Journal of Universal Computer Science, vol. 25, no. 12 (2019), 1589-1607
submitted: 30/4/19, accepted: 31/10/19, appeared: 1/12/19 © *J.UCS*

Badges for Peer Assessment of Teamwork in Organized Education

Zuzana Kubincová
(Comenius University in Bratislava, Slovakia
kubincova@fmph.uniba.sk)

Dana Šuníková
(Comenius University in Bratislava, Slovakia
dana.sunikova@gmail.com)

Martin Homola
(Comenius University in Bratislava, Slovakia
homola@fmph.uniba.sk)

Abstract: Team projects and group work have proven to be useful and rewarding educational activities that help students develop important abilities and skills. On the other hand, they have also proven to be difficult to grade in a balanced way which would be perceived as fair by the students. Numerical ratings such as those delivered by the Fink method provide ways to determine the shares of contributions of each individual team member. But such bare numerical evaluation needs also some secondary justification that helps team members to understand why their rating is high or low. Methods to collect such additional feedback often rely on surveys; however, extensive surveys may be perceived as tedious by the students, especially as they need to be filled out for each other team member participating in the team. This may have negative effect on the quality of feedback collected in form of surveys. We present a novel approach to peer assessment of a team-based project exploiting badges that represent individual contributions to the task and teamwork related traits. We show that this approach has positive influence on engagement in peer assessment compared to free-text open questions. We also study if the feedback obtained in this way is informative and to which extent it can serve as replacement for open free-text questions.

Key Words: badges, peer assessment, teamwork, negative feedback
Category: K.3.1, L.0.0, L.1.1, L.3.0, L.3.6, L.6.2

1 Introduction

Team-based projects employed as educational activities have several benefits for students: development of interpersonal skills, communication skills, pondering other points of view, ability to more effectively develop a complex project, insight into group dynamics, etc. [Wilson et al., 2018; Bailey et al., 2015; Ekblaw, 2017; Mello, 1993; Gatfield, 1999; Gibbs, 2009]. Working in teams represents also an interesting and effective way of learning [Conway et al., 1993] that improves students' engagement, performance and often also their marks [Gibbs, 2009]. However, for the teacher it is difficult to assess the individual's contribution to the group project. Grading all group members with the same mark or number of points is usually perceived as unfair. It can result in a variety of problems such as a sizable decrease in effort of particular students, freeloading, etc. [LaBeouf et al., 2016; Bacon et al., 1999; George, 1992].

There are various methods for reducing problems of this kind. Among others, peer assessment, or involving students in the assessment process in some way can be helpful. Such methods can be implemented in several diverse ways [Devlin et al., 2016; Cook et al., 2017; Murray, 2017; Gibbs, 2009].

In our course we introduced peer assessment of teamwork already four years ago [Kubincová et al., 2016; Homola et al., 2017]. The performance of particular team members was assessed using percentage evaluation of each other's contribution, in compliance with so called Fink method [Michaelsen and Fink, 2004]. We have employed also verbal justification in form of open textual questions to supplement the percentage rating. This secondary assessment is essential, because it provides formative feedback between the team members and a source of validation of the primary numerical rating. Students expressed their satisfaction with peer assessment of teamwork, however, they were not enthusiastic about writing the verbal evaluation.

To deal with this problem we experimented with employing gamification in the peer evaluation process. The study was conducted during two consecutive runs of a Master's level web design course that prominently features a team-based term-long project assignment. Open textual questions previously used in teamwork evaluation were replaced by the possibility to award badges between team members. We hypothesized that students would be much more willing to provide the feedback in form of badges. In addition, we wanted to verify if the badges were somehow representative with respect to the students' individual contribution within a team, and if this form of peer assessment was useful to the instructors.

While there were a few students who were not too enthusiastic about using badges, these were exceptions. In general most of the students were much more willing to provide feedback in this form, when compared to past years that featured open questions. While this kind of feedback is less accurate, it was still useful to the instructors. In the second of the two course runs we also experimented with adding a negative badge into the set. While the badge was only seldom used, we found out that the students considered it a useful addition.

2 Related Work

Along with an explosion of group-based activities in education [Gibbs, 2009; Ktoridou and Doukanai, 2016] a problem of assessment of individual contributions to the group work arose. As assigning the same grade to all group members often leads to freeloading; students perceive such an assessment as unfair, and are unable to take advantage of group work in their learning. Various approaches how to deal with this problem have been proposed [Synnott, 2016; Lejk and Wyvill, 2001; Gibbs, 2009; Hindle, 1993].

According to Gibbs [2009], there are a number of mechanisms that allow different grades to be given to individual students within a group. They are generally perceived as fair and can therefore lead to better use of the educational benefits of group work.

Based on analysing many other studies Gibbs suggested several options to decrease the problems related to freeloading caused by unsuitable evaluation of group work, such as (a) limiting the emphasis on group marks by allocating a substantial part of grading to individual assignments and tests; (b) assessing the outcomes of group work using individual examinations or assignments [Hindle, 1993]; (c) dividing the whole group task among group members and grading particular sub-tasks separately [Lejk and Wyvill, 2001]; (d) grading each group member individually based on the teacher's specific knowledge of the individual [Gibbs, 1995]; (e) moderating grades of the other group members by students themselves on the basis of their inside knowledge about the contribution of individuals to the group work.

Active involvement of students in the assessment process is being explored for more than two decades. A number of techniques have been designed and tested to evaluate the teammate's work [Planas-Lladó et al., 2018; Cook et al., 2017; Devlin et al., 2016; Murray, 2017; Gibbs, 2009].

The ability to evaluate the contribution of the others to a teamwork is considered one of the key skills required in IT area [Raban and Litchfield, 2012]. However, the responsibility to allocate individual scores to all teammates according to their individual contributions was shown to be too difficult for students who were not trained for such a task [Steensels et al., 2018] and in some cases resulted in egalitarian evaluation [Lejk and Wyvill, 2001]. Fortunately, peer assessment is becoming a widespread educational activity that helps students to develop the evaluation, feedback and review skills already during their studies [Raban and Litchfield, 2012; Wilson et al., 2018]. However, it seems that peer grades are more reliable when the grader knows the student she is evaluating, but the graded student does not know who is evaluating her work [Gibbs, 2009].

The peer assessment procedure is usually based on one of the two approaches [Lejk and Wyvill, 2001]: the holistic one or the category-based one. Using the holistic approach the reviewer rates her teammates by a single grade working only on the overall impression of their contribution to the group project. With the category-based approach several predefined criteria describing particular teamwork aspects are rated resulting in a single final score for each group member. Several studies argue for the latter one [Gatfield, 1999; Young and Henquinet, 2000], nevertheless, as shown by Lejk and Wyvill [2001] the holistic method is more in line with the aim of group assessment than the category specific approach. Moreover, it is even more positively accepted by students [Lejk and Wyvill, 2002]. Ohland et al. [2005] investigated also holistic approach with behavioral anchors outlining the performance of a student in a group, e.g. 'very good – consistently did what she was supposed to do, acceptable, well prepared and cooperative'. According to the findings in this paper the inter-rater reliability significantly improved when holistic peer assessment with behavioural anchors was used.

In general, the assessment can be summative or formative. The former one is used to evaluate learning outcomes at the end of a particular unit or activity and usually includes grading that indicates the performance level of the student. The role of the formative assessment is to summarize the students' development at a particular time with the aim to improve the teaching-learning process for students. Peer assessment can also be carried out either in summative or in formative way [Sridharan et al., 2019; Deeley, 2014; Orsmond et al., 2004]. The latter one allows for formative feedback that can help students to focus on their learning and enhance the learning experience.

Brooks and Ammons [2003] observed that peer assessment may have a role in forming the students' attitude towards their work. They pointed out the significance of an early and multiple points feedback during the group project development to the improvement of team members' contributions and reduction of freeloading. This finding was also affirmed by Kench et al. [2009] who used peer assessment of individual contributions to a group project to encourage all students' active participation in the group work. They used the holistic approach of peer assessment, however, students also answered twelve yes/no questions that were intended as a guide to the rating itself.

Several authors consider the peer evaluation a key strategy in team-based learning [Devlin et al., 2016; Murray, 2017; Cook et al., 2017]. They see its advantage in supporting contributions to the group work and avoiding the freeloading problem in the groups [Levine, 2008; Chin and Overton, 2005; Chang and Kang, 2016]. In the case of receiving peer feedback during a course, this evaluation can serve as formative feedback and can help the students to significantly improve their behavior in the team and become more effective co-workers [Levine, 2008].

According to Chin and Overton [2005] certain problems can arise if students perceive the received peer feedback being too subjective. However, such a problems can be avoided by preparing guidelines with the clearly stated information for the students about what is required of them, how they are being assessed, and what makes effective group work.

Multiple methods that can be used in peer assessment applied in team-based learning were studied [Michaelsen et al., 2004; Levine, 2008; Michaelsen and Fink, 2004; Wu et al., 2012; Koles et al., 2010]. They are mostly based on assigning scores, points, or percentage to team-members and justifying the evaluation. The differences lie in the method of dividing the points and calculating the overall score for a particular team-member and in the type of written feedback. One of the frequently used methods is the Fink method Michaelsen and Fink [2004] we also used in the research described in this paper.

Aiming to encourage and moderate the students' collaboration and to engage them in the assessment process, various gamification-based approaches were suggested. Some of them use badges – digital pictures with meta-data. Although digital badges originated in digital games, technologies for awarding digital badges outside of the game context were developed and badges started to be used as a motivation factor in a broader context, such as in internet banking, or in various mobile and web applications for different activities [Šuníková et al., 2015]. Digital badges are used in education for a few reasons

– they might be used as a motivation, as a recognition of status, clear, evidence-based credentials, a visual evidence of achievement easily communicated and understood by observers and as a guide or a signpost communicating relevant targets to learner [Gibson et al., 2015; Rughiniş and Matei, 2013].

Moccozet et al. [2013] introduced a framework for group work evaluation enhanced by gamification components – user points and scores. As the authors conclude, this learning platform not only encourages students to contribute and cooperate, but also covers the "free rider" problem as it provides the teacher with information about the student's individual contribution.

Hamari [2013] reported results of a field experiment conducted in a real existing service gamified through badges. According to the findings of this study participants who actively monitored their badges and badges of the others were showing significantly higher activity.

Tenório et al. [2016] proposed a gamified peer assessment model to deal with the lack of student engagement and motivation. Several gamification elements, such as points, rankings, badges, medals, and missions, were used. According to the authors' findings, gamification motivated students to participate in activities within this platform.

Badges as gamification elements were also used for summative peer assessment in a project-based learning scenario [Šuníková et al., 2017]. During the presentations of team projects other students were asked to reward their colleagues with badges according to the quality of their project.

Nevertheless, all of these studies used the gamification elements in peer evaluation of the others' work. In our research, badges were used for peer assessment of teammates' contribution and attitude.

3 Experiment

3.1 Course and Assignment

The experiments described in this paper were conducted during the two latest runs of a Master's level course concerned with web design methodology and user experience (in 2017 and 2018). The course is part of an Applied Informatics curriculum. As one of the main educational activities, it features a whole-semester practical project during which students design and develop web application of their own choice.

The assignment is team-based; students self-form 3–4 member teams. It is split into four consecutive rounds: (1) specification, (2) prototype development and testing, (3) application development, (4) content development.

During each round the students work in their team, then they conduct peer review of other teams' submissions, and consecutively they are able to take the feedback into account and finalize the submission for grading.

It is one of our goals to make the assignment similar to real web application development process. Therefore we allow the students to split the responsibilities in their teams arbitrarily, based on their own agreement.

To be able to asses the individual contributions of the team members, we have introduced teamwork reviews [Kubincová et al., 2016; Homola et al., 2017]. We rely on the Fink method [Michaelsen and Fink, 2004] wherein each team member splits 100 points among the remaining $n-1$ members of an n-member team. If the points are split equally, this means that in the evaluator's point of view all other team members contributed by the same amount of work. If the evaluator deems that someone contributed more than the others, she may adjust the points accordingly. No one evaluates herself. In the end, the points received by each member from all her colleagues are summed up. In the even case, everyone receives 100 points, hence the points may be viewed as per cent scales.

The evaluation of the project, which takes into account both the instructors' assessment of the submissions but as well the team review scores, is the main part of the course grading. Therefore it is important to assure that the team reviews are fair, and to have some additional justification to supplement the mere numeric evaluation respective to the Fink method as described above.

In the previous years we relied on textual input from the students. Besides for awarding the points, all students were asked additional open questions regarding each other team member. We have used the following three questions:

1. *Contribution summary:* Evaluate your colleague's contribution in this round of the project. Summarize what this person has done, and if you are satisfied by her contribution.

2. *Most valuable thing:* What is the single most valuable thing this person has done for your team?

3. *What can be improved:* What is the single most important thing this person could do to more effectively help your team?

While the outputs that we received from these open questions were useful (if provided), the task of filling out three questions per team member in each project round was viewed as laborious by many students and they were reluctant to fill in the answers properly. We resorted to awarding students a small number of evaluation points just for filling-in all the review forms, which worked to some extent, but was not ideal.

3.2 Badges

Two years ago, we have looked for ways how to improve the team review process by reducing the effort needed by the students and perhaps making this activity less boring and more fun; while still assuring some form of independent supplementary feedback besides for the split of points. As we explain in this section, we have resorted to allow students to evaluate their team members' contribution by awarding badges.

Figure 1: Set of badges; badge levels represented by different color

The teamwork review forms were modified as follows: (1) The split of points between all team members (based on the Fink method) – this part remained unchanged; (2) We added a new part in which students were able to award badges to each team member; (3) We kept one open textual field for any additional feedback to each team member, but it was optional.

The remaining parts of the project assignment were not changed, although the project was split into 4 rounds instead of 3 compared to the previous years.

Since the badges were to be used to supplement the evaluation of team members' individual contribution during each round, we have put together a set of 15 badges covering teamwork related skills and other skills needed to fulfil the task. The complete set of badges is showed in Fig. 1 where the badges are shown as in one of the student's reports; different colors distinguishing different levels of each badge. In the user interface each badge is supplemented with an explanation detailing its actual meaning, that is shown as tooltip. These explanations are listed in Table 1.

All the badges in the suggested set were positive – teammates were awarded a badge if they demonstrated the respective skill. However, at the end of the first course run with badges, the students demanded a negative badge. Therefore in the last course run the *Lazy* badge was introduced (Fig. 2). It replaced one of the former badges (*Neat coder*) that was not used frequently in the previous team review run.

In both course runs, the set of badges was the same during all four rounds. Therefore it was possible to be awarded each badge multiple times during each round but also again during next rounds. For this reason the badges were aggregated into four levels

Figure 2: The *Lazy* badge used in the last course run.

htp

Table 1: Meaning of badges

Badge	Meaning
Communicative	Has strong communication skills
Constructive	Gives constructive feedback
Creative	Proposes creative solutions
Engaged	Always eager to put effort into the project
Guru	Understands the project thoroughly
Hacker	Solves any programming problem like a piece of cake
Hard worker	Puts a lot of hard work into the project
Helpful	Constantly helps colleagues
Lazy	Too lazy, does not help much with the project
Leader	Shows excellent leadership skills
Motivator	Motivates others to deliver the best work
Neat coder	Writes clean code that is easy to read
Patient	Is patient with others, especially in stressful situations
Punctual	Does things on time
Responsible	Can be relied upon
Team player	Puts effort into working with colleagues

of each badge: regular (green), bronze, silver, and gold. If a student earned a particular badge in one round, the badge was shown as regular. Badges earned in two rounds were shown as bronze, silver for three rounds, and finally golden for earning the badge in four rounds. The number of distinct students who awarded the badge did not matter.

3.3 Data and Methodology

During the last two course runs we have conducted an experiment, during which we have introduced badges in teamwork reviews, as described above. There were 17 students taking the course in 2017 and 18 students in 2018. They formed ten 3–4 member teams (five teams in each course run). The aim of the experiment was to answer to the following research questions:

1. Did students use and did they like to use badges? Did they prefer badges to written feedback?
2. Were badges representative w.r.t. the students' contribution to the teamwork?
3. Were badges useful to the instructors?

The research questions were addressed by the following methods: (1) We collected and evaluated the data about the usage of badges. (2) At the end of each course run we surveyed students about their opinions, the survey was anonymous and voluntary. It focused on the student's perception of the task and its information value, and utility. All students who passed the exam participated in the survey (14 students in 2017 and 17 students in 2018). (3) We have also conducted interviews with the instructors of the course, especially in connection to the third research question.

The collected data were analysed using quantitative and qualitative methods. It should be noted, that it is not our goal in this experiment to measure the objective truth when it comes to individual team member's contribution to the project, especially w.r.t. the research question 2. That is, we are not concerned with comparison of the actual objective measure of team member's work contribution and their received badges. We are merely concerned with comparison of the contribution as perceived by the teammates (and to a lesser extent, as perceived by the instructors) and the received badges. This is justified by our goal to design a team evaluation strategy that is *perceived as fair* by the students, and thus to increase their satisfaction and overall good learning environment. The question whether such group work evaluation method does accurately gauge the individual team member's performance from the instructor's point of view is a different one, and outside of the scope of this research. The findings are presented in the next section.

4 Findings

4.1 Did Students Prefer Badges to Written Feedback?

Since we implemented badges as a reaction to students' complaints about written team-reviews, the foremost interest of our research was to find out if the students enjoyed this kind of evaluation and to what extent they used badges. Thus we analysed the collected data and also asked several related questions in the survey.

We found that each student awarded a badge 12–101 times. The number of times a student earned a badge was 4–105. See Fig. 3 for details. This translated into 449 aggregated badges: 83 regular, 94 bronze, 104 silver, and 168 gold. *Helpful*, *Team Player*, and *Responsible* were the most popular, while *Lazy* (used only in 2018), *Neat Coder* (used only in 2017), *Motivator*, *Leader*, *Guru* were used the least.

The first multiple-choice question in the survey targeted students' overall opinion on the whole team review system with the list of answers as follows: It was fair, I was

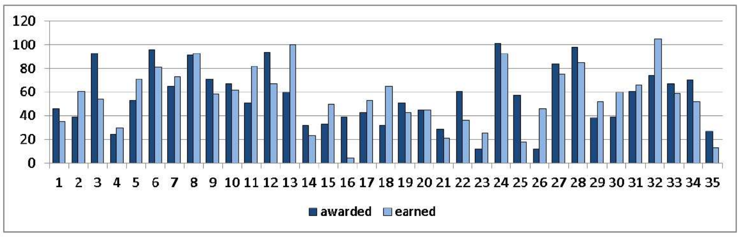


Figure 3: Badges awarded and earned per person (no. of times); 1–35 stand for individual students

satisfied; It helped us split project points fairly; It helped us improve teamwork; I did not consider it fair; It did not help improve anything; Other. Students were allowed to tick all the choices that were relevant. After sorting the "other" answers, we found that we obtained 80.9% positive, 10.6% neutral and 8.5% negative answers (see Fig. 4).

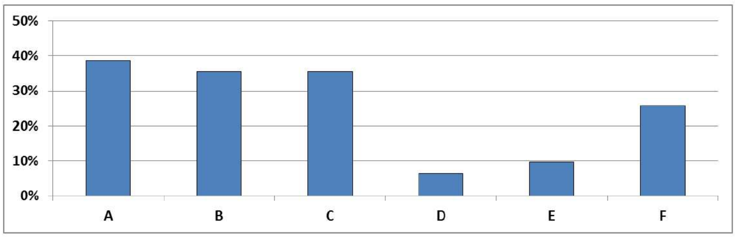


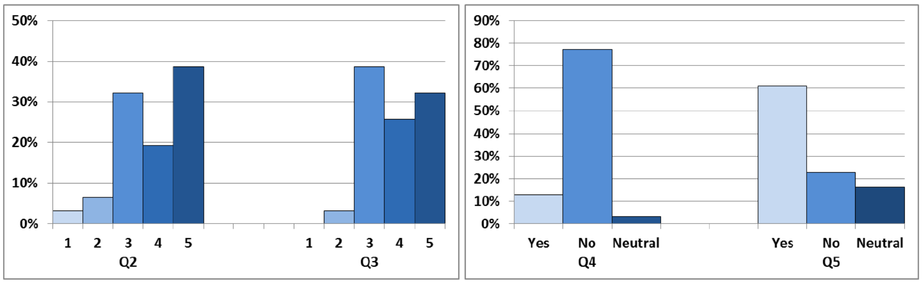
Figure 4: Q1: What is your overall opinion of our team-review system? A – It was fair, I was satisfied; B – It helped us split project points fairly; C – It helped us improve teamwork; D – I did not consider it fair; E – It did not help improve anything; F – Other

In the following two questions we investigated if students liked to earn badges (Q2) and if they liked to assess their colleagues with badges (Q3). The answers were on the scale from 1 (the most negative option) to 5 (the most positive option). As shown in Fig. 5(a) both questions were answered mostly positively or neutrally.

Students were also asked to justify their answers to these two questions. Three students who answered Q2 negatively saw badges as not entirely objective, not much satisfying, and useless because of not having any impact on the final grade. For the only one student, who answered Q3 negatively, this kind of assessment was not satisfying. Positive answers emphasized good feelings associated with earning an achievement, and getting feedback from peers. Students also appreciated that assessing with badges is fast and accurate, some of them considered badges to be cool, funny and interesting. Several students stated that earning badges had motivational effect on them.

We asked students if they would rather assess their colleagues by written comments (Q4 in Fig. 5(b)). Only 29 students answered, 83% of them preferred badges to verbal comments since they regard them as fast, less tiring, easy to use and accurate. They also worried that students would write minimal to nothing. Only 14% of students would prefer verbal feedback as they perceived it more objective, clearer and more reflective than badges.

The question if students would like to use badges in other courses (Q5 in Fig. 5(b)) was again answered mostly positively (61%). However, neutral and negative answers also occurred (16% and 23% respectively).



(a) (b)

Figure 5: (a) Q2: Did you like to earn badges? (Scale 1–5: I strongly disagree – I totally agree); Q3: Did you like to assess your colleagues with badges? (Scale 1–5: I did not like it at all – I really enjoyed it); (b) Q4: Would you rather assess your colleagues by written comments than by badges?; Q5: Would you like to use badges in other courses?

Based on the survey outcomes we can conclude that, in general, students used badges and liked to use badges. This finding was also confirmed by the teachers in the interview. Although rewarding teammates with badges was an optional part of team review, teachers were pleased with the amount of badges students awarded to each other.

From the data, it is evident that there were only few instances of a student assessing one of her peers in the particular round with no badge. In certain cases the students affirmed that this way they expressed their discontent with the teammate's contribution. In another significant case, when a team decided to always assess all team members in an egalitarian way, the team leader used the badges and verbal comments to distinguish hard-working teammates from those who have not contributed much to the teamwork. We observe that the badges provided some additional useful tool for this student to express feedback without affecting her colleague's score. Thus badges provided more flexibility in this case.

4.2 Can Badges Reflect Students' Contribution?

In the research question 2 we were concerned with whether the badges awarded to a student reflect her contribution within the team, at least to a certain extent. To evaluate this question, we have included four additional questions (Q6–Q9) in the survey in which we asked whether the students can match the awarded badges to their perception of the actual contribution of each team member. The students answered on the scale from 1 (I strongly disagree) to 5 (I totally agree). The results are shown in the Fig. 6.

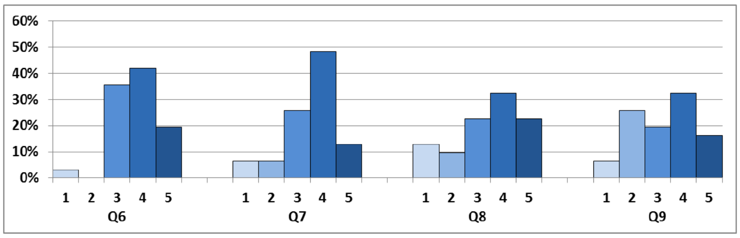


Figure 6: Q6: Do you think that the badges you got during team review represent your teamwork well? Q7: Do you think that the badges your teammates received represent the work they did on the project? Q8: Did the badges help you realize your strengths within teamwork? Q9: Did the badges help you realize what you need to improve within teamwork? (Scale 1–5: I strongly disagree – I totally agree)

Based on the answers to Q6 and Q7, we may conclude that the vast majority of students were positive (61% in both questions) or neutral regarding the representation of their own or teammates' work by the earned badges. Looking at Q8 and Q9 we observe that the responses are distributed to all types of answers, with a slight shift to positive responses. This shift is more pronounced in Q8, which means that according to the students the badges reflected better their strengths than their weaknesses. To some extent this is probably caused by the fact that there was no badge in our badge set that could be used to describe negative behaviour in 2017 and only one negative badge in 2018. Thus only the lack of awarded badges could be seen as a form of such feedback.

Looking at the same problem from the instructors point of view we obtain the research question 3. The instructors, however, had a more limited perception of students' split of responsibilities and their individual contributions. This perception is derived from their overseeing of the practical sessions during which part of the students' work was carried out. Therefore the research question is also more loosely formulated: we wanted to find out, if the badges are useful to the instructors, especially in order to understand who of the team members have contributed more and who have contributed less to the project.

The instructors truly confirmed that in some cases the badges helped them to identify the more active team members. This was helpful e.g. in situations when students expressed their discontent with the number of points the particular team members got in the final project evaluation.

We learned that the number of received badges and especially the number of golden and silver badges could also be seen as an indicative. The instructors noticed certain correlation between students' engagement in the project, the split of points earned from their teammates, and the number of badges they have earned. This correlation was also observable in time: a person who worked more in the beginning earned more badges then, while later on she earned less badges because she worked less.

In a few notable cases, they were able to match some of the students' badges with their perceived traits/contribution within the project. This was especially true with the badges such as *Leader*, *Creative*, or *Communicative*. But most of the badges were intended to be earned for a behaviour which was not always possible to observe by the instructors in the given timespan of practical sessions.

The instructors also concluded that, in general, if a larger set of leader traits (such as *Responsible*, *Team Player*, *Helpful*, *Creative*, *Communicative*) was taken into account they were able to better identify the team leader than by tracking only the *Leader* badge.

All in all the instructors concluded that the badges were useful to some extent. Textual answers would be more indicative, if only students have not been reluctant to provide them.

4.3 Negative Badges

After the 2017 course run, our survey included also an open question regarding other badges that might be useful. We have received mostly suggestions for some kind of negative badge, that could be awarded to teammates who did not contribute sufficiently, i.e., to be able to express also some form of negative feedback by badges. Only a few suggestions were towards positive badges. See Fig. 7 (a). To large extent this can be attributed to the fact that we have already provided the students with a large number of positive badges, and conversely with no negative ones. But it also shows that a large percentage of students would welcome some form of a negative badge.

Independently, this issue was brought up by the instructors during the interview. There was no clear consensus apart from that there were situations in which such a badge could be useful. However, two out of the three instructors expressed concerns that it could also be abused by the students. Still, they would be curious to try it at least as an experiment. When asked how should such a badge could be named, "Participation Trophy" came out as a possible candidate.

As already noted, in the consecutive course run we have decided to include one negative badge, the *Lazy* badge (see Fig. 2). We were interested to see if students embraced it, if they used it, and if it was useful to the instructors.

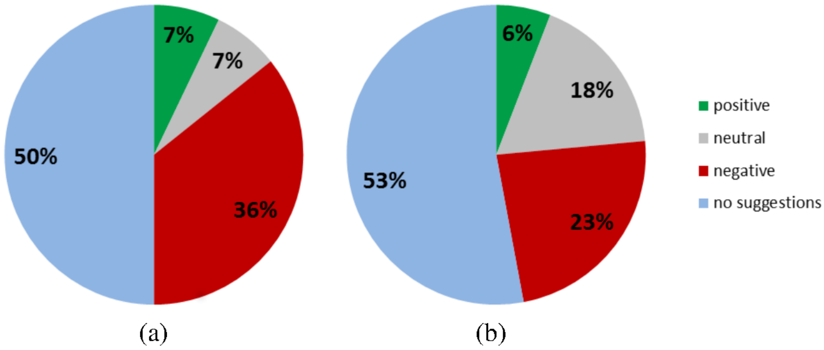


Figure 7: Survey findings (Q10) regarding missing badges: (a) 2017, (b) 2018

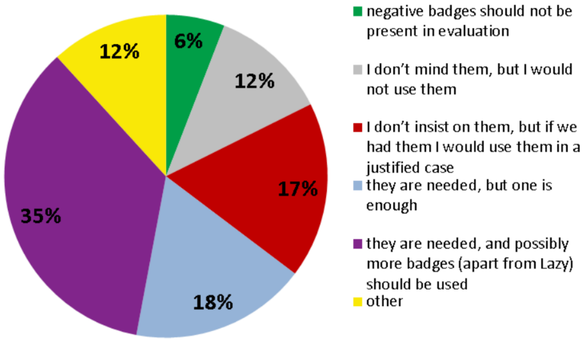


Figure 8: Q11: What is your opinion about negative badges, such as our Lazy badge?

As we learned, the badge was only used twice, and in fact each time it was awarded by the same student, to the same team member. This usage was justified, as we learned from the instructors.

When we asked the same question in the next survey, i.e., which other badges might be useful the percentage of negative suggestions decreased from 36 % in 2017 to 23 % in 2018. This may suggest that perhaps even if the students were hesitant to use the negative badge and only used it if really deserved, it may still be useful to include it for them to feel that they have the power to award it. Also the concerns that this kind of badge may be abused by the students was not justified.

The students of the 2018 course run already had some experience with at least one negative badge, so we explicitly asked them in the survey about their opinion on such badges. As plotted in Fig. 8 only 6 % of the students opposed negative badges, 29 % chose one of the two more or less neutral answers, and 52 % were in favour of negative badges out of which the larger part would recommend to add more negative badges.

In the interview the instructors also found the badge as useful, as even if it was only seldom awarded, it was in a justified case when the recipient student indeed deserved it. They also concluded, that perhaps the choice to name the badge *Lazy* was perhaps too strong – if there were multiple weaker negative badges, perhaps the students would not be as hesitant to use them as a form of constructive feedback. Other suggestions included to try to include negative badges that represent the opposite treats of the available positive badges, and perhaps also to distinguish negative badges by different colour. In the end the instructors concluded that they would definitely like to try to include more negative badges in the next course run. In a sense this is also justified by Q10 and Q11 results of the students' survey, as we already noted.

5 Conclusions

We have focused on the assessment of individual team member's contributions within a team-based project assignment. In such assignments, where some of the team members may contribute more and others less, it is important to select an appropriate evaluation model, so that the team members perceive the assessment as fair. Otherwise it can lead to adverse effects on students, decreased motivation, and overall quality of learning [Bacon et al., 1999; George, 1992].

We report on a new gamified approach to peer assessment of individual contributions to a team-based project. We rely on the Fink method [Michaelsen and Fink, 2004] by which the team members assign numerical rating to their peers. However for justification of the rating we have replaced free-text open questions, which the students were reluctant to answer in the past, with a set of badges representing traits relevant to the project and to teamwork.

Our first research question was concerned whether students actually used badges, if they liked it, and if it can be concluded that they prefer this form of feedback to open questions. We are able to answer this question positively to a large extent. Students were much more active in awarding badges when compared to the past experience with open questions. In the survey a large majority answered that they liked to earn and to award badges and they would not like to answer open questions instead.

The second research question was concerned whether the badges reflected the individual team members' contributions, especially as perceived by the students. We are able to answer also this question positively. The majority found the badges reflecting not only their perceived contribution but also how they perceived the contribution of their teammates. To some extent they were also able to recognize the utility of the badges as a form of formative feedback, an aspect that we would like to reinforce in the future.

Finally the last research question which we posed beforehand was concerned whether the badges were also useful to the instructors. The interview concluded that justification questions with open answers would be more indicative when it comes to the validation of the primary numeric ratings; but since students are reluctant to answer these questions, the badges are also helpful – at least partly. The instructors concluded that higher number of received badges, especially those of higher level, partly correlated with those students who the instructors perceived as more active in their respective teams.

There were also some additional takeaways aside from the formally formulated research questions. As the preliminary results after the 2017 course run indicated [Šuníková et al., 2018] an interesting possibility to introduce also a negative badge into the overall set, we tried one (*Lazy*) in the consecutive course run. The experience showed that the concerns that such a negative badge could also be abused by the students proved not to be justified. Students awarded it only seldom and in a justified case. Based on the survey we have also learned that many students would welcome further negative badges.

Also, most students – 24 out of 30 who answered the survey – considered the experience of peer-reviewing rather useful with regard to their future job.

Even if our study so far is limited in scale (35 students in total; 5 teams in each of the two years), these results encourage us to continue in this direction in the future. We consider the change from open textual answers to badges an improvement because in the former situation students were less willing to provide the feedback and hence all in all there was less information available to them (and also to the instructors) regarding the numerical evaluation they received. In the future we would like to especially focus on further increasing the amount of feedback delivered via the badges that can be useful to the students but to the instructors as well.

Acknowledgments

The authors would like to thank to their colleague Ján Kl'uka. This work was supported from the Slovak national VEGA project no. 1/0797/18.

References

[Bacon et al., 1999] Bacon, D. R., Stewart, K. A., and Silver, W. S. (1999). Lessons from the best and worst student team experiences: How a teacher can make the difference. *Journal of Management Education*, 23(5):467–488.

[Bailey et al., 2015] Bailey, S., Barber, L. K., and Ferguson, A. J. (2015). Promoting perceived benefits of group projects: The role of instructor contributions and intragroup processes. *Teaching of Psychology*, 42(2):179–183.

[Brooks and Ammons, 2003] Brooks, C. M. and Ammons, J. L. (2003). Free riding in group projects and the effects of timing, frequency. *Journal of Education for Business*, 78:5.

[Chang and Kang, 2016] Chang, B. and Kang, H. (2016). Challenges facing group work online. *Distance Education*, 37(1):73–88.

[Chin and Overton, 2005] Chin, P. and Overton, T. (2005). Assessing group work: advice and examples. *The Higher Education Academy, Physical Sciences Centre, Primer*, 6.

[Conway et al., 1993] