How (not) to Introduce Badges to Online Exercises

Lassi Haaranen, Petri Ihantola, Lasse Hakulinen and Ari Korhonen
Department of Computer Science and Engineering
Aalto University, School of Science
Espoo, Finland
firstname.lastname@aalto.fi

ABSTRACT
Achievement badges are increasingly used to enhance educational systems and they have been shown to affect student behavior in different ways. However, details on best practices and effective concepts to implement badges from a non-technical point of view are scarce. We implemented badges to our learning management system, used them on a large course and collected feedback from students. Based on our experiences, we present recommendations to other educators that plan on using badges.

Categories and Subject Descriptors
K.3.2 [Computers and Education]: Computer and Information Science Education—Computer science education

Keywords
Gamification, Achievement badges, Computer science education

1. INTRODUCTION
Gamification is a fairly new and rapidly growing field defined by Deterding et al. [5] as “the use of game design elements in non-game contexts.” Motivation behind this is to bring elements which are considered fun and engaging into tasks which might not inherently possess those qualities. Thus, perhaps, improving the user’s experience and leading to an enhanced performance, which can be anything from exercises submitted earlier than necessary to increased activity in a forum. Interest in gamification has increased over the past few years and it has even ‘earned’ its place in the Gartner’s Emerging Technologies 2012 Hype Cycle nearing the peak of inflated expectations [9]. Whether it can move in the Hype Cycle past the peak and ‘trough of disillusionment’ into something that provides considerable value remains to be seen.

A common way to bring game-like elements to educational systems is by adding badges (aka achievements, achievement badges, or trophies). Badges can be defined as optional sub-goals which are not required for completion of the main task [13]. Badges in gamification have their parallels in the real world, for instance, in the form of boy scout badges, which are awarded for acquiring a new set of skills or mastering certain skill levels. Even airline bonus point systems, which aim to encourage customer loyalty, employ game-like mechanics. If these real world achievements and points are considered gamification, then the concept dates further back than the actual name gamification.

Education in the past centuries has already had certain game like elements embedded into the system. Students are awarded points in tests and in different types of assignments. In universities, students complete courses and credit by credit work towards unlocking the last challenge – the final thesis. There are also certain known problems in education, such as procrastination of studying until the deadline and motivational problems in general. Gamification (e.g. badges), at least in the minds of its proponents, could offer some alleviation to these problems.

The purpose of this study is to increase understanding how badges may affect students when introduced to a large course where automatic assessment is already in place. Even though automatic assessment is a valuable tool, especially in large courses, it also might bring some negative side effects. For example, it can promote bad learning habits such as trial-and-error in some exercise types. In addition, we are looking for novel ways to teach some latent learning goals in our courses such as time management. Our research question related to this is:

How will the students react to gamification features such as badges and will it have any positive or negative effect on their behavior? We are especially interested in their attitudes towards badges.

Previous research related to this question is described in Section 2. After that, data collection and the research methods are described in Section 3. The course where badges were introduced had eight rounds of automatically assessed exercises and the badges were used only on the last four rounds. Right after the course we sent a survey to all students to learn their reactions. The survey results and other data we collected are described in Section 4. The results and badges in general will be discussed in Section 5. Finally, in Section 6 we make some recommendations based on our observations described in previous sections.
2. RELATED WORK

Gamification is closely related to toys, playful design and serious games. As illustrated in Figure 1, Deterding et al [5] separate gamification from these other terms by applying two axis: 1) designing something that is complete in itself vs. adding gameful or playful elements to something that already exists and 2) games vs. toys. Caillois [2] makes the distinction between game and play using terms ludus which encompasses games (usually with specific rules) and paidia which refers to more freeform play. The whole versus parts axis encapsulates the idea whether something implements a whole idea of either paidia or ludus or just parts of it. On the whole side of the spectrum are games and toys which focus completely on their core function.

Figure 1: Gaming versus playing and whole versus parts [5]

Gamification has been used in a variety of educational systems. An example of gamification in an educational system, which is open to everyone, is Khan Academy¹. In Khan Academy, badges are awarded to users, for example, for watching videos and completing exercises. Another free educational service that uses gamification is Duolingo². It aims at teaching a foreign language. Points and levels are used to track the users’ progress in learning a language. Points, which are represented by golden coins, are awarded for completing exercises and learning new words.

