So, H. J. et al. (Eds.) (2020). Proceedings of the 28th International Conference on Computers in Education.
Asia-Pacific Society for Computers in Education

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 behavior.

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 studies 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 behavioral changes 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 behaviors or attitudes, via two theoretical paths. Some gamification elements influence learning behaviors/attitudes, which in turn directly influence learning outcomes; thus, the learning behavior acts as a mediator. In other situations, the influence of students' behaviors or attitudes change 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 the time-on-task as the mediating behavior, 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 behavior. 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 behaviors 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 behaviors 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 Aleven (2014) explored the effects of two gamification features in Lynette (an ITS), which is re-practising of previously completed problems and rewards for each completed problem. The results showed that gamifying the 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 & Aleven, 2016), Lynette rewarded students by awarding stars and badges when they selected unmastered problems and showed perseverance on practising new problems. The gamification was shown to result in higher learning outcomes compared to the control condition, as well as improved knowledge of the problem-selection strategy.

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' behaviors or attitudes and 3) insufficient design guidelines are available due to a lack of empirical studies. Our study attempts to fill 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; 2019). 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 the 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 five consecutive 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.

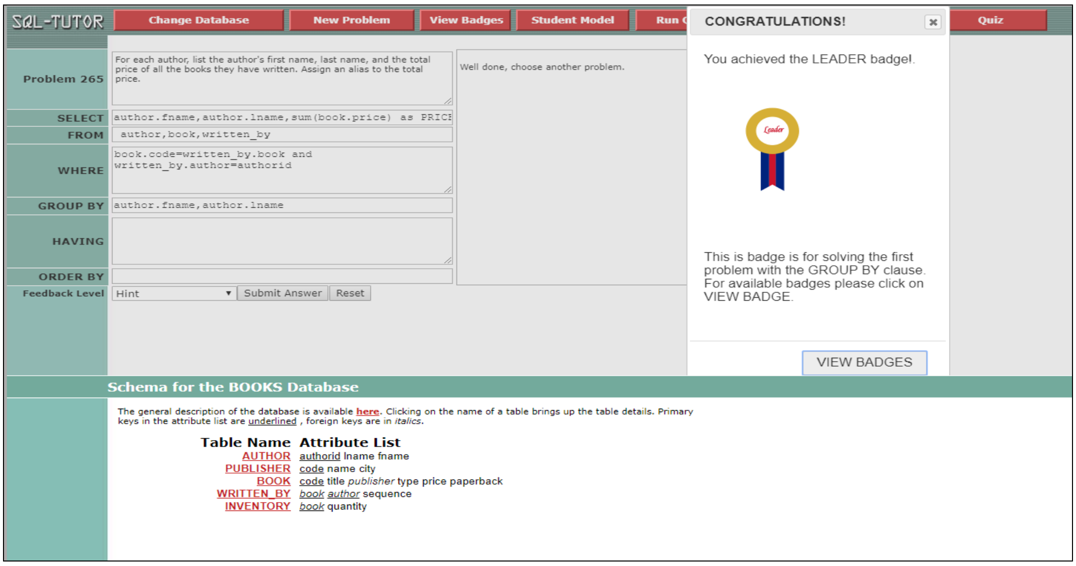

Figure 1. Notification of winning a badge

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.

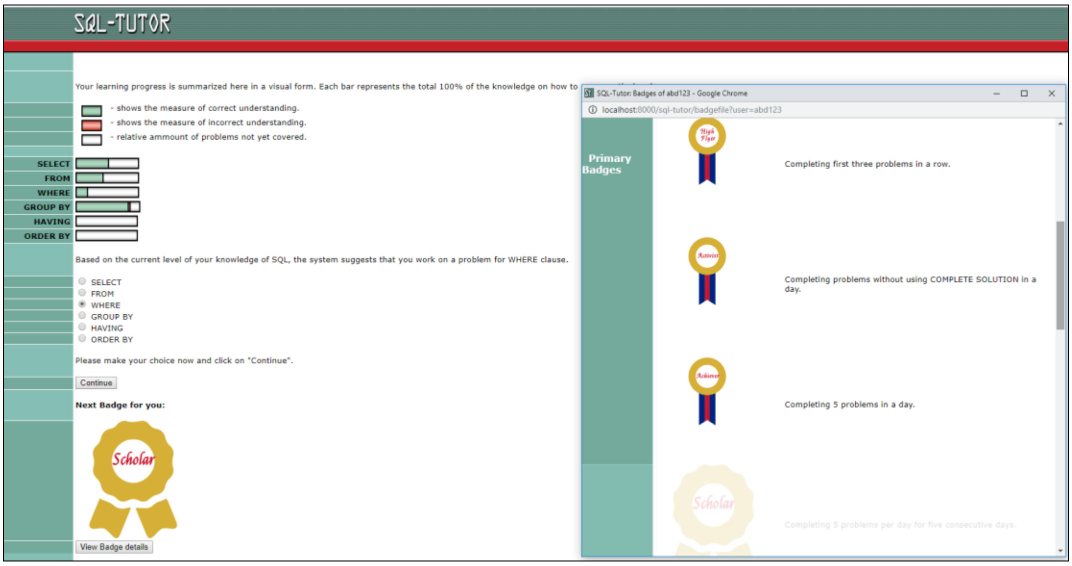

Figure 2. The OLM page, illustrating the next badge (left); the badge page (right)

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.

5.1. 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

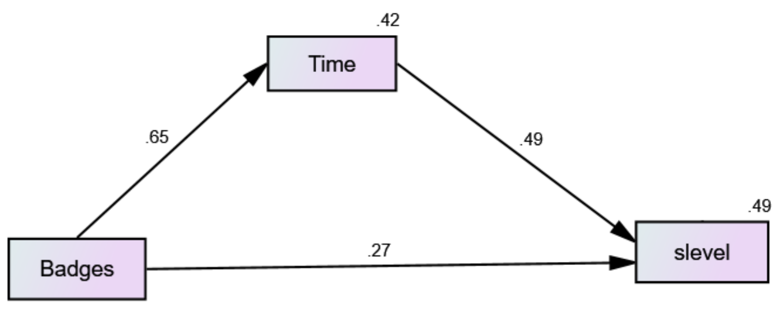

Figure 3. The mediation model, with standardized coefficients

< .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 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.

5.2. 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).

5.3. 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 (Table 4). 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 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 behavior (p = .01, partial η^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)
Exper.	0	28	3.00 (1.47)
	1	14	3.93 (1.77)

Table 5 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 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 in ITSs (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. *Proc. 47th Int. Conf. System sciences*, (pp. 3025-3034)

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.

Locke, E. A., & Latham, G. P. (1990). A theory of goal setting & task performance: Prentice-Hall, Inc.

Locke, E. A., & Latham, G. P. (2019). The development of goal setting theory: A half century retrospective. *Motivation Science*, *5*(2), 93-105.

Long, Y., & Aleven, V. (2013). Supporting students' self-regulated learning with an open learner model in a linear equation tutor. *Proc. Int. Conf. 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. *Proc. Int. Conf. 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. *Proc. Int. Conf. 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. *Proc. 26th Int. Conf. User Modeling, Adaptation and Personalization* (pp. 131-138). ACM.

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.