Evaluating the Effects of Introducing Three Gamification Elements in STEM Educational Software for Secondary Schools

PAMELA ANDRADE, University of Leicester, United Kingdom

EFFIE LAW, University of Leicester, United Kingdom

JUAN CARLOS FARAH, École Polytechnique Fédérale de Lausanne, Switzerland

DENIS GILLET, École Polytechnique Fédérale de Lausanne, Switzerland

Gamification aims to enhance the educational experience by enabling students to have fun with technology. Although research in the field has previously looked into the effectiveness of gamification, reviews of existing studies show that in the context of education, gamification has been largely applied at the university level, with diverse pedagogical approaches and outcomes. Furthermore, reports often refer to interactions with gamification elements in contexts outside STEM education. To bridge these research gaps, this paper reports on an empirical study involving 199 students from two secondary schools. Students had the opportunity to receive a digital reward (in the form of points, badges, or rankings) for their participation in an online physics lesson. Although no significant differences were found regarding student motivation, results confirm an impact—depending on the gamification element being introduced in the software—on three other important aspects: (i) perceived usability, (ii) student engagement, and (iii) learning performance.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**.

Additional Key Words and Phrases: gamification, educational software, motivation and engagement, learning by questioning

1 INTRODUCTION

One of the ongoing challenges that researchers in the field of games and Human-Computer Interaction (HCI) face is whether design and evaluation methodologies can help potentiate positive user experiences [72]. For the most part, it is still unclear *how*, *which*, or even *if* specific elements of games—or gamification—can help create more enjoyable and engaging software applications [46, 81].

In technology-enhanced education, gamification—which is generally defined as "*the use of game design elements in non-game contexts*" [31]—has the goal of capturing the attention of learners [49, 82] by including enjoyable and rewarding interactions in scholastic activities [83]. Although many empirical studies have been conducted in recent years, most of these have analysed the effects of gamification in isolation [51], treating it as a unified concept [80] and generalising findings from one scenario to another, irrespective of contextual differences that could be of influence [84, 96]. Similarly, surveys on HCI and gamification show that although plenty has been studied on the topic, data available

from related research have been collected predominantly from adult learners in the context of distance and higher education [25, 32, 33]. To bridge these research gaps, our study focuses on the effects of using the most popular elements of gamification and instant reward strategies (namely points, badges, and leaderboards) [20, 94] within Graasp[1], a learning platform that enables the virtual delivery of scientific lessons to secondary school students using multimedia resources and online labs.

Furthermore, to utilise the full potential of gamification techniques in technology-enhanced education, this study proposes that it is pertinent to support and enhance the implementation of gamification alongside well-established pedagogical practices to boost positive outcomes [28, 68]. Many pedagogical strategies (e.g. Flipped Classroom [50, 99], Scenario-Based Learning [63, 69]) have shown to improve not only students' attention but also general performance. After much consideration, the *Learning by Questioning* approach was selected for this study, as it lends itself for the introduction of gamification techniques. Although *learning by questioning* has shown potential to improve long-term cognitive skills [38], asking appropriate questions is a cognitively demanding task, as it entails in-depth information processing [52]. Students might therefore need extra encouragement through interactive and enjoyable strategies to sustain their mental effort for the task, and gamification can provide such encouragement by allowing students to have fun during the learning process.

In summary, our research aims to systematically study the effects of integrating a set of common gamification elements (Section 2.2) using the Graasp online learning platform (Sections 3.1-3.2), to reward students for their effort in creating questions (Section 2.1) related to a physics lesson (Section 3.3). The main contributions of this paper are: (i) to provide insights into the impact of three commonly used gamification elements (points, badges, and leaderboards) on the overall experience and learning performance of secondary school students using questioning-based techniques; and (ii) to develop a methodological approach to evaluate these gamification elements when integrated in educational software, in terms of their perceived usability, and their effects on user engagement, motivation, and learning performance.

2 RELATED WORK

This section presents our review of the literature concerning three major topics. First, we explore how students could benefit from *learning by questioning*, which is the pedagogical backbone of our research study. Second, we describe results from studies employing *gamified interfaces* in similar research. Third, we discuss the use of *gamification in educational contexts*, including its impact on student performance and self-reported measures of user experience, and we delve into the difference between student engagement with the gamified software during the learning activity and their motivation to use it in the long term. Finally, we introduce our research question as derived from the literature review.

2.1 Learning by Questioning

Learning by quality questioning is defined as a dynamic process through which students are engaged in cognitive and meta-cognitive processes that focus on learners' thinking and questioning of specific content knowledge [93]. It allows students to explore a topic's meaning, logic, or line of reasoning, so as to help them draw their own conclusions [92]. A number of studies suggest that questioning strategies may contribute to the development of higher cognitive skills, since learners perceive the difference between information they have already studied and what they are learning at that moment [53, 67]. According to the related research, the use of questioning methods could lead to beneficial conceptual connections and comparisons that transfer learning in the brain from short-term into long-term memory [21, 38].

---

[1]Graasp (https://graasp.eu) is an online platform used to create learning spaces supporting interactive lessons.

Recent studies have been therefore drawn by these potential benefits, combining questioning and gamification in technology-enhanced education to potentiate higher learning gains [29].

2.2 Gamified Interfaces

As mentioned in Section 1, gamification is a technique that has increasingly been utilised in education—albeit with different levels of success—to motivate students to engage more actively in the learning process [36]. While there can exist various gamification elements that are fit for different purposes [80], we have selected three—points, badges and leaderboards—due to their usage in a broad range of contexts [24, 55, 73]. All three gamification elements are extrinsic motivators of behavioural intentions [42] and are sometimes referred to as the *PBL triad* [80]. Points are earned for student performance [97]. Badges visually represent student merits [5]. Leaderboards sort students according to their relative success of interaction within the software [80].

Some studies have already evaluated the impact of these gamification elements in different contexts and with various scales. Research on leaderboards shows mostly a positive impact as a motivator, for example, in task performance [57] or software use [40]. Other studies have shown similar positive results for points [9, 64] and badges [14, 45]. Our choice of these three elements was inspired by the insights of these studies, as we aimed to further investigate the impact that the *PBL triad* could have on student experience and performance, to enable more effective uses of such elements in future educational software design. Nevertheless, we must emphasise that not all empirical results from similar research support the use of gamification in education [27, 34, 47], as it has been found to have negative effects on student motivation in the medium to long term [70], and on those students at the bottom of the class [80].

To counteract this issue, we propose the integration of pedagogical methods alongside engagement techniques to boost positive effects. As described in Section 2.1, *learning by questioning* could help transfer short-term storage into knowledge, but being a demanding cognitive task, it could benefit from the encouragement provided by gamification. Our work is based on a number of studies [45, 48, 65] and surveys [1, 4] that have found that gamification has positive effects on several educational aspects, including user experience and learning performance. We also build on recent studies that have looked into the qualitative and quantitative effects of implementing gamification [6, 88], including work conducted using similar questioning-based techniques [28]. Hence, supported by the literature and the findings of these studies, we aim to investigate the effects of combining gamification and questioning-based techniques in technology-enhanced education to shed light on how students can learn in a more enjoyable manner [22, 98] while enhancing the storage of lasting information in their memory [38].

2.3 The Impact of Gamification in Educational Contexts

Alongside the analysis of student performance—which is certainly important as a goal of gamification in education—other measures, including usability and user engagement, have proven highly relevant in determining the effects of implementing gamification [81]. Exploring the relationships between the perceived *usability* of a gamified educational application and the impact it has on student *motivation* and *engagement* could not only help design better interfaces, but also have an impact on student learning gains.

Motivation and engagement are two concepts so tightly connected they often overlap [4, 35], but the former usually refers to the psychological inclination to take part in an activity [79], whereas the latter refers to the enthusiasm and attention that a person shows during a task [8, 78]. Motivation and engagement can influence the persistence of students to master skills [61]. Hence, it is important to design tools that utilise gamification as a technique to boost motivation [30], allowing students to feel more inclined to use gamified educational software to learn. Users' experiences

with gamified software can be measured with different scales, most of which are grounded in relevant conceptual frameworks such as Self-Determination Theories (SDTs) [26]. Other researchers [46] have likewise suggested studying social contextual factors—alongside the technological aspects—as the social environment of users could differently influence their motivation and engagement (e.g. due to the willingness of user groups to participate more actively [2]), depending on their motivational learning behaviour and user type [17].

Strengthening motivation and engagement through gamification in education could have a decisive impact on student performance, according to some recent research [4, 10, 19]. Most studies have shown positive effects [75, 81], but some have—on the contrary—found unfavourable outcomes of using gamification in the long term [47, 56].

In HCI venues, a number of people argue for using game elements in interface design to create a pleasant learning atmosphere where students can relax, engage, and have fun [46, 60]. However, many other authors advise that the use of gamification elements in educational software design could distract students from their main goals, resulting in various negative effects, including lower intrinsic motivation and performance [47, 86], especially on those users who are at the bottom of the group [97]. More empirical research is therefore claimed to be needed to clarify these conflicting outcomes and contrast results observed in previous work.

Based on our literature review, we aimed to investigate the following research question with an empirical study: ***When integrated into a questioning-based physics education activity for secondary school students, to what extent does each of the gamification elements—points, badges, and leaderboards—influence students' perceptions of the usability of the Graasp online learning platform, as well as their user engagement, motivation, and learning performance during the activity?***.

3 DESIGNING THE APPS

We developed two web-based applications (apps) to enable teachers to assign points and badges, and to rank students on leaderboards. Before conducting the empirical study outlined in this paper, we first carried out a formative usability test with key stakeholders. The apps were then published and embedded in the Graasp learning platform, where the online learning activities for our study were hosted. This section outlines the architecture of the apps and the design of the user interfaces.

3.1 Software Architecture

Both the app for points and badges (titled *Rewards*), and the app for leaderboards (titled *Activity Score*) were developed using web frameworks that enabled them to run within a browser. In order to run within Graasp, the apps connected to the platform's Application Programming Interface (API). This connection provided a read/write data store for information related to the online lesson, such as a list of students participating in the learning activity and the rewards assigned to each student. Furthermore, the API allowed the apps to be embedded both within the platform's *Authoring View* and its *Live View*. The *Authoring View* is a password-protected environment where teachers integrate and configure the resources that they will use to create their online lessons. Within this teacher-centric view, the API gives read/write access to apps. The *Live View* is an environment directed at students, accessible through a randomly-generated link, and where students can exploit the activities prepared by teachers. Within this student-centric view, the API gives the apps read-only access to data.

3.2 User Interfaces

The *Rewards* and *Activity Score* apps featured two main interfaces, one for teachers and one for students. As shown in Figure 1, the first interface of the *Rewards* app enabled teachers to assign or revoke rewards to students, as well as to visualise a list of rewards that had already been assigned to students in a class. This interface was only accessible through the *Authoring View* and was therefore protected from unauthorised access. A second read-only interface presented students with a *Trophy Room* where they could visualise the rewards that they had earned. As shown in Figure 6, the app supported assigning multiple rewards, of different types, to the same student. Within this view, students were only able to see their own rewards and not those of others. As shown in Figure 5, the *Activity Score* app displayed a leaderboard—accessible to students and teachers—ranking all participants in a class.
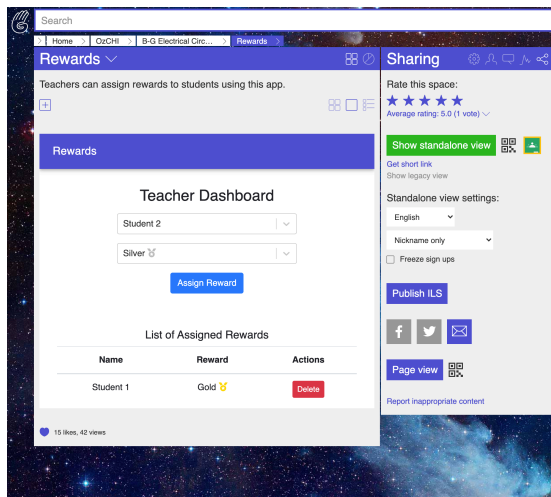

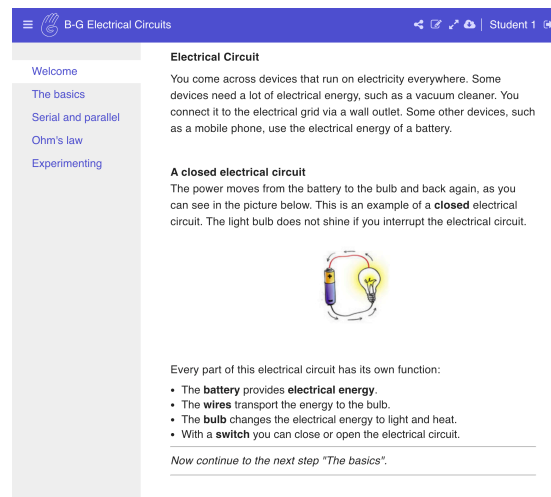Fig. 1. Teacher View of the *Rewards* App

Fig. 2. Initial View of the Online Lesson

3.3 The Online Physics Lesson

For the purpose of this study, we selected an online lesson in Graasp that was aligned with the relevant secondary school physics curricula. The online lesson covered concepts and components of electrical circuits and was adapted to incorporate *learning by questioning*. The questioning technique provided students the opportunity to deepen their understanding of the topic as they were prompted to organise and synthesise information available to them in order to create appropriate questions [58]. *Learning by questioning* commonly refers to the process of articulating conceptual thoughts through written or verbal expressions in the form of interrogations. This technique was selected based on a number of studies, which suggested that the act of questioning could possibly help develop higher cognitive skills [23].

As depicted in Figures 2-6, the *Live View* provided the context for the learning activity. Students could navigate through the different steps or *phases* of the activity using the menu on the left side of the screen. The phases prepared for this study were progressively displayed to students according to the task they were asked to perform. An initial introduction to the topic of electrical circuits was presented in the first phase of the online lesson (see Figure 2). Moving along the phases, students learned about current, tension, electrical power, serial and parallel circuits, and Ohm's
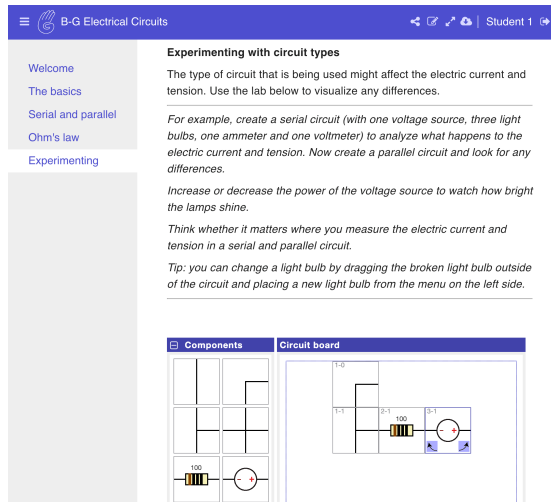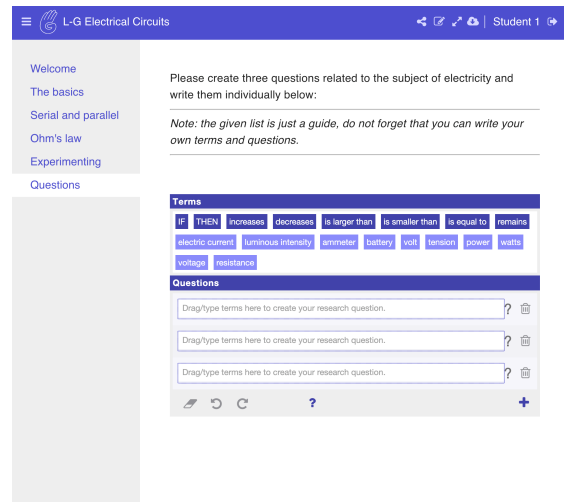
Fig. 3. Description of the Experiment

Fig. 4. Question Creation Phase

law. Students also had access to an online laboratory where they could test the concepts just learned (see Figure 3). Guidelines on how to use the lab were given throughout the lesson to facilitate a connection between the scientific topics addressed and the experimental phase.
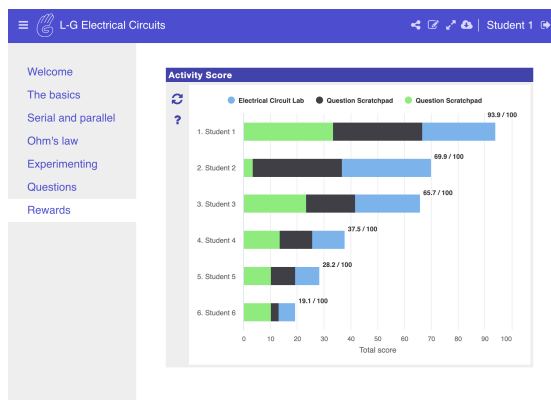
Fig. 5. Student View of the *Activity Score* App
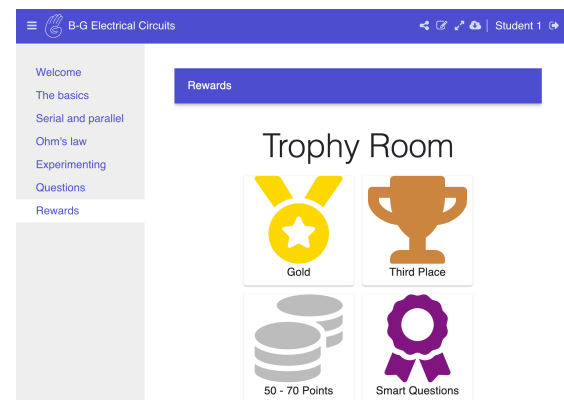
Fig. 6. Student View of the *Rewards* App

After students had gone through the online lesson, they were asked to write down the three best questions they could formulate on the topic of electrical circuits. For this purpose, a new phase was made available for students in the *Live View* of the activity (Figure 4). The three gamification elements—points, badges, and leaderboards—were then used to reward, in each respective group, the effort each student put when creating the set of questions. Displayed on a newly available phase (see the navigation menu on the left hand side of Figures 5 and 6), these rewards were shown to students over a second session, where they were asked again to create three questions about the topic. This

questioning-based gamification approach gave students a second opportunity to get a reward, providing valuable input with respect to possible changes in student motivation and performance.