Clearly there are many kinds of badges. They look different and are rewarded for different reasons. Hamari et al. [8] have presented a method of describing badges. They divide them in three separate parts: signifier, completion logic, and reward. Signifier consists of the visual image and name of the badge. They found three categories for rewards: achievement (meta-)game, in-game, and out-game. The first category of rewards is part of a bigger meta-game (such as the Xbox Live’s gamerscore) of collecting badges while the second one is a reward that stays within the game. The last category is external rewards that exist outside the system. An example of rewards in the last category would be to offer badges which they could earn by doing exercises with certain additional restrictions, such as completing the exercises a full week before deadline or getting exercises completely correct on the first try. Quantitative analysis of the results revealed that student’s behavior can be affected, at least in some cases, with badges. An example of affected behavior was submitting correct answers with fewer attempts. However, significant effect on the grades were not observed between the treatment and control groups.

A gamification study done on a course for “Qualification for users of ICT” found qualitative and quantitative differences between control (N=80) and treatment (N=131) groups [6]. 58 students from treatment group opted for the gamified version. Those students had higher scores in initial activity and practical exercises in various modules. However, they found that the treatment group had lower scores on the final examination of the course and the participation, which was also scored. In their setup, students were able to choose whether or not they would participate in the gamified version of the course, which might have had a significant impact on the results. They argue that the gamified version improved the practical competence but at the same time had a negative effect on understanding theoretical concepts. Qualitatively they also found that gamification has a significant emotional and social impact. While some students liked public leaderboards and found them motivating, some students disliked the leaderboards and did not appreciate the competitive element they created.

University of Central Florida’s course, “Adventures in Emergent Media”, implemented badges for their online course in which part of their grade was determined by how many achievements they got during the course [12]. After the course, they held a survey about the course in general and badges in particular to which 138 students responded. Although, they found the effects of badges to be slightly positive, many students reported frustration because some of the badges were hidden and had to be discovered, which proved to be too difficult in some cases. They also noted a social aspect to badges where half of the students said they were motivated to get a badge if they noticed that one of their friends had achieved it.

Denny conducted a large-scale (N > 1000) controlled experiment on a course where they implemented badges to PeerWise platform, which is aimed at students creating and answering questions [4]. They found a positive motivational effect. The treatment group was active during more days and answered more questions. However, badges did not have an effect on the number of questions created. On a survey after the course, the majority of students preferred the system with badges. They did not observe any negative effects, and thus conclude that badges can be useful motivators in educational contexts.

3. METHODOLOGY

This is a follow-up study to the aforementioned study in Aalto University [7]. We implemented badges to a course that is described in Section 3.1. To map out how students perceived badges, we collected and analyzed data from various sources as described in Section 3.2.

3.1 Course description

The data was collected from a 5 ECTS Data Structures and Algorithms course taught at Aalto University in spring
2013. The course is a bachelor level course with basics of programming as prerequisites. Part of the course were individual online exercises that are assessed and graded automatically. For the CS major students, the online exercises contribute 20 per cent to their final course grade and for the minor students, the online exercises contribute 30 per cent to their final grade.\(^3\) Online exercises were divided into eight rounds with distinct deadlines. Separate rounds did not have any minimum requirements, but only the sum of the total points from all rounds was used to determine the grade. 50 % of the maximum points was required to pass the exercises with grade 1 and 60, 70, 80 and 90 % of total points resulted correspondingly to grades 2, 3, 4 and 5. Thus, students were able to pass the exercises by doing only the first four rounds with full points. There were several exercises on each round and resubmissions were also allowed.

Badges were added as a plugin to A+ learning management system\(^10\) that we used to deliver the online assignments. On the first four exercise rounds, there were no badges at all and students were unaware that they would be revealed after round four. Starting from round five, students could achieve various badges as described in Section 3.1.1. Badges were visible on the side bar when viewing course schedule, exercise page or submissions for an exercise. Only newly achieved badges were shown, or if no new badges were achieved the latest achieved badge was visible. In addition to the view in the side bar, a summary page to visualize all achieved and available badges was also implemented. These are illustrated in Figure 2.

