

User-based Evaluation of Gamification Elements in an Educational Application

Pamela Andrade
University of Leicester
University Road, Leicester
LE1 7RH, United Kingdom
pyas2@leicester.ac.uk

Effie Lai-Chong Law
University of Leicester
University Road, Leicester
LE1 7RH, United Kingdom
lc19@leicester.ac.uk

Gamification has increasingly been utilised to motivate students to use technological tools and applications to learn. Although the existing research seems to recognise the benefits that gamification has in and outside classroom, little has been explored about individual gamification elements. To bridge this gap, we conducted a user-based evaluation of an educational application, Knowma+, with different gamification elements to understand which are more desirable to use from the teacher and student perspectives. Knowma+ is built upon relevant pedagogical concepts such as learning by questioning that fit well with the inclusion of gamification elements to enhance learning experience. Results of eye-tracking data suggest that more visually attractive gamification elements could capture the initial attention of both teacher and student participants. However, there were slight differences between the two groups in self-reported preferences with regard to the perceived usefulness of each gamification element as part of their teaching or learning approaches.

Gamification elements, education, user experience, interaction design, eye tracking.

1. INTRODUCTION

Gamification is nowadays a popular concept being applied in a number of fields (Rojas, Kapralos & Dubrowski, 2013). It is generally defined as “*the use of game design elements in non-game contexts*” (Deterding et al., 2011, p. 10).

While gamification has become a widely adopted approach in marketing and business to preserve customers’ loyalty to a brand or a product (Roth, Schneckenberg & Tsai, 2015; Muntean, 2011), the potential of gamification in education being realised through digital technology has only recently been studied more systematically. Recent studies have reported that when applied in the realm of education, gamification can have a substantial impact (Dicheva et al., 2015).

One of the main goals of gamification in education is to gain the attention of learners and motivate them to actively participate in the learning process (Kapp, 2012; Shaffer, 2006). This is usually achieved by adapting enjoyable gamification elements to fit in scholastic activities (Simões, Redondo & Vilas, 2013). Similarly, gamified computer systems can impact education by examining how users interact with them as part of their routines, by understanding how these applications affect people, and finally, by

adapting to newly discovered educational needs (Richards, Thompson & Graham, 2014).

According to the related literature in Human-Computer Interaction (HCI) and gamification studies (e.g. Nacke & Deterding, 2017; Arnab et al., 2015; van Roy & Zaman, 2015), gamified systems have only been examined within specific environments, contexts and user groups. It has been argued that there is an explicit need to analyse the impact of gamification in different and perhaps more significant ways.

Among other issues, it remains unclear which gamification elements can result in more enjoyable, attractive and effective learning systems (Hamari, Koivisto & Sarsa, 2014), or psychological effects that gamification can have on individuals and their social relationships (de-Marcos et al., 2016). Moreover, as most empirical studies have analysed gamification as a group of elements, no concrete conclusions have been presented so far regarding the contribution of individual gamification elements from students’ and teachers’ perspectives (Mekler et al., 2017).

Another important weakness identified in prior research associated with gamification (Cheong, Filippou & Cheong, 2014) includes the difficulty to give useful feedback to students. Most educational

applications are dedicated to supporting online training for adult users, who usually end up dropping a course because they do not receive timely or meaningful feedback for their involvement.

Based on the limitations identified in previous studies as described above, one contribution our research aims to achieve is to compare individual gamification elements in an educational application. The goal is to understand which of these elements have a more significant effect in the perceived effectiveness, attractiveness and acceptance from the user (i.e., student/teacher) perspective. To make the comparison manageable, it was necessary to narrow down gamification elements to the most commonly used ones in the literature (Section 2.2.1.1). Three elements stand out: badges, leaderboards and points (Bunchball, 2010; Dicheva et al., 2015; Sailer et al., 2017) and have been regarded as the basic elements of gamification strategies, being an integral part of games and instant reward (Wang & Sun, 2011).

Furthermore, our work can contribute to the understanding of the issues pertaining to feedback (Section 2.2.1.2) by including supporting features in the application Knowma+ we developed (Section 3) and by providing meaningful responses to students grouped according to their performance level. In gamification, feedback mechanisms contribute to increased positive affect and self-efficacy by making users feel directly responsible for their success (Tondello, Premasukh & Nacke, 2018).

2. BACKGROUND

Knowma+ has been built on some relevant pedagogical concepts (Section 2.1) using appealing techniques to enhance user involvement (Section 2.2). With both subjective (self-reported surveys) and objective (eye-tracking) data collection methods (Section 2.3), we aim to understand which specific gamification elements are perceived more attractive for users of educational applications.

2.1 Pedagogical Concepts

New pedagogical approaches emerge with the continual evolution of didactics. The constructivist and socio-cultural approaches, for example, focus on experience, interactivity, and cultural context to understand the complex dynamics of knowledge systems (Spector, 2000).

Methods and rules created for classic scholastic practices are no longer sufficient, especially since the evident transformation that major technological improvements have built in the educational systems (Emeling, 2010). Students have become more independent, inquisitive, active, and involved in the educational process, thanks to the ever-increasing

easy access to various sources of information through technologies (Rebolj, 2009).

To address the new pedagogical needs and guidelines, we propose that it is pertinent to implement certain established and viable teaching practices as a technological solution. After much research about the advantages of different teaching strategies, Knowma+ has been built upon the following concepts.

2.1.1. Learning by questioning

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 present, the use of questioning methods leads to beneficial connections and comparisons that can help transfer learning from short-term into long-term memory (Ecclestone et al., 2010).

Additionally, the act of questioning allows students to exchange and contrast ideas with other classmates, articulate their conceptual thoughts through verbal or written expressions, making them more participative in the learning process (Colburn, 2006). Likewise, when formulating questions, the quality of self-induced clarification and investigation can increase significantly. Learners are prompted to organize and synthesize their own understanding of the topic in order to create appropriate questions to share with others (Lehmann & Chase, 2015).

With our understanding of the learning by questioning approach, one of the functionalities that Knowma+ embraces in its design is the option for students to create questions on a specific topic and share them with others, who are enabled to comment and assess these questions (see more details in Sections 3.1 & 3.3).

2.1.2. Formative assessment

Another crucial objective of the design of Knowma+ is to help enhance students' commitment to their own learning. This is done with formative assessment, which is not only a means to certify competence, to rank accomplishments, or to keep track of goals but is also a fundamental activity that provides both students and teachers with feedback that can inform how teaching needs to be adapted to meet learning needs (Beldarrain, 2006).

An integral part of formative assessment is the delivery of proper feedback, with the goal of making learners aware of their own strengths and weaknesses, thereby identifying strategies (e.g., seeking additional learning resources) to improve their work (Buehl, 2013).

Formative assessment is not necessarily restricted to formal tests; it can include more casual

interactions such as comments, conversations or exchanges of opinions (see Sections 3.1 & 3.3). It can help teachers to evaluate how well learners understand the given educational concepts; and for the students it represents an opportunity to diagnose their own progress and decide if or how to acquire more knowledge (Black & Wiliam, 2010).

2.2 Gamification in Learning

Apart from the cognitive aspect of learning and performance discussed above, it is crucial to address the importance of positive learning experience. Innovative trends and research areas have emerged alongside the increasing availability and influence of different types of games in everyday life. But the concept of using rewards and feedback as gamification elements to increase the active participation of students in learning activities is nothing new (Becker & Nicholson, 2016).

Research studies on real-time interactions and games suggest that the act of playing can result in learning if students find an educational system enjoyable to use (Shin, 2006). Therefore, learning could be more appealing to students with the right combination of educational content, challenges and fun (Vassileva, 2008).

Nowadays, a number of methods aiming to enhance learning processes are supported by the interactivity of new educational technologies (Burbules, 2018). Gamification is one of those methods; it can lead to desirable learning outcomes or attitudinal as well as behavioural changes through motivational mechanisms of habit reinforcement and rewards (Robson, et al., 2015).

One of the main goals of gamification (as defined by Deterding et al., 2011; Zichermann & Cunningham, 2011) in education is to gain the attention of learners and involve or motivate them to participate in learning activities. To do so, gamification embraces three reinforcement choices: conflict, cooperation, or competition (Koster, 2005). Cooperation would be to work with others to accomplish mutually aspired goals. A conflict can involve a challenge to win against an opponent, typically outperforming their score or playing against them. Competition is the ability to win not by impeding the opponent from winning, but by optimizing the learner's own performance to achieve a desired goal.

2.2.1. Gamifying educational systems

Several authors have attempted to define a classification of game elements (Iosup & Epema, 2014; Bunchball, 2010), game models (Heintz & Law, 2015; Bista et al., 2012) or frameworks (Nah, Telaprolu & Rallapalli, 2013; Zagal et al., 2007). Although none of these structures have yet become generalised, common concepts to all these have come to be the basis of gamification design.

A gratifying experience for students in gamified educational systems should embrace a wide range of gamification tools, including but not limited to support competition, goal achievements, real-time feedback and more (Deterding, 2012).

With the use of diverse gamification elements and mechanics, teachers could either use external factors to extrinsically (Section 2.2.1.1) motivate students (Deci & Ryan, 2010) or students themselves could be intrinsically (Section 2.2.1.2) motivated to participate and learn more (Banfield & Wilkerson, 2014).

2.2.1.1. The gamification triad.

Points, badges and leaderboards are three of the most common *reward-based* gamification elements employed in both game and non-game contexts (Mekler et al., 2013; Werbach & Hunter, 2012; Zichermann & Cunningham, 2011). They use extrinsic motivation to modify user's behaviour (Goshevski & Hatzia Apostolou, 2017). Various studies have presented short-term benefits of this particular group of game elements (Landers, Bauer & Callan, 2017; Attali & Arieli-Attali, 2015; Meckler et al., 2013), providing evidence of their significance in classrooms.

Points are usually earned for successful interactions of students within the gamified environment (Werbach & Hunter, 2012). Badges are visual representations of merits and accomplishments (Anderson et al., 2011). Leaderboards rank users according to their performance (Sailer et al., 2017).

Rewarding performance and boosting competition are the main features of this type of gamification design. Badges and points are available in Knowma+ as part of the peer-review process (see Figure 3). The (two) leaderboards are updated either when students receive points from peers on the questions they create, or after students take tests and challenges in a module (See Section 3.1). These gamification elements are mainly based on competition and reward mechanics (Werbach & Hunter, 2012).

2.2.1.2. Meaningful gamification

In contrast to the previous group, this approach does not provide rewards but intends to engage users by increasing intrinsic motivations instead (Nicholson, 2015). This type of gamification design cares about long-term benefits that gamification elements can have on students (Goshevski & Hatzia Apostolou, 2017).

The main aim of meaningful gamification is to transform students into active participants in learning practices, enhancing their awareness of themselves and their responsibilities and strengthening their drive to learn more (Becker & Nicholson, 2016). Hence, meaningful feedback is one of its key components.

Encouraging students to reflect and take control of their own learning is the basic characteristic of this type of gamification. In this context, meaningful feedback can be given in Knowma+ either as a result of completing tests (Figure 4) or as comments from other peers who are known to each other, eliciting fun and excitement, especially when comments are shared with witty humour and mutual respect (Figure 3). The former drives user engagement by giving information (explanations and resources) for students to reflect on their work (Nicholson, 2015) whereas the latter appeals to social aspects of interaction and altruism (Werbach & Hunter, 2012).

2.3 Eye movement analysis

Tracking eye movements can assist HCI researchers to identify usability problems of an interactive system as gaze data show which user interface (UI) objects users focus on when interacting with the system. Furthermore, scanpaths (i.e., gaze sequences) can help the analysis of cognitive processes in decision making (Orquin, Ashby & Clarke, 2016).