4 METHODOLOGY

The empirical study was approved by the Research Ethics Committee of the University of Leicester. All sessions took place in school facilities where students had their own computer for the whole duration of the experiment. Throughout a session, a single researcher and the science teacher of each respective class were present to monitor the process, but they did not give any unsolicited help. The four versions of the gamification apps under evaluation in this study (see Section 3), were: non-gamified [NG], gamified with leaderboard [GL], gamified with points [GP], and gamified with badges [GB].

4.1 Methods and Instruments

Eight different instruments were digitally used to collect data from participants to address our main research question as stated in Section 2.3. Four of these instruments were standardised questionnaires to measure: (i, ii) predisposition to learn science, (iii) usability, (iv) user engagement, and (v) motivation. Additionally, homegrown questionnaires were used to assess (vi) learning performance and (vii, viii) user perceptions.

(i) The **Attitudes Towards Science Questionnaire (ATS)** [51] assesses components of attitudinal predisposition of young students towards learning science. The ATS is composed of 34 items and six factors: Learning Science in School, Practical Work in Science, Science Outside of School, Importance of Science, Self-Concept in Science, and Future Participation in Science. Results are presented on a scale of 0-1.

(ii) The **Science Motivation Questionnaire II (SMQ)** [41] is designed to assess students' motivation to learn science in college and secondary school courses. It contains 25 measurement items rated on a 0 (never) to 4 (always) scale and comprised of four factors: Intrinsic Motivation, Self-Determination, Self-Efficacy, and Career Motivation. Results are presented on a scale of 0-1.

(iii) The **System Usability Scale (SUS)** [18] is a ten-item questionnaire using a five-point Likert scale to measure usability. SUS results are presented as scores with a range of 0-100; the higher the score, the better usability the system under evaluation has.

(iv) The **User Engagement Scale - Short Form (UES-SF)** [71] measures user engagement in terms of four different factors (Focused Attention, Perceived Usability, Aesthetic Appeal and the Reward Factor). Each item is rated with a five-point Likert scale. Individual factors can be calculated by averaging the scores of the respective three items, but more commonly, an overall user engagement score is calculated by adding all 12 items and dividing them equally. Results from UES-SF are given on a scale of 1-5.

(v) The **Situational Motivation Scale (SIMS)** [44] is based on SDTs and is composed of different measurement factors of situational intrinsic and extrinsic motivations, rated on a seven-point Likert scale ranging from one (not at all in agreement) to seven (completely in agreement). SIMS evaluates user responses to the items of four factors: Intrinsic Motivation, Identified Regulation, External Regulation, and Amotivation. Results are reported either with an overall score averaged over all 16 items, or factor-specific scores averaged over the four corresponding items per factor.

(vi) **Pre- and post-knowledge tests (KT)** were composed of 12 questions closely related to the physics lesson used in the experiment. The pre- and the post-tests had identical questions, which were all multiple choice with a single possible correct answer. As an example, the following was a question from the knowledge test: "*What is the unit of measurement for the resistance to the flow of electricity? (a) Ohm, (b) Ampere, (c) Watt, (d) Volt*".

(vii) **Homegrown Questionnaires (HQs)** included questions to collect data specifically useful for our research. Students answered two HQs during this study: before and after the intervention. Questions referred to the participants' background, their self-reported measures of IT competence and usage, their perceptions of the enjoyability and usefulness of educational software they have used prior to this experiment, their opinions of the gamification element they had interacted with in this experiment, and other feedback.

(viii) **Group Discussions (GDs)** and general observations are often useful approaches to collect data for evaluating how successful is the interaction between users and a given system [11]. For this study, group discussions were audio-recorded with the participants' consent and notes were taken by the researcher throughout the whole session. Following the experiment, a group discussion was triggered by an online poll (created with Kahoot, Quizizz, or similar tools) where students responded to multiple choice questions. After answering all the questions, and with the results displayed on a screen, participants had the chance to express and share their opinions out loud. The following is an example of a question from this poll: *"Would you be more motivated to use technology to learn if it offered a gamified style? (a) Yes, (b) Maybe, (c) No, (d) I prefer not to answer"*.

4.2 Hypotheses

Following the traditional statistical methods for hypothesis testing, we formulated four null hypothesis (Table 1), which were directly related to the main research question of this study.

Overall, we aim to investigate if specific gamification elements can make a significant impact on the learning experience and performance of students in secondary schools, in the context of questioning-based physics education.

Table 1. List of Null Hypotheses Classified per Research Question and Method or Instrument

Hypothesis	Instrument
H1: There are no significant differences in the perceived usability of the Graasp software among the NG, GL, GP, and GB groups.	SUS
H2: There are no significant differences in student self-reported engagement among the NG, GL, GP, and GB groups.	UES-SF
H3: There are no significant differences in the motivation of students using the NG, GL, GP, and GB versions of the software.	SIMS
H4: There are no significant differences in student performance before and after the intervention among the NG, GL, GP, and GB groups.	KT

4.3 Participants

School A is a state school and *School B* is a private boarding school in the same county in the East Midlands of England. According to the official website of the UK government [87], the term *state school* in England refers to education that is funded by the government and as such is free of charge to students. A *private school* refers, on the other hand, to an independent school not economically supported by the government that instead charges pupils a fee for their attendance. A *boarding school* refers to the option of paid accommodation on school premises throughout the academic year.

Students in both schools have access to good scientific and technological facilities. A total of 199 students were recruited for this study: 108 students in *School A* (53 male, 52 female) and 91 students in *School B* (60 male, 25 female); 9 people preferred not to specify their gender. The mean student age was 13.76 ($SD = 0.47$).

For the learning activity, schools selected the topic of electrical circuits from a predefined list. Eight classes of students were involved in this study, four in each school. Each class was allocated to one of the four groups (non-gamified [NG], gamified with leaderboard [GL], gamified with points [GP], and gamified with badges [GB]). Students interacted with the educational software application for an average of two school hours. The distribution in School A is summarised as follows: 25 students in the [NG] group, 25 in the [GL] group, 31 in the [GB] group, and 27 in the [GP] group. In School B there were: 22 students in [NG], 24 in [GL], 23 in [GB], and 22 in [GP].

Data of School A and School B were analysed separately throughout this paper given the highly significant difference in the initial knowledge test (or pre-test) and the results of the ATS and SQM questionnaires (see Sections 5.7, 5.2, and 5.3, respectively). These significant differences suggest that the schools' social and educational environments could be factors influencing student performance and, quite possibly, their perception of gamified educational software.

4.4 Procedure

In the days prior to the sessions planned for this study, participants filled in the ATS and SMQ questionnaires under supervision of their schools to measure their attitudes and motivation towards learning science. As shown in Table 2, the first session of this experiment started with a general introduction and other relevant instructions (Activity 1 or A1). Afterwards, students filled in a background data questionnaire (A2). All students then interacted with a copy of the same online physics lesson on electrical circuits (A3) and completed a knowledge test to assess how much they learned (A4). Using the software, students were asked to create a set of three questions based on the lesson's subject (A5), with students not being aware that their participation could be (or not be) rewarded with points, badges, or in leaderboards. They later completed the SIMS questionnaire (A6) and had a 1–72 hour break.

Table 2. Protocol of the Study Highlighting the Average Duration *t* along with Instruments and Methods used by the Participating Non-Gamified [NG] and Gamified [G] Groups.

Code	Activity	Resources	*t* [NG]	*t* [G]
A1	General Information	Slides	5'	5'
A2	Personal Background	DQs	3'	3'
A3	Physics Lesson	Lesson Material	15'	15'
A4	Pre-Test	KT1	5'	5'
A5	Question Creation (1)	Software	10'	10'
A6	Motivational Test	SIMS1	4'	4'
	Break	-	-	-
A7	Reward Visualization	Software	-	5'
A8	Question Creation (2)	Software	10'	10'
A11	Self-Reported Perceptions	SUS + UES-SF	10'	10'
A10	Post-Test	KT2	5'	5'
A9	Motivational Test	SIMS2	4'	4'
A12	User Feedback	PIQ + GDs	20'	15'

For the second session, all students (except the ones in the control groups [NG], as they were not given any reward) were presented with results in the form of the gamification element under evaluation in their respective groups (A7). Points, badges, and leaderboards were displayed in a new phase of the online activity for the three questions they had created before the break. This meant that the GL, GB and GP groups became then fully aware that the questions they had created were assessed and rewarded. Afterwards, all students were asked to create a second set of three questions (A8) and they completed once again the standardised motivational test (SIMS) to measure if there was a change in

their behavioural intention (A9). A knowledge post-test (which contained the same questions as the pre-test) was also administered to test if the use of gamification and questioning in the classroom had any impact on the learning outcomes (A10). Then, all students filled in the UES-SF and SUS questionnaires (A11). Lastly, each group responded to an online poll to give general feedback and participants were given the opportunity to express and contrast opinions about their group and personal experience through a moderated discussion (A12).

5 RESULTS

Eight instruments (see Section 4.1) were applied in this study in order to test the hypotheses listed in Table 1. Results of the main methods and instruments are presented in the sections below, preceded by some descriptive statistics based on the demographic data. Note that *not* all participants answered all the questionnaires; some experienced technical difficulties (e.g. the computer crashed at some point) and some were not present at the time of the activity (e.g. because of other duties at school). Hence, the number of total participants per questionnaire varied.

5.1 Demographic Data

In HQ1 (see Section 4.1) students rated their IT skills on a scale of 1-10. In School A the mean was 6.66 ($SD = 1.78$) and in School B the mean was 6.84 ($SD = 1.77$). When asked to estimate how many hours per day they spent on an electronic device—such as a tablet or computer—40 students (37.03%) in School A claimed to use 1-3 hours daily, 47 students (43.51%) reported 4-7 hours of use, and 21 students (19.44%) over 7 hours. In School B there were 37 students (40.66%) who reported to be spending 1-3 hours, 50 students (54.95%) who reported 4-7 hours, and 4 students (4.40%) who said to spend more than 7 hours per day on an electronic device. Additionally, a high number of students in both schools ($n = 166$ out of 199) said to have used educational software in the past, with over 75% of them being able to access said software from both their home and school.

5.2 Results from the Attitudes Towards Science Questionnaire

All 199 students involved in the study completed the ATS (see Section 4.1). School A had a mean score of 0.63 ($SD = 0.12$) and School B a mean score of 0.73 ($SD = 0.10$). Results of the Shapiro-Wilk test indicated that the data were not normally distributed in all groups (School A: $W(108) = 0.968, p = 0.010$; School B: $W(91) = 0.983, p = 0.282$). Results of the non-parametric Mann-Whitney test show a significant difference between schools ($U = 8514, Z = -5.901, p < 0.001$).

Table 3. Results of the ATS Questionnaire Classified by School and Group

School	Group	N	Mean	SD
School A	Control [NG]	27	0.59	0.10
	Leaderboards [GL]	25	0.63	0.11
	Badges [GB]	29	0.66	0.13
	Points [GP]	28	0.66	0.11
School B	Control [NG]	21	0.71	0.13
	Leaderboards [GL]	24	0.76	0.11
	Badges [GB]	22	0.69	0.08
	Points [GP]	23	0.77	0.07

5.3 Results from the Science Motivation Questionnaire

A total of 196 students completed the SMQ (see Section 4.1). School A had a mean score of 0.64 ($SD = 0.15$) and School B a mean of 0.76 ($SD = 0.12$). Results of the Shapiro-Wilk test show that the data were not normally distributed in School A ($W(107) = 0.915, p < 0.001$), but they were normally distributed in School B ($W(89) = 0.989, p = 0.669$). Similarly to the ATS results described above, the non-parametric Mann-Whitney test shows a significant difference between schools in the SMQ scores ($U = 8206, Z = -5.902, p < 0.001$).

Table 4. Results of the SMQ Classified by School and Group

School	Group	N	Mean	SD
School A	Control [NG]	27	0.55	0.14
	Leaderboards [GL]	24	0.62	0.18
	Badges [GB]	29	0.70	0.13
	Points [GP]	27	0.68	0.12
School B	Control [NG]	22	0.74	0.14
	Leaderboards [GL]	24	0.77	0.13
	Badges [GB]	22	0.71	0.09
	Points [GP]	21	0.82	0.09

5.4 Results from the System Usability Scale

Of the 199 students involved in the study, 175 completed the SUS (see Section 4.1). Overall, School A reported a mean SUS score of 54.43 ($SD = 14.67$) and School B a mean score of 63.90 ($SD = 15.49$). Results of the Shapiro-Wilk test indicated that the SUS data of seven out of eight groups were normally distributed. Parametric tests were then used to verify **H1**.

Table 5. SUS Mean Scores in School A and School B

School	Group	N	Mean	SD
School A	Control [NG]	25	50.00	*12.29*
	Leaderboards [GL]	23	50.75	*14.66*
	Badges [GB]	19	52.37	*15.88*
	Points [GP]	26	63.46	*12.49*
School B	Control [NG]	18	59.58	*20.15*
	Leaderboards [GL]	24	62.81	*13.19*
	Badges [GB]	17	63.38	*13.49*
	Points [GP]	22	68.80	*14.63*

The control groups [NG] in both schools gave the learning software application the lowest usability scores. The order of the other three groups (the gamified groups) is also consistent across both schools: [GL] in third place, [GB] in second place, and [GP] in first place (see Table 5).

Combined data from both Schools showed that overall, there were significant differences among the four groups when using the one-way ANOVA test ($F(3, 171) = 5.386, p = 0.001$). As revealed by a Tukey post-hoc test, the points

group (M = 65.97, SD = 13.66) had a significantly higher SUS score than the control group (M = 54.01, SD = 16.54). No significant difference was found between the control group and the leaderboards group (M = 56.91, SD = 15.06), or the control group and the badges group (M = 57.57, SD = 15.62).

A more detailed analysis of the results of the one-way ANOVA reveals that for **School A**, there were significant differences in the SUS scores among the four groups ($F(3, 89)$ = 5.305, p = 0.002). The Tukey post-hoc tests also showed that there was a significant difference between the points group [GP] and the three others: the badges [GB] (p = 0.043), the leaderboards [GL] (p = 0.009), and the control group [NG] (p = 0.004). For **School B**, results of the one-way ANOVA test showed that there were no significant differences among the four groups ($F(3, 78)$ = 1.293, p = 0.283).

Furthermore, we observed a high correlation between software usability and the previous attitudes and motivations of students towards science (as collected by the SUS, ATS, and SMQ questionnaires). Simple linear regression showed a significant relationship ($p < 0.001$) between previously collected student attitudes and their perception of software usability. The R^2 value was 0.411, which indicates that 41.1% of the variation in the usability scores could be explained by the attitude of students towards learning science. Likewise, previously reported motivations of students towards science were found to have a significant relationship in this study with the perception of software usability ($p < 0.001$). The R^2 value was 0.182, which indicates that 18.2% of the variation in the usability scores could be explained by the previous self-reported motivation in students.

Based on the analysis of results of the SUS data in this experiment, the **null hypothesis H1** was **partially rejected**. A significant difference on the perceived usability of the application was found between groups in School A.

5.5 Results from the User Engagement Scale

A total of 174 students responded to the UES-SF. Overall, School A had a mean score of 2.81 (SD = 0.75) and School B a mean score of 2.91 (SD = 0.76), out of the five possible points in the UES-SF (see Section 4.1). All the UES-SF data were normally distributed. Hence, parametric tests were used to address **H2**.

Both schools had the lowest engagement scores in the control groups [NG], as shown in Table 6. As with the SUS results in Section 5.4, the UES-SF scores of the control groups [NG] are followed, in ascending order, by the leaderboards groups [GL], then the badges groups [GB], and—with the best scores—the points groups [GP].

Table 6. UES-SF Mean Scores in School A and School B

School	Group	N	Mean	SD
School A	Control [NG]	25	2.52	0.70
	Leaderboards [GL]	20	2.66	0.93
	Badges [GB]	19	2.98	0.61
	Points [GP]	26	3.07	0.65
School B	Control [NG]	21	2.40	0.84
	Leaderboards [GL]	24	3.02	0.69
	Badges [GB]	17	3.09	0.52
	Points [GP]	22	3.14	0.71

Results of the one-way ANOVA test showed a significant difference among groups in **School A** ($F(3, 86)$ = 3.077, p = 0.032). The same test also showed significant differences on the UES-SF scores among the four groups ($F(3, 80)$ = 4.904, p = 0.004) in **School B**. Scores from the UES-SF questionnaire are shown for both schools in Table 6.

The Tukey post-hoc test showed that in School A there was a significant difference between the points group [GP] and the control group [NG] ($p = 0.041$), but no significant differences were found between the control groups and the two other groups (leaderboards and badges). The same Tukey post-hoc test in School B showed significant differences between the control groups and all the gamified groups. More specifically, $p = 0.024$ for the leaderboards group, $p = 0.019$ for the badges group, and $p = 0.006$ for the points group.

Additionally, a simple linear regression showed a significant relationship ($p < 0.001$) between the usability of the software and the engagement students had with the learning activity in this study. The R^2 value was 0.679, which indicates that 67.9% of the variation in the engagement scores could be explained by the student perceptions of usability of the software. As suggested by previous research, the nature of the learning application could have influenced the level of user involvement in the activity, according to their expectations of the system [66, 89].

Given these results, the **null hypothesis H2** was **rejected**. Significant differences were found in the UES-SF factors among at least one of the gamification elements and the control groups in both schools.

5.6 Results from the Motivational Scale

During the experiment, students were asked to fill in the SIMS (see Section 4.1) questionnaire twice: before and after knowing that they would receive a reward for their performance. A total of 179 students responded to both questionnaires. Two of the four datasets of the pre-test in School A, as well as three out of four in School B were not normally distributed according to the Shapiro-Wilk test. In the post-test, one group in School A and all four in School B did not have a normal distribution. Therefore, non-parametric tests were used to verify **H3**.

Combined results from both schools in the first SIMS (or SIMS1) ($M = 3.85$, $SD = 0.76$) indicated lower student motivation compared to the scores of the second SIMS (or SIMS2) ($M = 3.87$, $SD = 0.80$). However, no significant improvement was found between the two using a Wilcoxon Signed Rank test ($Z = -0.234$, $p = 0.815$).