### 3.1.1 Completion Logic and Reward for Badges

Badges were awarded in three different categories: learning, time management, and carefulness. Basis for the different categories for the badges were derived from the previous research\(^7\). All of the badges come in three different levels (bronze, silver, gold) based on an increase in difficulty in completion logic and stricter criteria. Badges did not have any external rewards and they did not affect the course grade in any way. Since the badges could not be shared with friends via social media they did not have social aspect to them\(^5\).

In the learning category, badges were achieved by completing the exercises with 50, 75 or 100 % of the score from the rounds maximum score. A factor that affected the completed score of the round was that submissions which arrived after the deadline suffered a 50 % penalty to the score.

Time management badges encouraged submitting earlier. They were awarded for completing 50 % of the exercise round’s score at least a week (gold), three days (silver) or a day (bronze) before the deadline. Learning category badges were somewhat dependent on time management because late submissions were allowed, but penalized 50 %.

Final category of badges was carefulness. These badges were awarded for completing exercises with perfect score and no more than one submission per exercise, meaning that the student needed to submit a perfect answer on the first try. The aim was to encourage the students to submit well thought out answers on the first try and subsequently curb trial-and-error method of solving problems. The levels to achieve bronze, silver or gold were 5, 10 or 15 respectively. The total number of exercises which counted towards this badge was 23.

### 3.2 Research Method

After the course, a link to an anonymous online survey was sent to all students. A total of 162 students (53 %), out of 306 registered to the course, responded to the survey that contained multiple choice questions and an option for freeform feedback. Results of this survey are presented in Section 4.1. We used data based content analysis on all the textual answers by manually and iteratively grouping them to categories that emerged from the data\(^11\).

In addition to the survey, we also had the students’ submissions, submission times, and feedback (including points) that a student got from his or her submission. Moreover, every time a user viewed the badge summary view (on right in Figure 2) was also recorded. 252 students (82 %) passed the online assignments with at least minimum points. The log data dissected in Section 4.2 focuses on these students.

### 4. RESULTS

#### 4.1 Survey

Numeric questions used a likert scale from 0 (completely disagree) to 4 (strongly agree) with 2 as the neutral option. Translated questions and answers to those are presented in Table 1.

A majority of the students liked the look of the badges and about half of the students were satisfied with the completion logic. Correspondingly, most of the students did not find that the badges disturbed their work. About one third of the students agreed with the statement that badges were motivating, one third disagreed, and one third simply thought they were neither motivating nor demotivating. Still, 52 % of students were in favor of having badges in the next year’s course while 20 % were opposed to that.

Altogether 88 students responded with written feedback to the last question of “Please give additional comments about the badges”. Approximately half of those felt indifferent towards the badges. A quarter of students gave positive feedback while ten assessed them negatively. This was also reflected in the numerical answers where the majority of students did not find achievement badges particularly motivating or demotivating. We analyzed the written feedback and found some common themes in them.

#### Suggesting improvements

A lot of the responses contained ideas and suggestions for improvement. Seven of the suggestions mentioned an UI glitch, where two exercises in earlier rounds were partially covered by achievement badges on some browsers. In addition to problems with UI, four students did not notice badges or were confused and not certain why they got badges or if it affected course grades in any way, commenting: “You could never really know when one of them would appear.”\(^5\)

#### Badges did not impress

28 of the students reported that they were indifferent towards badges or “don’t see the usefulness of achievement badges.”

\(^3\)The rest of the final grade is determined by other projects and the final examination.

\(^4\)Technically sharing the URL for the summary page was possible. But there was no designed way to share achieved badges.

\(^5\)Quoted feedback is translated
Table 1: Numeric feedback regarding achievement badges (N=162)

<table>
<thead>
<tr>
<th>Question</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>I found the badges motivating</td>
<td>20%</td>
<td>24%</td>
<td>17%</td>
<td>24%</td>
<td>14%</td>
</tr>
<tr>
<td>Badges disturbed my work</td>
<td>61%</td>
<td>15%</td>
<td>15%</td>
<td>6%</td>
<td>3%</td>
</tr>
<tr>
<td>Trying to achieve badges had an effect on my behavior</td>
<td>40%</td>
<td>21%</td>
<td>13%</td>
<td>15%</td>
<td>11%</td>
</tr>
<tr>
<td>Visual look of the badges was good</td>
<td>3%</td>
<td>6%</td>
<td>22%</td>
<td>42%</td>
<td>27%</td>
</tr>
<tr>
<td>I was satisfied with the criteria for awarding badges</td>
<td>3%</td>
<td>8%</td>
<td>42%</td>
<td>29%</td>
<td>18%</td>
</tr>
<tr>
<td>I think that badges should be used in A+ for the next year’s course as well.</td>
<td>8%</td>
<td>12%</td>
<td>28%</td>
<td>22%</td>
<td>30%</td>
</tr>
</tbody>
</table>

badges”. However, seven of those students said that they would have been motivated if badges affected the course results somehow or if they would just get bonus points without the badges. “They did not really affect my studying at all. It would be more motivating to give bonus points as a reward.”

**Emotions**

Three students wrote very negative feedback on the badges. One person noted that they should be taken away and that he “died internally” every time he saw them. Nine of the responses described a positive reaction towards the achievements: “Wonderful idea. Would it be possible to add some funny badges?”

**Social**

Even though the system did not have a way of sharing badges with other students, at least some students compared with their friends how many badges they got. Three students mentioned it in the feedback. Particularly one student noted: “I did not particularly care about achievement badges because I realized they did not have any real meaning, and that’s good. I did, however, have some natural instinct to collect more of them than my friend.”

**Affecting behavior**

In total, 14 students reported that the implemented badges affected their behavior. Time management was mentioned in five responses, such as: “[badges] brought immediately feeling of achieving that motivated me to do the online exercises on time and with care.” Seven responses pertained to carefulness and the number of submissions per exercise. One student told that before badges he “did not care how many times he had to try an exercise before passing”, but after badges were implemented he “always tried to get a perfect submission on the first try.”

Few students noted that the completion logic of the badges was unimaginative and boring, “Badges that appeared in the middle of the course had a little forced feeling to them ... I would make badges more epic by adding familiar concepts and humor from games”. Completion logic of the badges resembled closely the goals of the online exercises. Carefulness badges had the only completion logic that was not somehow tied to course grading and they were most mentioned when students discussed their changed behavior.

**4.2 Log data**

Since students only needed to complete 50% of the total amount of points, some of them do not work on exercises after achieving that limit. This leads to decreasing number of students submitting in each round. In Figure 3, we can see the average score for each round as well as the percentage of students submitting solutions. In the first round, almost all students enrolled in the course submit at least one exercise. In the last round, however, only 75% of students submit something. In addition, the average score in each round is also decreasing, which indicates that students do more exercises in the beginning of the semester than at the end of semester. Some of the fluctuation is caused by the level of difficulty, though. Typically, easier rounds collect more submissions.

The amount of badges achieved for the first four rounds (badges not visible) was calculated with same criteria as the implemented badges for the last four rounds (badges visible). The decreasing number of submissions towards the
end of the course, however, hinders our ability to assess the effect of the badges. More submissions in the beginning of the course means also more badges achieved (even though the students did not know about them). In addition, the exercises in the first four rounds were easier than the rest of the exercises. Thus, it was easier to achieve some of the badges as well. Thus, comparing the number of achieved badges in the first four rounds to the last four rounds showed no statistically significant improvements. Despite this, however, some (small portion of) students indeed improved their badge-performance by collecting more badges on the second half of the course.

The students’ interest in the badge system was also measured by the number of visits to the summary page where all achieved and available badges were shown. Figure 4 shows the amount of badges earned on the last four rounds and the number of times the summary view was visited. The plotted linear correlation indicates a connection between the amount of badges earned and the number of times information about them was viewed. 23% of the students that passed the online exercises did not visit the summary page at all, 21% visited it once and 56% visited it multiple times. There is also a separate group of students who viewed the summary page a lot more often than others, indicating interest in the badges.

