In *Gamification in education and business* (pp. 1–20): Springer.
- 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*.
- Pereira, F. D., Toda, A., Oliveira, E. H., Cristea, A. I., Isotani, S., Laranjeira, D., Almeida, A., & Mendonça, J. (2020). Can we use gamification to predict students' performance? A case study supported by an online judge. In *International conference on intelligent tutoring systems* (pp. 259–269). Springer.
- Prensky, M. (2003). Digital Game-Based Learning. *Computers in Entertainment (CIE)*, 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.
- van Roy, R., Deterding, S., & Zaman, B. (2018). Uses and gratifications of initiating use of gamified learning platforms. In *CHI conference on human factors in computing systems* (pp. 1–6). ACM.
- 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. *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.
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, 7(4), 422–445.
- Suh, A., Wagner, C., & Liu, L. (2018). Enhancing user engagement through gamification. *Journal of Computer Information Systems*, 58(3), 204–213.
- Susi, T., Johannesson, M., & Backlund, P. (2007). Serious games: An overview
- Tahir, F., Mitrovic, A., & Sotardi, V. (2019). Towards adaptive provision of examples during problem solving. In: Chang, M. et al. (Eds.) (2019). In *Proceedings of the 27th international conference on computers in education* (pp. 57–62). Taiwan: Asia-Pacific Society for Computers in Education.
- 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). Cham: Springer.
- van Harsel, M., Hoogerheide, V., Verkoeijen, P., & van Gog, T. (2019). Effects of different sequences of examples and problems on motivation and learning. *Contemporary Educational Psychology*, 58, 260–275.
- van Lehn, K. (2006). The behavior of tutoring systems. *Artificial Intelligence in Education*, 16(3), 227–265.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