In addition, no significant changes were found in student motivation between SIMS1 and SIMS2 using Wilcoxon Signed Rank tests separately for **School A** ($Z = -1.061$, $p = 0.289$) and **School B** ($Z = -0.894$, $p = 0.371$). Out of seven possible points, the average SIMS1 score in School A was 3.67 ($SD = 0.86$), which improved to 3.74 ($SD = 0.88$) in SIMS2. The average SIMS1 score in School B was 4.07 ($SD = 0.57$), with a decrease to 4.04 points ($SD = 0.67$) in SIMS2 (see Table 7 for more details).

Table 7. Mean Scores for School A and School B SIMS Data, Before and After the Intervention

Questionnaire	Group	School A			School B		
		N	Mean	SD	N	Mean	SD
SIMS1	Control [NG]	24	3.73	0.61	18	4.03	0.74
	Leaderboards [GL]	21	3.83	1.05	24	3.94	0.50
	Badges [GB]	26	3.30	1.02	17	4.37	0.60
	Points [GP]	28	3.85	0.61	21	4.00	0.37
SIMS2	Control [NG]	24	3.71	0.60	18	4.06	0.99
	Leaderboards [GL]	21	3.78	1.05	24	3.94	0.48
	Badges [GB]	26	3.57	0.91	17	4.26	0.74
	Points [GP]	28	3.88	0.93	21	3.97	0.44

These results **supported** the **null hypothesis H3**. No significant differences were found on the motivational intentions of students as measured by the SIMS scale.

Although motivation did not significantly change with our intervention, it is important to note that the scores of this metric were considerably higher in School B than in School A ($t(177) = -3.564, p < 0.001$). Environmental factors could have influenced these results, as a high correlation was found between the motivational intentions of students and their previously collected attitudes ($p = 0.010$) and motivations towards science ($p = 0.012$). A significant correlation was also found between this metric and the software usability ($p = 0.002$).

5.7   Results from the Knowledge Tests

Of the 199 students that were involved in this study, 95 in School A and 82 in School B completed both the pre- and post-knowledge tests. As previously explained in Section 4.3, data of School A and School B were analysed separately throughout this paper given the highly significant difference in the initial knowledge test (or pre-test) as shown by the result of the Mann-Whitney test ($U = 2173, Z = -6.066, p < 0.001$). For School B, data were normally distributed in all groups except the [GL] in the pre-test and the [GB] in the post-test, as shown by the results of the Shapiro-Wilk test ($W(82) = 0.939, p = 0.001$). Data were normally distributed for all groups both in the pre-test and the post-test in School A.

Table 8. Mean Scores of the Pre- and Post- Knowledge Tests in School A and School B

Test	Group	School A			School B		
		N	Mean	SD	N	Mean	SD
KT1	Control [NG]	24	6.33	0.48	19	8.58	0.48
	Leaderboards [GL]	22	6.50	0.55	24	9.83	0.23
	Badges [GB]	23	6.26	0.44	17	8.18	0.40
	Points [GP]	26	8.88	0.42	22	9.59	0.31
KT2	Control [NG]	24	5.88	0.41	19	8.32	0.54
	Leaderboards [GL]	22	7.73	0.40	24	9.67	0.30
	Badges [GB]	23	6.83	0.55	17	9.41	0.30
	Points [GP]	26	8.96	0.43	22	9.95	0.28

Students in **School A** achieved a mean of 7.05 ($SD = 2.46$) out of 12 possible points in the pre-test (KT1) and an average of 7.38 ($SD = 2.45$) in the post-test (KT2). Significant differences were found in the learning outcomes among groups in School A using a paired sample t test ($t(94) = -2.19, p = 0.031$). The GL group had an increase of 1.23 points, with a significant difference between the pre- and post-tests ($t(21) = -4.67, p < 0.001$)). Although the GB (+0.57) and the GP (+0.08) groups in School A also indicated an improvement on the scores of the post-test as compared to the pre-test, neither had a significant statistical difference ($[t(22) = -1.43, p = 0.168]$ and $[t(25) = -0.39, p = 0.703]$], respectively). Students in **School B**, on the other hand, had an overall mean score of 9.13 points ($SD = 1.71$) on the pre-knowledge test (KT1) and they scored a mean of 9.38 points ($SD = 1.72$) on the post-test (KT2). A paired sample t test showed no significant difference between the two means in this school ($t(81) = -1.12, p = 0.265$)). Nevertheless, two groups had an increase in the KT scores in School B. The GB group showed a significant positive difference (+1.23) in the KT scores ($t(16) = -3.18, p = 0.006$)) whereas the GP group had an insignificant increase (+0.36) on learning performance ($t(21) = -0.95, p = 0.351$)).

Given these analysis results, the **null hypothesis H4** was **partially rejected**. Significant differences were found in the learning performance between the two knowledge tests in School A. Results from the [GB] group were the only ones that showed a significant difference in the learning gains in School B.

6 DISCUSSION

Our work focused on whether the use of a set of gamification elements could have significant effects on students' perception of the usability of the Graasp online lesson [13, 59], and how this overall quality assessment could be related to student motivation and engagement [74], to consequently affect learning outcomes [56]. Our work also contributed to substantiating this applied body of knowledge by gathering more empirical evidence to evaluate the impact of implementing gamified designs in educational software together with a *learning by questioning* approach [58].

Data highlighted in Section 5 show that, when used in secondary schools to teach physics with a questioning-based approach, gamified interfaces can have a positive effect on the perception of usability of the software. Results also suggest that, to some extent, the *PBL* elements can improve student engagement and performance. Moreover, strong correlations were found between the perception of usability of the software and student engagement and motivation. However, no correlations were found between these metrics and the learning gains of students after the intervention, and no significant differences were shown in the motivational responses of students in this study. Overall, the use of points to improve user perceptions of usability and engagement in exploratory scenarios, as well as the use of badges and leaderboards to improve performance in competitive classrooms appear to be effective when teaching physics. These outcomes are discussed in more detail below.

6.1 About the Usability Results

Students of School A reported in the demographic questionnaire (Section 5.1) to have had less prior experience with other educational software as compared to their counterparts in School B. In the same questionnaire, students from School A also rated themselves slightly less skilled with computers. Factors of this nature could have influenced student perception of software usability as suggested by previous studies [12, 85], although no statistical results were fully supportive of this statement in this research. As an example, a positive—although insignificant ($p = 0.086$)—correlation was found between the self-reported IT skills of students in both schools and their perception of usability of the software (the higher their skills, the easier it was to use the software).

Additionally, as observed in findings from previous research [46, 66], students in School A might have felt more attracted to the gamified interfaces due to the novelty effect, thus scoring the usability of the gamified software higher. The broader experience with educational software of students in School B could have enabled them to interact with our apps with more ease, without perceiving a significant differentiation of the software's usability regardless of whether the interface was gamified (with points, badges, or leaderboards) or not.

6.2 About Student Engagement

Prior to this study, School B had already in place both virtual and tangible prizes to acknowledge student merit in a number of school activities, which might help explain why students in School B were found to be significantly more engaged with using the gamified versions of our software. Previous findings (e.g. [62]) have shown that students with prior access to rewards schemes are better at recognising the benefit these bring to their own learning practices.

Additionally, the strong correlations found between previous attitudes and motivations of students to learn science and their engagement with creating scientific questions in this study suggest that personal preferences and circumstances

could affect student involvement in scholastic activities. The social context of each school might have also influenced student perceptions, on different levels, depending on particular environmental factors, as suggested by previous studies [95]. Pre-established social interactions at each school, for example, could have helped shape the role that a given student—as a part of a group—took up during this study [17]. Based on recent research [4] and from an observational point of view in this study, students in School A seemed to mainly enjoy interacting with others online, whereas students in School B mostly enjoyed competing with others to achieve a higher status due to their higher performance.

6.3   About the Motivational Results

In line with previous findings [70, 90], we infer that the absence of significant differences in the motivation of students before and after the intervention could be due, among other factors, to the limited time that students had to interact with the software application. We also agree with some researchers that have in this sense suggested the necessity of medium and long-term empirical research to understand how and why particular game elements act as extrinsic or intrinsic motivators in educational contexts [47, 66, 88].

   In addition, given the large difference between the student self-reported measures of *engagement* and *motivation* in the results of our study, we also support the arguments of previous researchers (e.g. [4, 8]) that have emphasised the importance that both psychological processes have on learning, differentiating them as individual constructs. As discussed in previous research, engagement mainly refers to the level of involvement that a person has when performing a task or activity [77, 78], whereas motivation refers to the personal drive and decision to participate in the activity in the first place [43, 79].

6.4   About Learning Performance

Our results in the sections above have shown that the inclusion of gamification elements in questioning-based educational practices could have a positive influence on student behaviour and perceptions due to environmental factors, as previously analysed in similar research [91, 98]. However, specifically referring to learning performance, no significant relationships were found in this study between the learning gains and other measures of usability, engagement, and student attitudes and motivations towards learning science.

   Results from similar research help explain how user traits could have affected the learning outcomes in this study. First, as students in School A had a lower score on the pre-test (compared to the pre-test score of students in School B), they had more room for improvement in this respect [37]. Second, students in School A had a higher learning gain when they were ranked in the leaderboards whereas students in School B had a better performance when they were awarded badges in the software. This underlying preference could be connected to their social and personal characteristics [4, 39], as discussed in Section 6.2. Third, an explanation for why the non-gamified groups in both schools had a lower score on the knowledge test after the intervention could be that for these students, who were asked to create questions twice, the process of filling out several questionnaires without a clear goal might have been perceived as boring and repetitive [3], as opposed to the gamified groups.

6.5   About Learning by Questioning

To finalise with our analysis, it is important to remark that the quality of the second round of questions improved considerably in both schools, regardless of the gamification element being evaluated. Hence, our use of a questioning-based gamified activity for this study appears effective, which is consistent with previous research [28, 29]. Questions were manually assessed in terms of novelty, clarity, and relevance on a scale of 1-5. Both groups showing significant

16

learning gains in this study (GL in School A and GB in School B) had an increase in the quality of their questions of at least 15%. Quotes from the group discussions in this regard included comments like: *"Having goals to be achieved [in the software] feels encouraging and makes me want to try harder"*. During these group discussions, the majority of students in both schools also claimed to believe that technology could enhance their learning experience when learning about physics. More specifically, a total of 71.59% of students in School A and 86.75% of students in School B said that they would be more motivated to learn physics if the software was gamified (compared to 14.47% of students in School A and a 9.64% in School B who said it would not make any difference to them, with the rest being unsure or declining to answer). The questioning activity was also well-received by the majority of students (74.84%) who showed—regardless of whether they were at the bottom or at the top of the class—an improvement in the quality of their questions. However, students proposed that it would be more beneficial for their learning if they could see and respond to the questions created by their peers as well as their own. This feedback supports the suggestion of previous research to introduce peer questioning and assessment to broaden students' knowledge [92] and increase engagement with the activity [93].

6.6   Limitations and Future Research

It is possible to highlight three factors that could have played a pivotal role in obtaining insignificant results in this study: (i) the science topic selected for the online lesson, (ii) the relatively low social visibility of the rewards, and (iii) the design of the gamification elements embedded in Graasp.

(i) All eight participating groups reported during the group discussions that they would have felt differently about the software if the topic of the lesson would have been different or less complicated. Hence, this study cannot entirely prove that questioning-based gamification would be the preferred technique when teaching physics (and specifically a lesson on electrical circuits) at school. (ii) The way in which the rewards were presented to students could have also played a role in their perception of the software. For example, whereas all students received points and a position on the leaderboard, the fact that only the top three students were awarded badges could have limited the effect of this gamification element. Furthermore, being high on the leaderboards or receiving a badge could have been particularly encouraging only for those students concerned, while the rest of the class could have potentially been negatively affected. (iii) Questions created in the points and badges groups were qualitatively—but manually—assessed in terms of relevance, clarity, and novelty, whereas leaderboards were automatically generated according to the number of interactions students had in a question creation app. Automating the process of delivering points and badges, as well as qualitatively assessing the ranks of students in the leaderboards, are limitations not overcome in the present study, which ideally could be addressed in future research. Moreover, this constraint could have influenced students' perception of the gamification elements, for instance, due to confusion caused by the rubric with which their questions were being assessed. This could have also affected the reflection level of students during the second question creation round, minimizing the benefits of the *learning by questioning* approach described in Section 2.1.

Finally, due to space constraints, we presented only a few salient results from the group discussions and observational methods rather than a more elaborated qualitative data analysis. It could be interesting to analyse, both quantitatively and qualitatively, how these three and other gamification elements could affect student behaviour in different ways and different contexts (for instance, by using competition to motivate student participation, or by providing higher visibility of their achievements [76]). Especially in terms of intrinsic motivation, the analysis of behavioural changes could be more effectively identified with longitudinal studies that assess the medium to long-term effects of regular use of gamified software [7, 66]. Also, extending our research on a cross-cultural level could be potentially beneficial to identify if differences between demographics are relevant when deciding how to design more attractive and engaging

graphic interfaces for educational purposes. Looking into the source of personal preferences in relation to learning styles (e.g. traditions and habits, governmental policies, access to educational technologies) could help decide what practices are best for designing gamified software [54]. Another next step could be to analyse the particular goals and needs of end-users to influence user enjoyment and behavioural change [16] using adaptive gamification design [15].

7 CONCLUSIONS

Results from this study show that the individual implementation of the gamification elements—*points*, *badges*, and *leaderboards*—into the Graasp online learning platform had positive effects not only on the perceived usability of the software in some cases, but more importantly, on the engagement of students with the learning activity in both participating schools. Moreover, groups interacting with the gamified questioning-based technique in our software presented some significant gains on their learning performance, as opposed to the null improvement that the control groups showed. However, not all results from this study are encouraging, as insignificant findings were found concerning student motivation, which suggests the need for future long-term research of these effects.

As not all results were consistent or significant among the gamified groups in the two participating schools, this study can only partially support the premise that technology-enhanced learning can stimulate more optimal conditions for learners by using game elements and a questioning-based approach. With the aim of helping secondary school students attain higher learning gains and enjoy a better user experience, further exploration is necessary to improve the suitable selection of gamification elements according to specific contexts, socio-demographic characteristics, and possibly even user personality traits.

ACKNOWLEDGMENTS

REFERENCES

[1] Azita Iliya Abdul Jabbar and Patrick Felicia. 2015. Gameplay Engagement and Learning in Game-Based Learning: A Systematic Review. *Review of Educational Research* 85, 4 (2015), 740–779.

[2] Icek Ajzen. 1991. The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes* 50, 2 (1991), 179–211.

[3] Donald L Alderman. 1981. Student Self-Selection and Test Repetition. *Educational and Psychological Measurement* 41, 4 (1981), 1073–1081. https://doi.org/10.1177/001316448104100416

[4] Raed S Alsawaier. 2018. The Effect of Gamification on Motivation and Engagement. *The International Journal of Information and Learning Technology* 35, 1 (2018), 56–79.

[5] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. 2013. Steering User Behavior with Badges. In *Proceedings of the 22nd International Conference on World Wide Web* (Rio de Janeiro Brazil). ACM, New York, NY, USA, 95–106.

[6] Pamela Andrade and Effie Lai-Chong Law. 2018. User-Based Evaluation of Gamification Elements in an Educational Application. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference*. BCS Learning & Development, Swindon, UK, 1–13.

[7] Andrés Francisco Aparicio, Francisco Luis Gutiérrez Vela, José Luis González Sánchez, and José Luis Isla Montes. 2013. Gamification: Analysis and Application. In *New Trends in Interaction, Virtual Reality and Modeling*. Springer, London, UK, 113–126.

[8] James J Appleton, Sandra L Christenson, Dongjin Kim, and Amy L Reschly. 2006. Measuring Cognitive and Psychological Engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology* 44, 5 (2006), 427–445.

[9] Yigal Attali and Meirav Arieli-Attali. 2015. Gamification in Assessment: Do Points Affect Test Performance? *Computers & Education* 83 (2015), 57–63.

[10] Gabriel Barata, Sandra Gama, Joaquim Jorge, and Daniel Gonçalves. 2013. Engaging Engineering Students with Gamification. In *2013 5th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES)* (Poole, UK). IEEE, 1–8.

[11] H Russell Bernard, Amber Wutich, and Gery W Ryan. 2017. *Analyzing Qualitative Data: Systematic Approaches.* Sage Publications, Thousand Oaks, CA, USA.

[12] Nigel Bevan and Miles Macleod. 1994. Usability Measurement in Context. *Behaviour & Information Technology* 13, 1-2 (1994), 132–145.

[13] Upasna Bhandari, Tillmann Neben, Klarissa Chang, and Wen Yong Chua. 2017. Effects of Interface Design Factors on Affective Responses and Quality Evaluations in Mobile Applications. *Computers in Human Behavior* 72 (2017), 525–534.

[14] Sanat Kumar Bista, Surya Nepal, Nathalie Colineau, and Cecile Paris. 2012. Using Gamification in an Online Community. In *8th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)* (Pittsburgh, PA, USA). IEEE, 611–618.

[15] Martin Böckle, Isabel Micheel, Markus Bick, and Jasminko Novak. 2018. A Design Framework for Adaptive Gamification Applications. In *Proceedings of the 51st Hawaii International Conference on System Sciences* (Waikoloa, HI, USA).

[16] Rokia Bouzidi, Antonio De Nicola, Fahima Nader, and Rachid Chalal. 2019. A Systematic Literature Review of Gamification Design. In *Proceedings of the 20th Annual Simulation and AI in Games Conference (GAME-ON 2019)* (Breda, The Netherlands). EUROSIS, Ostend, Belgium.

[17] Klaudia Bovermann and Theo J Bastiaens. 2020. Towards a Motivational Design? Connecting Gamification User Types and Online Learning Activities. *Research and Practice in Technology Enhanced Learning* 15, 1 (2020), 1.

[18] John Brooke. 1996. SUS: A 'Quick and Dirty' Usability Scale. In *Usability Evaluation in Industry*, Patrick W Jordan, Bruce Thomas, Ian L McClelland, and Bernard A Weerdmeester (Eds.). Taylor & Francis, London, UK, Chapter 21, 189–194.

[19] Patrick Buckley and Elaine Doyle. 2016. Gamification and Student Motivation. *Interactive Learning Environments* 24, 6 (2016), 1162–1175.

[20] Bunchball, Inc. 2010. *Gamification 101: An Introduction to the Use of Game Dynamics to Influence Behavior.* Technical Report. Bunchball, Inc.

[21] Lucas Payne Butler, Samuel Ronfard, and Kathleen H Corriveau. 2020. *The Questioning Child: Insights from Psychology and Education.* Cambridge University Press.

[22] David Codish and Gilad Ravid. 2015. Detecting Playfulness in Educational Gamification through Behavior Patterns. *IBM Journal of Research and Development* 59, 6, Article 7330105 (2015).

[23] Alan Colburn. 2006. *What Teacher Educators Need to Know about Inquiry-Based Instruction.* Technical Report. California State University, Long Beach, CA, USA.

[24] Tommaso Dal Sasso, Andrea Mocci, Michele Lanza, and Ebrisa Mastrodicasa. 2017. How to Gamify Software Engineering. In *2017 IEEE 24th International Conference on Software Analysis, Evolution and Reengineering (SANER)* (Klagenfurt, Austria). IEEE, 261–271.

[25] Simone de Sousa Borges, Vinicius H. S. Durelli, Helena Macedo Reis, and Seiji Isotani. 2014. A Systematic Mapping on Gamification Applied to Education. In *Proceedings of the 29th Annual ACM Symposium on Applied Computing* (Gyeongju, Republic of Korea) *(SAC '14)*. ACM, New York, NY, USA, 216–222. https://doi.org/10.1145/2554850.2554956

[26] Edward L Deci. 1975. *Intrinsic Motivation.* Plenum Press, New York, NY, USA.

[27] Edward L Deci, Richard Koestner, and Richard M Ryan. 2001. Extrinsic Rewards and Intrinsic Motivation in Education: Reconsidered Once Again. *Review of Educational Research* 71, 1 (2001), 1–27.

[28] Paul Denny. 2013. The Effect of Virtual Achievements on Student Engagement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France). ACM, New York, NY, USA, 763–772.

[29] Paul Denny, Fiona McDonald, Ruth Empson, Philip Kelly, and Andrew Petersen. 2018. Empirical Support for a Causal Relationship between Gamification and Learning Outcomes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* 1–13.

[30] Sebastian Deterding. 2012. Gamification: Designing for Motivation. *Interactions* 19, 4 (2012), 14–17.

[31] Sebastian Deterding, Rilla Khaled, Lennart E Nacke, and Dan Dixon. 2011. Gamification: Toward a Definition. In *Proceedings of the 2011 CHI Gamification Workshop* (Vancouver, BC, Canada). ACM, New York, NY, USA.

[32] Christo Dichev and Darina Dicheva. 2017. Gamifying Education: What is Known, What is Believed and What Remains Uncertain: A Critical Review. *International Journal of Educational Technology in Higher Education* 14, 1 (2017), 9.

[33] Darina Dicheva, Christo Dichev, Gennady Agre, and Galia Angelova. 2015. Gamification in Education: A Systematic Mapping Study. *Journal of Educational Technology & Society* 18, 3 (2015).

[34] Sarah Diefenbach and Annemarie Müssig. 2019. Counterproductive Effects of Gamification: An Analysis on the Example of the Gamified Task Manager Habitica. *International Journal of Human-Computer Studies* 127 (2019), 190–210.

[35] Zoltán Dörnyei and Ema Ushioda. 2013. *Teaching and Researching: Motivation.* Routledge.

[36] Stayc DuBravac. 2012. Game Mechanics for Classroom Engagement. In *Increasing Student Engagement and Retention Using Immersive Interfaces: Virtual Worlds, Gaming, and Simulation*, C Wankel and P Blessinger (Eds.). Cutting-Edge Technologies in Higher Education, Vol. 6 Part C. Emerald Publishing, Bingley, UK, 67–94. https://doi.org/10.1108/S2044-9968(2012)000006C006

[37] John Dunlosky, Katherine A Rawson, Elizabeth J Marsh, Mitchell J Nathan, and Daniel T Willingham. 2013. Improving Students' Learning with Effective Learning Techniques: Promising Directions from Cognitive and Educational Psychology. *Psychological Science in the Public Interest* 14, 1 (2013), 4–58.

[38] Kathryn Ecclestone. 2010. *Transforming Formative Assessment in Lifelong Learning.* McGraw-Hill Education, Maidenhead, UK.

[39] David Folmar. 2015. *Game It Up!: Using Gamification to Incentivize Your Library*. Library Technology Essentials, Vol. 7. Rowman & Littlefield, Lanham, MA, USA.

[40] Panagiotis Fotaris, Theodoros Mastoras, Richard Leinfellner, and Yasmine Rosunally. 2016. Climbing Up the Leaderboard: An Empirical Study of Applying Gamification Techniques to a Computer Programming Class. *The Electronic Journal of e-Learning* 14, 2 (2016), 94–110.

[41] Shawn M Glynn, Peggy Brickman, Norris Armstrong, and Gita Taasoobshirazi. 2011. Science Motivation Questionnaire II: Validation with Science Majors and Nonscience Majors. *Journal of Research in Science Teaching* 48, 10 (2011), 1159–1176. https://doi.org/10.1002/tea.20442

[42] Dimitar Goshevski, Joana Veljanoska, and Thanos Hatziapostolou. 2017. A Review of Gamification Platforms for Higher Education. In *Proceedings of the 8th Balkan Conference in Informatics* (Skopje, Macedonia). ACM, New York, NY, USA, 1–6.

[43] Stephen Grossberg. 1987. A Psychophysiological Theory of Reinforcement, Drive, Motivation, and Attention. In *The Adaptive Brain I: Cognition, Learning, Reinforcement, and Rhythm*, Stephen Grossberg (Ed.). Advances in Psychology, Vol. 42. Elsevier, Chapter 1, 3–81.

[44] Frédéric Guay, Robert J. Vallerand, and Céline Blanchard. 2000. On the Assessment of Situational Intrinsic and Extrinsic Motivation: The Situational Motivation Scale (SIMS). *Motivation and Emotion* 24, 3 (2000), 175–213. https://doi.org/10.1023/A:1005614228250

[45] Juho Hamari. 2017. Do Badges Increase User Activity? A Field Experiment on the Effects of Gamification. *Computers in Human Behavior* 71 (2017), 469–478.

[46] Juho Hamari, Jonna Koivisto, and Harri Sarsa. 2014. Does Gamification Work? — A Literature Review of Empirical Studies on Gamification. In *Proceedings of the 47th Hawaii International Conference on System Sciences* (Waikoloa, HI, USA). IEEE, 3025–3034. https://doi.org/10.1109/HICSS.2014.377

[47] Michael D Hanus and Jesse Fox. 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 (2015), 152–161.

[48] Biyun Huang and Khe Foon Hew. 2015. Do Points, Badges and Leaderboard Increase Learning and Activity: A Quasi-Experiment on the Effects of Gamification. In *Proceedings of the 23rd International Conference on Computers in Education* (Hangzhou, China). 275–280.

[49] Karl M Kapp. 2012. *The Gamification of Learning and Instruction: Game-Based Methods and Strategies for Training and Education*. Pfeiffer, San Francisco, CA, USA.

[50] Jared Keengwe, Grace Onchwari, and James N Oigara (Eds.). 2014. *Promoting Active Learning through the Flipped Classroom Model*. IGI Global, Hershey, PA, USA.

[51] Per Kind, Karen Jones, and Patrick Barmby. 2007. Developing Attitudes Towards Science Measures. *International Journal of Science Education* 29, 7 (2007), 871–893.

[52] Alison King. 1990. Enhancing Peer Interaction and Learning in the Classroom through Reciprocal Questioning. *American Educational Research Journal* 27, 4 (1990), 664–687.

[53] Alison King. 1994. Guiding Knowledge Construction in the Classroom: Effects of Teaching Children How to Question and How to Explain. *American Educational Research Journal* 31, 2 (1994), 338–368.

[54] Jonna Koivisto and Juho Hamari. 2014. Demographic Differences in Perceived Benefits from Gamification. *Computers in Human Behavior* 35 (2014), 179–188.

[55] Jonna Koivisto and Juho Hamari. 2019. The Rise of Motivational Information Systems: A Review of Gamification Research. *International Journal of Information Management* 45 (2019), 191–210.

[56] Elias Kyewski and Nicole C Krämer. 2018. To Gamify or Not To Gamify? An Experimental Field Study of the Influence of Badges on Motivation, Activity, and Performance in an Online Learning Course. *Computers & Education* 118 (2018), 25–37.

[57] Richard N Landers, Kristina N Bauer, and Rachel C Callan. 2017. Gamification of Task Performance with Leaderboards: A Goal Setting Experiment. *Computers in Human Behavior* 71 (2017), 508–515.

[58] Chris Lehmann and Zac Chase. 2015. *Building School 2.0: How to Create the Schools We Need*. Jossey Bass, San Francisco, CA, USA.

[59] Gitte Lindgaard, Gary Fernandes, Cathy Dudek, and Judith Brown. 2006. Attention Web Designers: You Have 50 Milliseconds to Make a Good First Impression! *Behaviour & Information Technology* 25, 2 (2006), 115–126.

[60] Jemma Looyestyn, Jocelyn Kernot, Kobie Boshoff, Jillian Ryan, Sarah Edney, and Carol Maher. 2017. Does Gamification Increase Engagement with Online Programs? A Systematic Review. *PLOS ONE* 12, 3 (2017), 1–19.

[61] Annette Majnemer. 2011. Importance of Motivation to Children's Participation: A Motivation to Change. *Physical & Occupational Therapy in Pediatrics* 31, 1 (2011), 1–3.

[62] Helen M Marks. 2000. Student Engagement in Instructional Activity: Patterns in the Elementary, Middle, and High School Years. *American Educational Research Journal* 37, 1 (2000), 153–184.

[63] Naomi McGrath and Leopold Bayerlein. 2013. Engaging Online Students through the Gamification of Learning Materials: The Present and the Future. In *30th ASCILITE Conference 2013 Proceedings* (Sydney, Australia). ASCILITE, Macquarie University, 573–577.

[64] Elisa D Mekler, Florian Brühlmann, Klaus Opwis, and Alexandre N Tuch. 2013. Disassembling Gamification: The Effects of Points and Meaning on User Motivation and Performance. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1137–1142.

[65] Elisa D Mekler, Florian Brühlmann, Klaus Opwis, and Alexandre N Tuch. 2013. Do Points, Levels and Leaderboards Harm Intrinsic Motivation? An Empirical Analysis of Common Gamification Elements. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications* (Toronto, ON, Canada). 66–73.

[66] Elisa D Mekler, Florian Brühlmann, Alexandre N Tuch, and Klaus Opwis. 2017. Towards Understanding the Effects of Individual Gamification Elements on Intrinsic Motivation and Performance. *Computers in Human Behavior* 71 (2017), 525–534.

[67] Naomi Miyake and Donald A Norman. 1979. To Ask a Question, One Must Know Enough to Know What is Not Known. *Journal of Verbal Learning and Verbal Behavior* 18, 3 (1979), 357–364.

[68] Cristina Ioana Muntean. 2011. Raising Engagement in e-Learning through Gamification. In *Proceedings of the 6th International Conference on Virtual Learning* (Cluj-Napoca, Romania). Bucharest University Press, Bucharest, Romania, 323–329.

[69] Som Naidu, Mohan Menon, Chandra Gunawardena, Dayalatha Lekamge, and Shironica Karunanayaka. 2007. How Scenario-Based Learning can Engender Reflective Practice in Distance Education. In *Finding Your Online Voice: Stories Told by Experienced Online Educators*, J Michael Spector (Ed.). Lawrence Erlbaum Associates, Mahwah, NJ, USA, Chapter 4, 53–72.

[70] Scott Nicholson. 2015. A RECIPE for Meaningful Gamification. In *Gamification in Education and Business*, Torsten Reiners and Lincoln C Wood (Eds.). Springer, New York, NY, USA, 1–20.

[71] Heather L O'Brien, Paul Cairns, and Mark Hall. 2018. A Practical Approach to Measuring User Engagement with the Refined User Engagement Scale (UES) and New UES Short Form. *International Journal of Human-Computer Studies* 112 (2018), 28–39. https://doi.org/10.1016/j.ijhcs.2018.01.004

[72] Randy J Pagulayan, Keith R Steury, Bill Fulton, and Ramon L Romero. 2018. Designing for Fun: User-Testing Case Studies. In *Funology 2*. Springer, 419–433.

[73] Sungjin Park and Sangkyun Kim. 2018. Patterns Among 754 Gamification Cases: Content Analysis for Gamification Development. *JMIR Serious Games* 6, 4 (2018), e11336.

[74] Wei Peng, Jih-Hsuan Lin, Karin A Pfeiffer, and Brian Winn. 2012. Need Satisfaction Supportive Game Features as Motivational Determinants: An Experimental Study of a Self-Determination Theory Guided Exergame. *Media Psychology* 15, 2 (2012), 175–196.

[75] Lisa-Maria Putz, Florian Hofbauer, and Horst Treiblmaier. 2020. Can Gamification Help to Improve Education? Findings from a Longitudinal Study. *Computers in Human Behavior* 110, Article 106392 (2020), 12 pages.

[76] Amon Rapp. 2017. Designing Interactive Systems Through a Game Lens: An Ethnographic Approach. *Computers in Human Behavior* 71 (2017), 455–468.

[77] Johnmarshall Reeve, Hyungshim Jang, Dan Carrell, Soohyun Jeon, and Jon Barch. 2004. Enhancing Students' Engagement by Increasing Teachers' Autonomy Support. *Motivation and Emotion* 28, 2 (2004), 147–169.

[78] Johnmarshall Reeve and Ching-Mei Tseng. 2011. Agency as a Fourth Aspect of Students' Engagement During Learning Activities. *Contemporary Educational Psychology* 36, 4 (2011), 257–267.

[79] Richard M Ryan and Edward L Deci. 2000. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology* 25, 1 (2000), 54–67.

[80] Michael Sailer, Jan Ulrich Hense, Sarah Katharina Mayr, and Heinz Mandl. 2017. How Gamification Motivates: An Experimental Study of the Effects of Specific Game Design Elements on Psychological Need Satisfaction. *Computers in Human Behavior* 69 (2017), 371–380. https://doi.org/10.1016/j.chb.2016.12.033

[81] Katie Seaborn and Deborah I Fels. 2015. Gamification in Theory and Action: A Survey. *International Journal of Human-Computer Studies* 74 (2015), 14–31. https://doi.org/10.1016/j.ijhcs.2014.09.006

[82] David Williamson Shaffer. 2006. *How Computer Games Help Children Learn*. Palgrave Macmillan, New York, NY, USA.

[83] Jorge Simões, Rebeca Díaz Redondo, and Ana Fernández Vilas. 2013. A Social Gamification Framework for a K-6 Learning Platform. *Computers in Human Behavior* 29, 2 (2013), 345–353.

[84] Andrew Stott and Carman Neustaedter. 2013. *Analysis of Gamification in Education*. Technical Report. Simon Fraser University, Surrey, BC, Canada.

[85] Ronald L Thompson, Christopher A Higgins, and Jane M Howell. 1994. Influence of Experience on Personal Computer Utilization: Testing a Conceptual Model. *Journal of Management Information Systems* 11, 1 (1994), 167–187.

[86] Armando M Toda, Pedro HD Valle, and Seiji Isotani. 2017. The Dark Side of Gamification: An Overview of Negative Effects of Gamification in Education. In *Higher Education for All. From Challenges to Novel Technology-Enhanced Solutions* (Maceió, Brazil). Springer, 143–156.

[87] UK Government. [n.d.]. *Types of School in the UK*. https://www.gov.uk/types-of-school [online; accessed 23 October 2020].

[88] Niels van Berkel, Jorge Gonçalves, Simo Hosio, and Vassilis Kostakos. 2017. Gamification of Mobile Experience Sampling Improves Data Quality and Quantity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–21.

[89] Hans van der Heijden. 2004. User Acceptance of Hedonic Information Systems. *MIS Quarterly* 28, 4 (2004), 695–704.

[90] Rob van Roy and Bieke Zaman. 2018. Need-Supporting Gamification in Education: An Assessment of Motivational Effects Over Time. *Computers & Education* 127 (2018), 283–297.

[91] Edda Vélez Meza, Alexis Guzmán Terán, Mónica Gallegos Varela, and Jacinto Méndez Urresta. 2019. Digital Gamification in Basic General Education Students. In *International Conference on Knowledge Society: Technology, Sustainability and Educational Innovation* (Ibarra, Ecuador). Springer, 143–156.

[92] Stacy E Walker. 2003. Active Learning Strategies to Promote Critical Thinking. *Journal of Athletic Training* 38, 3 (2003), 263.

[93] Jackie Acree Walsh and Beth Dankert Sattes. 2011. *Thinking through Quality Questioning: Deepening Student Engagement*. Corwin Press, Thousand Oaks, CA, USA.

[94] Hao Wang and Chuen-Tsai Sun. 2011. Game Reward Systems: Gaming Experiences and Social Meanings. In *Proceedings of the 2011 DiGRA Conference: Think Design Play* (Hilversum, The Netherlands).

[95] Ming-Te Wang and Jacquelynne S Eccles. 2013. School Context, Achievement Motivation, and Academic Engagement: A Longitudinal Study of School Engagement Using a Multidimensional Perspective. *Learning and Instruction* 28 (2013), 12–23.

[96] Siaw-Chui Wee and Weng-Wai Choong. 2019. Gamification: Predicting the Effectiveness of Variety Game Design Elements to Intrinsically Motivate Users' Energy Conservation Behaviour. *Journal of Environmental Management* 233 (2019), 97–106. https://doi.org/10.1016/j.jenvman.2018.11.127

[97] Kevin Werbach and Dan Hunter. 2012. *For the Win: How Game Thinking can Revolutionize your Business*. Wharton Digital Press, Philadelphia, PA, USA.

[98] Ibrahim Yildirim. 2017. The Effects of Gamification-Based Teaching Practices on Student Achievement and Students' Attitudes Toward Lessons. *The Internet and Higher Education* 33 (2017), 86–92.

[99] Zamzami Zainuddin. 2018. Students' Learning Performance and Perceived Motivation in Gamified Flipped-Class Instruction. *Computers & Education* 126 (2018), 75–88.