Recordings of gazes can provide a dynamic trace of where in a visual display a person's attention is directed at (Poole & Ball, 2006). Hence, usability problems can be derived from eye-tracking data such as fixation duration, scanning behaviour, heat maps indicating (non)attended UI objects of a webpage, hesitation, etc. (Ehmke & Wilson, 2007). Specifically, in our study the use of the eye-tracking technique can result in crucial insights, both for usability problems in Knowma+ (Section 3), and especially with regards to the difference in user attention paid to the gamification elements implemented.

2.4 Research Questions

Knowma+ is an educational application based on key pedagogical strategies (Section 2.1) integrated with a gamified approach (Section 2.2). With our analytic and empirical work, we aim to address two main research questions (RQs):

- RQ1:** Which gamification elements are perceived as more attractive for students and teachers as users of an educational application?
- RQ2:** What feedback type (i.e. reward-based or meaningful) is perceived as more useful for teachers and students?

To achieve the aim, we have used mixed methods for data collection (Section 4.2) to compare the responses of teachers and students during and after their interaction with the application, for instance, by contrasting participants' verbal responses to their gazes at UI objects on screen (Section 2.3).

3. DEVELOPMENT OF KNOWMA+

This section refers to the overall design of our educational application and its main features (described in Section 2). Knowma+ was built with usability principles in mind, involving iterative cycles and prototype testing where possible.

To clarify, some variations and synonyms are used across the document to refer to the same element. For example, points that can be obtained in the shape of stars are also mentioned as ratings; points that are earned for completing tests and challenges are displayed on the leaderboards, which sometimes can be referred also as rankings; feedback on questions for peer review can be called comments; results and resources given to students after taking a test are known as meaningful or instant feedback; and medals can be referred as badges or prizes.

3.1 Goal

Knowma+ (Knowing more by asking more [+]) has been developed by the first author of this paper to implement the concepts pertaining to learning by questioning (Section 2.1.1) and assessment (Section 2.1.2). The application is dedicated to supporting summative and formative assessment through teacher- and peer-evaluation with the use of gamification elements (Section 2.2) to enhance learning experience.

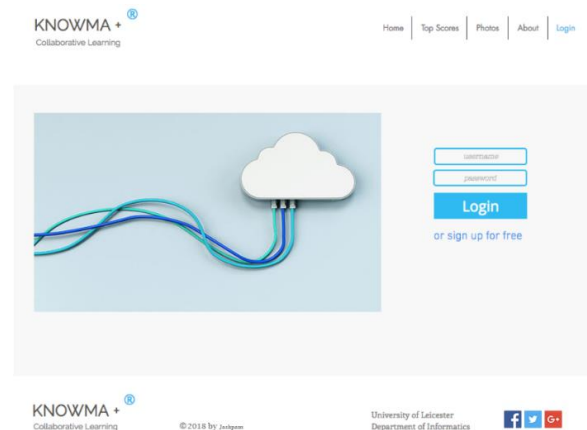


Figure 1: Knowma+ website's login area

The prototype supports three main roles:

- **Administrator:** This role has all the rights to create and adjust the structure of any educational institution. For instance by creating, modifying or renaming departments, categories, modules and users.
- **Teacher:** This role can comment, review or award badges to questions created by students. Teachers are able to create, reuse, share and manage tests, questions and topics. Other functionalities include

sending and receiving personal messages and modifying relevant settings on their modules. They are also able to customize an enhanced response to be instantly sent to students after a test.

- Student: This role can do regular educational tasks such as taking tests and receiving personalized feedback for their performance, including supportive resources and multimedia content. Additionally, students become initiators of reasoning and evaluation when they create their own questions and challenges to grade each other or to gain knowledge about specific academic topics. This effort and voluntary involvement can be rewarded with badges and meaningful feedback from teachers, comments from other students, and points that contribute to the ranking of individual students in the leaderboards.

Considering the results and limitations of previous research, we decided to analyse and include the following elements in Knowma+: points, badges, leaderboards and feedback.

Points and leaderboards are updated in two different ways: synchronously after taking a (teacher) test or a (student) challenge, as the scores are calculated immediately; or asynchronously, when students submit a question for peer-review and need to wait for an interaction to happen (i.e. rating).

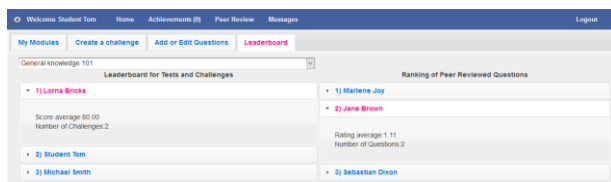


Figure 2: Student view of leaderboards by module

Feedback is also presented on two different cycles: immediately after taking a test, or once peer and teacher assessment (on questions created by students) has been provided. When students take a test, they receive information, resources and links related to their performance on each topic contained in that specific test. But when students create questions, they may need to wait to receive comments and remarks on the quality of the questions they have created and shared for a module (see Section 3.3 for more details).

In addition, teachers are the only ones that can award badges, after assessing the level of complexity of the questions created by their students. The medals obtainable are: gold, silver, bronze (or none) and they serve as special recognition which students may perceive as less valuable if they were awarded by peers.

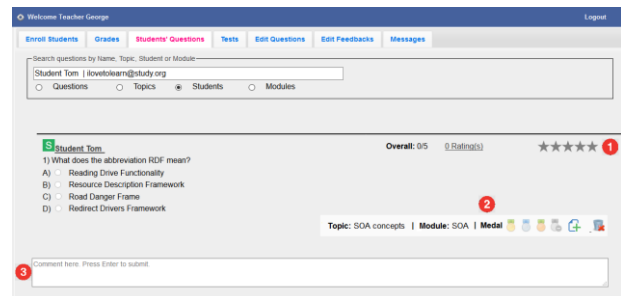


Figure 3: Teacher view of various gamification elements ([1] ratings, [2] medals, [3] comments) for questions created by students for peer and teacher evaluation.

3.2 Target groups

The target groups of Knowma+ are relatively broad, ranging from secondary school students up to university students and their teachers/lecturers. In the current user-based evaluation, we recruited a convenient as well as symbolic sample of genuine university students and practising lecturers.

3.3 Scenarios

As Knowma+ serves several educational purposes, it can be used in different ways depending on the user's role and goals. For example, a basic task can involve teachers enrolling students to their modules and awarding grades. To do the latter, teachers can create questionnaires for assessment.

When teachers create tests for students to take, different settings can be selected for that specific test (e.g. duration, instructions, passcode, availability). More importantly, when adding questions to a test, teachers can tag them with a topic (created by themselves in the past or shared from a colleague's module) that specifies the area to which the question belongs. If none of the existing topics fits the current question, teachers can generate a new one.

For this specific user-based evaluation we created two main modules with various general knowledge questions (e.g. "How much is the 20% of £800?" for "Basic Maths", or, "Where is the Acropolis located?" for "Geography") to minimize the impact of the content on user perception of Knowma+. For instance, questions on domain-specific knowledge can frustrate users lacking such knowledge, who may attribute their inability to respond to the application's design. This phenomenon has been studied with reference to the Attribution Theory (Depping & Mandryk, 2017).

Each topic can have as many levels as necessary for teachers to assess student performance. Teachers can write for each level different feedback and include additional files or web resources to encourage students to search and learn what they specifically need. This will be used later on to give automatic feedback messages to students

depending on the scores obtained for each of the topics included in a test (Figure 7).

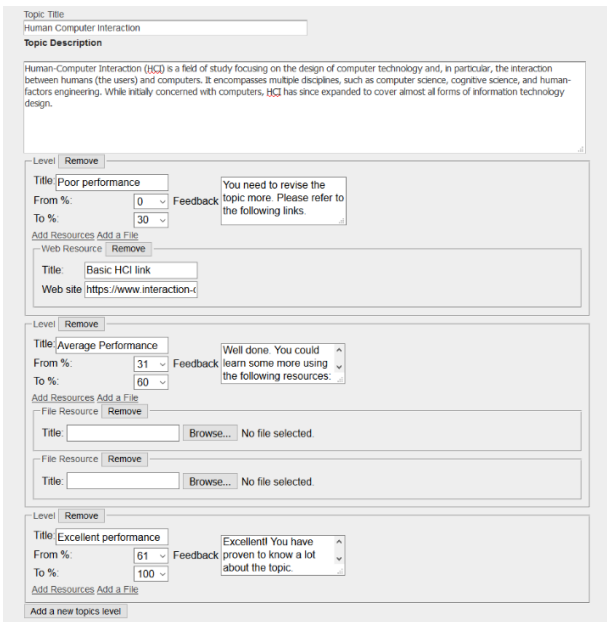


Figure 4: Example of feedback levels created by a teacher for a topic inside a module

In this way, after taking a test, students will not only be presented their score but will also receive more personalised feedback. One tab shows a summary of errors and grades (Figure 5) and another tab shows the information linked to all topics covered in a test (Figure 7). Therefore, resources displayed to students vary with the score they got on individual topics addressed in a test.

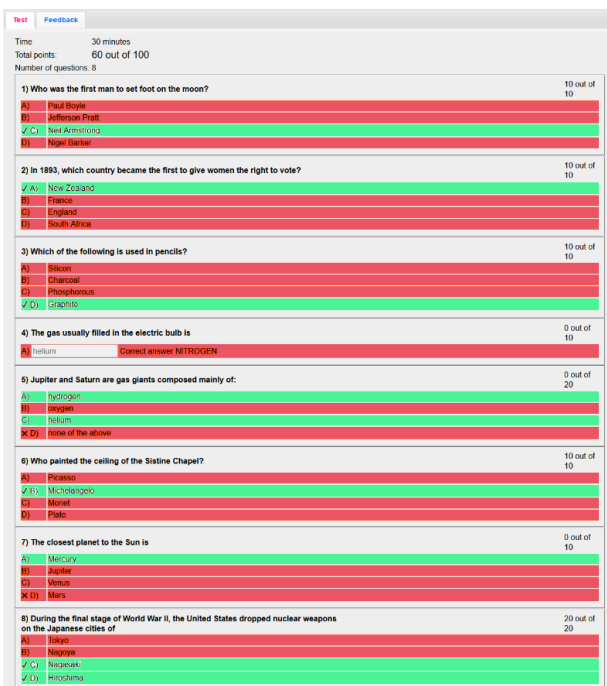


Figure 5: Screenshot of tabs with test results

Another option for students to gain more knowledge about specific topics –or areas of knowledge– is with the creation of ‘challenges’ and the publication of questions that they share with other peers. *Challenges* are a collection of questions created with a structure similar to normal tests, but with fewer features (e.g. passcode and topic creation are not allowed). *Single questions* are (stored to be reused in a challenge or) shared with others. When students create challenges, they can also tag individual questions with topics that have been created by the teacher of that module.

Questions created by students can be used in two ways: i) to earn points or ii) to receive feedback. When students respond to a challenge, they will obtain points for the accuracy of their answers. These points are used to contribute to their ranking in the module’s leaderboard (Figure 2). On the other hand, when students share single questions with their peers, those questions will be evaluated in terms of their quality and relevance to the topic while the quality of the answers to such questions is not evaluated. For this purpose, all students taking that module (and the teacher in charge) can either rate those questions out of five stars (and see an overall rating), or give comments and suggestions to those questions (Figure 3).

All past achievements by module are available for students to see. Notifications are also shown on this page (and as a number at the top menu) for newly acquired ones (i.e. badges, ratings, scores).

Other activities that can be executed in Knowma+ include: sending private messages as a means of communication between users, configuration of (some) personal details and password, editing questions and tests before submission, and search and reuse of content.

4. USER-BASED EVALUATION

The main purpose of this evaluation was to collect data in order to compare the acceptance of individual gamification elements available in Knowma+. With the mixed methods approach for gathering qualitative and quantitative data, we aimed to analyse behavioural and attitudinal responses of students and teachers when interacting with Knowma+ to understand their preferences for certain gamification elements and feedback styles.