5. DISCUSSION

The numerical feedback suggests that the system did not hinder or annoy the majority of students in a way that would have been detrimental to the learning outcome. Furthermore, we did not observe any negative effects on learning results either. Overall the implemented badges did not have a significant effect on the course results or student behavior because some students stop doing the exercises once they have enough points to pass or have achieved their desired grade. However, the written feedback suggests that students perceive badges in varied ways.

Based on the results from the freeform questions in the survey, there seemed to be two axis where most of the students can be placed: emotional and affect. Students’ emotions towards badges ranged from very negative to very positive with majority of reactions being indifferent. The other axis ranges from badges having no effect on behavior to having a positive effect. We did not observe negative effect from badges. Placing students on these two axis would be interesting to study in the future.

Gamification is not unanimously viewed as a positive contribution to learning. It poses challenges especially in the area of motivation. While adding points and badges to a core service might sound like an easy way to improve user experience and engagement, it is not as trivial in practise. Gamification has been criticized for having a negative effect on intrinsic motivation by granting extrinsic rewards which undermine the underlying intrinsic motivation. Meta-analysis on the literature has shown that in educational context giving out external rewards reduces internal motivation [3]. If adding badges discourages those who have a deep motivation to learn for the sake of learning and understanding, then clearly they should be used with caution. However, we did not find any negative effects on learning that could be tied to badges and this was also the case in Denny’s experiment [4].

Nicholson has presented a theoretical framework for meaningful gamification [14]. He concluded that meaning is defined differently by each individual so customization is required to achieve better outcomes. Furthermore, meaningful gamification should focus more on the aspects of playing and less on the aspect of scoring. We found this to be true in our case as well. Students experience varied, thus customization would have been beneficial, for example we could have provided an option to turn the badges off. In addition, based on the feedback, adding a social element, for instance a way of sharing badges with friends, would have also improved the system.

Another risk related to intrinsic and extrinsic motivation is that gamification might induce unwanted behavior where gamification becomes more important than the actual core function [1]. When this happens in an educational context, students start to care more about acquiring points and
badges than knowledge. An example of this type of behavior is a site called whatbadgenext\(^7\) which scans through your current progress for different badges in Khan Academy and suggests activities that will lead to a new badge with least amount of time and effort.

Moreover, it is possible that by encouraging students to achieve competitive time management badges might have had a negative effect on the carefulness [7]. This has the implication that achieving badges was becoming more important than the actual core service. This type of unwanted behavior can be especially detrimental in an educational context. In our study, none of the badges were competitive in the sense that only certain fraction of students would be able to get them, so this did not lead to negative results.

6. CONCLUSIONS

To find out how students react to gamification implemented with badges, we introduced badges to a large course with automatically assessed exercises in place. We collected and analyzed data from survey and from logs. Based on the observations from our experiment, we make some recommendations to other educators planning to bring badges to their teaching:

- We found that some students have strongly negative emotions towards badges. Therefore, one should make it technically easy to opt out from badges or make them opt in.

- Although we did not plan so, our students shared badges and discussed them. Therefore, one should make badges social in the first place. However, at the same time, it should be kept in mind that some students do not like to compete with peers.

- If there is already an automatic assessment system with points in place, badges telling how many points a student got are of little novelty. Instead, it would be a good idea to make badges something completely orthogonal to the grading policy. In addition, one should encourage latent aspects that are beneficial to learning but not graded. Moreover, badges should be fun (e.g., one could give a Santa badge if an exercise was submitted at Christmas).

- Our badges did not affect grading, but feedback from points and badges can confuse students. Therefore, one should give students clear instructions what else, if anything, is affected by badges.

- One should always make sure that adding badges does not hinder the usability of learning platform where the exercises and other core tasks related to learning are carried out.

Based on our log data, we were able to confirm Denny’s [4] observation that there is a positive correlation between the number of badges collected and how often a user views information about badges. In addition, although our survey hinted so, we were not able to find statistically significant results to support the claims that time management badges guide students to submit their work earlier and carefulness badges encourage students to avoid bricolage approach (trial-and-error) in problem solving.

\(^7\)http://whatbadgenext.appspot.com/

7. REFERENCES