4.1 Participants

For the user-based evaluation of Knowma+ we recruited 10 participants (5 teachers, 5 students). Although 7 of them were not native English speakers, all have studied or worked in the UK, therefore, they spoke and read English fluently. Four were female and the average age of the group was

29.5 years ($SD= 6.43$). Participants recruited had different academic backgrounds to ensure the heterogeneity of the sample: 6 people taught or studied in fields related to Computer Science, 2 in Engineering, 1 in Medicine, and 1 in Modern Languages.

Regarding the IT competence, on a 7-point scale (1: lowest; 7: highest), nine participants judged themselves as being quite experienced with computers, with the remaining participants self-assessed as below average ($Mean=5.2$, $SD=1.03$). Additionally, eight people have used educational software before, but in general participants have not frequently used applications for any learning purpose (Never= 2; Yearly= 4; Monthly= 1; Weekly= 3; All the time= 0).

Their participations were voluntary without involving any kind of compensation.

4.2 Instruments

The mixed methods approach allows researchers to analyse numerical data typical of quantitative research combined with verbal data customary for qualitative research, in order to gain more in-depth insight of a phenomenon (Williams, 2011). We used 4 different methods to collect data from participants:

- (i) **Eye tracking:** Eye movement research enables the capture of diverse interactions with system interfaces, such as fixation durations and the sequence in which users look at specific areas of interest at any given time (i.e., scanpath) (Margolis & Pauwels, 2011). Tobii T60 (desktop) and Tobii Studio 3.3.2.1150 (software analysis) were used. We recorded participants' gaze data on the screen to infer usability problems as well as user preferences of the gamification elements visualised in the application being evaluated.
- (ii) **Observation:** Researchers gather qualitative data by taking notes of what they hear and see in a specific context. This method also allows the collection of quantitative data with the use of a coding scheme for defining specific behaviours of which frequencies are tallied (Bernard, Wutich & Ryan, 2016). Participants in this evaluation were not only observed while interacting with the application, but they were also asked to think aloud while doing so, with the opportunity for the researcher to ask quick questions to clarify their verbal as well as gestural behaviours.
- (iii) **Questionnaires:** Two standardized questionnaires were deployed to evaluate perceived usability (SUS) and user engagement (UES-SF). The System Usability Scale (SUS) (Brooke, 1996)

consists of 10 statements with each being rated with a 5-point Likert scale. The User Engagement Scale - Short Form (UES-SF) (O'Brien, Cairns & Hall, 2018) is a 12-item scale with a four-factor structure (i.e., Focused Attention, Perceived Usability, Aesthetic Appeal, and Reward).

- (iv) **Interview:** While trying to remain impartial and non-judgmental to responses, an interviewer asks a series of questions to interviewees in order to delve into certain aspects of the matter being explored (Tashakkori & Teddlie, 2010). This method is very useful for clarification and confirmation after concluding an evaluation. We conducted semi-structured interviews with mixed (open and close-ended) questions to gather qualitative and quantitative data about usability and preferences.

4.3 Procedure

The user-based evaluation was conducted on an individual basis. Participants were divided into two groups according to their real-life roles. As part of the introduction, they were asked to fill in a consent form and a short survey with demographic questions. Participants also received a brief explanation about the purpose of the application and were given the chance to ask any questions. Afterwards, the eye-tracker had to be set up and calibrated for recording users' eye movements when interacting with Knowma+.

For the evaluation, participants were asked to follow a list of instructions. Students had 12 tasks and teachers 10. In both cases, the tasks were the same for each group of participants but randomised when interacting with the different types of gamification elements to avoid any bias in their responses due to the order effect. During the individual session, participants were asked to think aloud while they were interacting with Knowma+. The experimenter, who was present in the testing room throughout the session, could then take notes about participants' comments and also note down other observations.

Table 1: Average duration t (in minutes) per phase per user role (Note: Q&I = Questionnaire and Interview)

Role	Intro	1 st part	Break	2 nd part	Q&I	Avg. (t)
S	12	22	10	12	16	72
T	11	27	-	-	17	55

The student group session was split in two (with a short break of 10 minutes in between) as the questions created by this role had to be evaluated by another (teacher) user. The teacher group

worked through the tasks in a single session (see Table 1).

After finishing using Knowma+, participants filled in standardized questionnaires (Section 5.3). Then they were debriefed about the goals of this evaluation along with an explanation of what gamification is. Participants were also given the opportunity to ask new questions and at the end of the session they responded to a semi-structured interview, which was audiotaped.

5. RESULTS AND DISCUSSION

In this section we are going to analyse data relevant to the individual gamification elements of Knowma+ as well as the usability problems found with the user-based evaluation of the application.

As before discussed (Section 4.2), there were four different sources of data to address the two research questions (Section 2.4):

5.1 Eye-tracking

This technique was used to associate data collected from the participants' gaze data with the results from other methods in order to identify usability issues (Section 5.2) and gather improvement suggestions on these problems (Section 5.4).

We also analysed gaze data to compare 3 gamification elements included on a single page of Knowma+: ratings, medals, and comments. The leaderboards and the meaningful feedback were part of different zones in the application so they had to be compared and evaluated through different methods.

Results of *time to first fixation* (i.e. an eye-tracking metric) from all 10 users indicated that among the three Areas of Interest (AOIs) – comments, medals and ratings - participants looked first at the ratings with a mean of 2.37 seconds after the webpage (i.e. the stimulus; Figure 3) was shown. Closely followed by the medals with 4.76 seconds. At last were the comments with a mean of 15.06 seconds. However, no significant difference, as shown by the result of the non-parametric Friedman test ($\chi^2_{(2)} = 4.2, p > 0.5$), was found regarding the *first fixation duration* (i.e. duration of the first fixation on an AOI) for each of these gamification elements ($Mdn_{ratings} = 0.17$, range = 0.14-0.24; $Mdn_{medals} = 0.15$, range = 0.09-0.17; $Mdn_{comments} = 0.10$, range = 0.08-0.14).

We did not take into account for this evaluation other metrics such as the *total fixation duration* as we asked participants to carry out some tasks that could have interfered with the time they spent on an AOI (e.g. to give comments on a question takes longer than to give a medal).

In addition to these metrics, heatmaps provide us with visual representations of the relative intensity of gaze at different areas of a page in Knowma+. In Figure 6 we can see the favourite spots of the teachers when navigating on the Peer and Teacher Review of single questions created by students (to share with others). The areas with the highest gaze intensity on that page were –apart from the questions themselves- the ratings and medals. In Figure 7 we can see the heatmap of students' gazes when reading the feedback received after completing a test. The areas with the highest gaze intensity were the feedback and the resources and websites attached to that topic.



Figure 6: Heatmap of gamification elements from a teacher point of view



Figure 7: Heatmap of instant feedback for students after taking a test

5.2 Thinking aloud

Verbal data from think-aloud enable us to understand better the gaze data (Section 5.1). With the help of eye-tracking data we were able to infer some usability issues. For instance, participants that did not look at the correct areas on the page where they were trying to accomplish a particular task were more likely to express this difficulty out loud, evidencing the connection between data from both methods.

Participants were later on able to explain their motives to enjoy or dislike certain UI objects while interacting with the application during the interview (see Section 5.4 for more details).

Usability problems gathered from observation and think-aloud were grouped by the type of issue identified (adapted from Bruun & Stage, 2015): (i) interface design and navigation issues causing time delays; (ii) software bugs; (iii) test moderator's assistance needed for completing a task; and (iv) confusion or errors due to the mismatch between the user's mental model and the application's design (e.g., users were lost as they could not make sense of the interaction sequence).

Also, usability problems and their importance were classified (Table 2) in three levels (Picard, 2000). 'Low' for issues that might have affected the overall sense of the quality of the interface, but would not hinder participants significantly from completing their tasks. 'Medium' for issues that may have confused, delayed or distracted users briefly. And 'High' for issues that were an obstacle for participants, either preventing them from completing their tasks, or causing some significant delay, disruption, confusion or annoyance.

Table 2: A combined list of problems encountered by all users according to the impact on usability

Usability problems N=41	Impact on users		
	Low	Medium	High
Delays due to the interface design	6	7	3
Software bugs	1	1	1
Explanation/help needed	2	5	-
Mismatched mental model	10	4	1
Total = 41	19	17	5

5.3 Questionnaires

We asked participants to fill in both the SUS and the UES-SF (Section 4.2).

The average SUS score is 77 ($SD= 10.19$), suggesting that the usability of Knowma+ is reasonable. The statements where the application scored the lowest were related to the learnability of the system (items 4, 8 & 9), indicating that effectiveness and efficiency could be enhanced with help and documentation (Nielsen, 1994) such as a short video tutorial.

The overall UES-SF score (in a five-point scale) for all participants is 3.71 ($SD= 0.67$) with the strongest dimensions being PU (Perceived Usability) with an average score of 4.33 and RW (Reward) with a mean of 4.20. These results indicate that participants felt highly interested and involved, and the application was perceived as engaging (O'Brien et al., 2018). The dimension with the lowest score was AE (Aesthetic Appeal) with an average of 3.1 points, suggesting that some improvements are needed on the visual appeal and attractiveness of the application's interface (see Sections 5.2 & 5.4 for more details).

5.4 Interview

Findings were inferred from the experimenter's observational notes (Section 5.2) as compared to the answers in the interview. Combining these methods we were able to realise that:

- Students who seemed self-motivated were more likely to try harder to obtain better rewards and grades (e.g. by spending extra time creating and answering questions), whereas students that seemed less interested in the tasks did not push boundaries to overcome their limitations (e.g. their score on the leaderboards).
- Half of the participants found the task of assessing the quality of questions created for peer review difficult to do without some predefined criteria to follow. Especially when talking about badges, the participants thought it would have been advantageous to provide a set of well-defined rules and criteria.
- Most participants skimmed through the whole list of students' questions before starting giving comments/ratings/medals to compare their quality. For instance, teachers usually gave longer comments (e.g. with improvement suggestions) to students' questions that were considered as "easy". Instead, students generally gave shorter comments to those questions and longer responses (e.g. stating reasons for liking them, or explaining how such formulation made them reflect more) to questions assessed as more "difficult".
- The majority of students (4 out of 5) ranked the meaningful feedback as most useful for learning, but they did not agree on a most attractive gamification element.
- On the other hand, (4 out of 5) teachers thought the best gamification element was the ratings, but did not agree on which specific element or if the meaningful feedback was most useful for learning.
- Both teachers and students (6 out of 10) rated badges as the gamification element they liked the least.
- Just one participant suggested the inclusion of other gamification elements (statistics and a progress bar) and the remaining participants thought that the selected elements were enough for this application.
- 9 out of 10 participants said that they would like to use this (or a similar) application in their usual teaching/learning practices.

Also, most participants agreed on the same suggestions on how to improve the application, by:

- Altering certain objects of the interface (e.g. the appearance of the leaderboards on Figure 2, or the colour scheme on Figure 5).
- Adding new objects to the UI (e.g. a button to submit comments, Figure 1).
- Redesigning the flow and look of the page for creating tests.
- Improving the presentation of (old and new) achievements earned by students.
- Including documentation or tips to help users to find out more about the features of the application.

5.5 Discussion

While some findings indicate that graphic gamification elements could capture the user's initial attention, there were slight differences between the self-reported preferences of teachers and students with regards to the perceived usefulness of each of the gamification elements and the meaningful feedback.

Preliminary results based on eye-tracking data (Section 5.1) suggest that the most attractive gamification element in Knowma+ was the *rating* of peer-review questions (RQ1, Section 2.4). For the teachers, this observation matches with the findings from the self-reported data collected through the other methods (Sections 5.2 & 5.4). However, this predilection does not coincide with that of the students who could not decide on a most attractive and enjoyable gamification element. Similarly, although participants rated the attractiveness and effectiveness of *badges* rather low, it captured their high attention as shown by the eye-tracking data.

On the other hand, although results based on the think-aloud data (Section 5.2) indicate that all participants agreed on the convenience of having meaningful feedback and comments, teachers and students had dissimilar opinions about its usefulness (Section 5.4). The teachers seemed more attracted to the benefits that the *reward-based* gamification could have for engaging students, especially with the use of stars to rate student work, while the students rated *meaningful feedback* as most useful for learning in real-life scenarios (RQ2, Section 2.4).

6. CONCLUSION

The contribution of our work presented in this paper is twofold: (1) the improvement of applied knowledge for the gamification research by undertaking a user-based evaluation to compare different gamification elements in an educational application. This helps to address the issue on the lack of empirical evidence for the specific effects that individual gamification elements can have on user attention in educational environments; (2) The development of the

educational application Knowma+, which is grounded in relevant pedagogical concepts and has the potential to motivate learners to know more by asking more and better questions.

The findings from our user-based evaluation can be of practical interest to educationalists from secondary schools up to universities looking for viable means to enhance similar learning experiences in their students. Our results can also be relevant to researchers interested in understanding the attractiveness of individual gamification elements used in educational applications.

6.2 Limitations and future work

Plenty has been done in the field of technology-enhanced learning. However, most software applications are mainly developed for adult learners. Likewise, when speaking about gamified educational systems, most research has dedicated to studying the effect of gamification in few specific contexts. For instance, analyses of gamification in educational environments (Sousa Borges et al., 2014) indicate that gamification has been used to support Higher Education at least five times more than for Elementary Education. Hence, the potential of digital technologies with gamified approaches for secondary schools has been scarcely analysed, implying that more research in the area needs to be conducted.

Our work is limited in this study by the number of gamification elements we used in Knowma+. Different combinations of elements could help to assess the value of each group or specific elements in the context of gamification in learning. In addition, as the sample size for this study was small, for our future research work we aim to recruit a larger number of participants, thereby enabling us to draw more solid conclusions. Apart from increasing the sample size, we plan to involve participants with different age groups, given that Knowma+ is designed to support a broad range of learners.

To address the issue of ecological validity, it will be beneficial to evaluate the hypothesized effects of individual gamification elements in real-life educational environments. This is to check to what extent the findings reported of this paper can be replicated with different teachers and learners in different contexts. Additionally, the learning effect of the gamification elements can be measured with a systematic control with pre- and post-knowledge tests to analyse if they can make actual impact when they are implemented in classroom.

Our future work will also aim to improve Knowma+ iteratively with systematic user-based evaluations, focusing on both the cognitive and experiential aspect of learning and teaching that this viable gamification tool aims to support.

7. REFERENCES

- Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2013, May). Steering user behavior with badges. In Proceedings of the 22nd international conference on World Wide Web (pp. 95-106). ACM.
- Arnab, S., Lim, T., Carvalho, M. B., Bellotti, F., Freitas, S., Louchart, S. & De Gloria, A. (2015). Mapping learning and game mechanics for serious games analysis. *British Journal of Educational Technology*, 46(2), 391-411.
- Attali, Y., & Arieli-Attali, M. (2015). Gamification in assessment: Do points affect test performance?. *Computers & Education*, 83, 57-63.
- Banfield, J., & Wilkerson, B. (2014). Increasing student intrinsic motivation and self-efficacy through gamification pedagogy. *Contemporary Issues in Education Research (Online)*, 7(4), 291.
- Becker, K., & Nicholson, S. (2016). Gamification in the Classroom: Old Wine in New Badges. *Learning, Education and Games*, 61-67.
- Beldarrain, Y. (2006). Distance Education Trends: Integrating new technologies to foster student interaction and collaboration. *Distance Education*, 27(2), pp.139-153.
- Bernard, H. R., Wutich, A., & Ryan, G. W. (2016). *Analyzing qualitative data: Systematic approaches*. SAGE publications.
- Black, P. and Wiliam, D. (2010). Inside the Black Box: Raising Standards through Classroom Assessment. *Phi Delta Kappan*, 92(1), pp.81-90.
- Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4-7.
- Bruun, A., & Stage, J. (2015, September). An empirical study of the effects of three think-aloud protocols on identification of usability problems. In *Human-Computer Interaction* (pp. 159-176). Springer, Cham.
- Buehl, D. (2013). *Classroom strategies for interactive learning*. Newark, USA: International Reading Association.
- Bunchball, I. (2010). Gamification 101: An introduction to the use of game dynamics to influence behavior. White paper, 9.
- Burbules, N. (2018). *Watch IT: The risks and promises of information technologies for education*. Routledge.
- Cheong, C., Filippou, J., & Cheong, F. (2014). Towards the gamification of learning: Investigating student perceptions of game elements. *Journal of Information Systems Education*, 25(3), 233.
- Colburn, A. (2006). What teacher educators need to know about inquiry-based instruction. Paper presented at the annual meeting of the Association for Education of Teachers in Science, Akron, OH.
- Deci, E. L., & Ryan, R. M. (2010). *Self-determination*. John Wiley & Sons, Inc..
- Depping, A. E., & Mandryk, R. L. (2017, May). Why is This Happening to Me?: How Player Attribution can Broaden our Understanding of Player Experience. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 1040-1052). ACM.
- Deterding, S. (2012). Gamification: designing for motivation. *interactions*, 19(4), 14-17.
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011, September). From game design elements to gamefulness: defining gamification. In Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments (pp. 9-15). ACM.
- Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in education: a systematic mapping study. *Journal of Educational Technology & Society*, 18(3), 75.
- Ecclestone, K., Davies, J., Derrick, J. and Gawn, J. (2010). *Transforming formative assessment in lifelong learning*. Berkshire, England: Open University Press.
- Ehmke, C., & Wilson, S. (2007, September). Identifying web usability problems from eye-tracking data. In Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it-Volume 1 (pp. 119-128). British Computer Society.
- Erneling, C. (2010). *Towards discursive education*. New York: Cambridge University Press.
- Goshevski, D., Veljanoska, J., & Hatziapostolou, T. (2017, September). A Review of Gamification Platforms for Higher Education. In Proceedings of the 8th Balkan Conference in Informatics (p. 28). ACM.
- Hamari, J., Koivisto, J., & Sarsa, H. (2014, January). Does gamification work?--a literature review of empirical studies on gamification. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on* (pp. 3025-3034). IEEE.
- Heintz, S., & Law, E. L. C. (2015). Game Elements-Attributes Model: a First Step towards a Structured Comparison of Educational Games. In *DiGRA Conference*.
- Iosup, A., & Epema, D. (2014, March). An experience report on using gamification in technical higher education. In Proceedings of the

- 45th ACM technical symposium on Computer science education (pp. 27-32). ACM.
- Koster, R. (2005). *A theory of fun for game design*. Scottsdale, AZ: Paraglyph Press.
- Landers, R. N., Bauer, K. N., & Callan, R. C. (2017). Gamification of task performance with leaderboards: A goal setting experiment. *Computers in Human Behavior*, 71, 508-515.
- Lehmann, C. and Chase, Z. (2015). *Building school 2.0*. John Wiley & Sons.
- de-Marcos, L., García-López, E., García-Cabot, A., Medina-Merodio, J. A., Domínguez, A., Martínez-Herráiz, J. J., & Díez-Folledo, T. (2016). Social network analysis of a gamified e-learning course: Small-world phenomenon and network metrics as predictors of academic performance. *Computers in Human Behavior*, 60, 312-321.
- Margolis, E., & Pauwels, L. (2011). *The Sage handbook of visual research methods*. Sage. (pp.433-435)
- Mekler, E. D., Brühlmann, F., Tuch, A. N., & Opwis, K. (2017). Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior*, 71, 525-534.
- Mekler, E. D., Brühlmann, F., Opwis, K., & Tuch, A. N. (2013, October). 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* (pp. 66-73). ACM.
- Muntean, C. I. (2011, October). Raising engagement in e-learning through gamification. In *Proc. 6th International Conference on Virtual Learning ICVL (Vol. 1)*. sn.
- Nacke, L. E., & Deterding, S. (2017). The maturing of gamification research. *Computers in Human Behavior*, 71, 450-454.
- Nah, F. F. H., Telaprolu, V. R., Rallapalli, S., & Venkata, P. R. (2013, July). Gamification of education using computer games. In *International Conference on Human Interface and the Management of Information* (pp. 99-107). Springer, Berlin, Heidelberg.
- Nicholson, S. (2015). A recipe for meaningful gamification. In *Gamification in education and business* (pp. 1-20). Springer, Cham.
- Nielsen, J. (1994, April). Enhancing the explanatory power of usability heuristics. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 152-158). ACM.
- O'Brien, H., Cairns, P., & Hall, M. (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, 28-39.
- Orquin, J. L., Ashby, N. J., & Clarke, A. D. (2016). Areas of interest as a signal detection problem in behavioral eye-tracking research. *Journal of Behavioral Decision Making*, 29(2-3), 103-115.
- Picard, R. W. (2000). Toward computers that recognize and respond to user emotion. *IBM systems journal*, 39(3.4), 705-719.
- Poole, A., & Ball, L. J. (2006). Eye tracking in HCI and usability research. *Encyclopedia of human computer interaction*, 1, 211-219.
- Rebolj, V. (2009). E-education between Pedagogical and Didactic Theory and Practice. *Organizacija*, 42(1).
- Richards, C., Thompson, C. W., & Graham, N. (2014, October). Beyond designing for motivation: the importance of context in gamification. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play* (pp. 217-226). ACM.
- Robson, K., Plangger, K., Kietzmann, J., McCarthy, I. and Pitt, L. (2015). Is it all a game? Understanding the principles of gamification. *Business Horizons*, 58(4), pp.411-420.
- Roth, S., Schneckenberg, D., & Tsai, C. W. (2015). The ludic drive as innovation driver: Introduction to the gamification of innovation. *Creativity and Innovation Management*, 24(2), 300-306.
- Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior*, 69, 371-380.
- Seel, N. (2012). *Encyclopedia of the sciences of learning*. Germany: Springer.
- Shaffer, D. W. (2006). How computer games help children learn. Macmillan.
- Shin, N. (2006). Online learner's 'flow' experience: an empirical study. *British Journal of Educational Technology*, 37(5), 705-720.
- Simões, J., Redondo, R. D., & Vilas, A. F. (2013). A social gamification framework for a K-6 learning platform. *Computers in Human Behavior*, 29(2), 345-353.
- de Sousa Borges, S., Durelli, V. H., Reis, H. M., & Isotani, S. (2014, March). A systematic mapping on gamification applied to education. In *Proceedings of the 29th Annual ACM Symposium on Applied Computing* (pp. 216-222). ACM.
- Spector, J. M. (2000). *System dynamics and interactive learning environments: Lessons*

- learned and implications for the future. *Simulation & Gaming*, 31(4), 528-535.
- Tashakkori, A., & Teddlie, C. (Eds.). (2010). *Sage handbook of mixed methods in social & behavioral research* (pp 305-306). Sage.
- Tondello, G. F., Premasukh, H., & Nacke, L. E. (2018). A Theory of Gamification Principles Through Goal-Setting Theory. In *Proceedings of the 51st Hawaii International Conference on System Sciences (HICSS)*.
- Tynjälä, P. (1999). Towards expert knowledge? A comparison between a constructivist and a traditional learning environment in the university. *International journal of educational research*, 31(5), 357-442.
- Wang, H., & Sun, C. T. (2011, September). Game reward systems: Gaming experiences and social meanings. In *DiGRA Conference*.
- Werbach, K., & Hunter, D. (2012). *For the win: How game thinking can revolutionize your business*. Wharton Digital Press.
- West, D. (2012). *Digital schools*. Washington, D.C.: Brookings Institution Press.
- Williams, C. (2011). Research methods. *Journal of Business & Economics Research (JBER)*, 5(3).
- Zagal, J. P., Mateas, M., Fernández-Vara, C., Hochhalter, B., & Lichti, N. (2007). Towards an ontological language for game analysis. *Worlds in Play: International Perspectives on Digital Games Research*, 21, 21.
- Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in web and mobile apps*. O'Reilly Media, Inc.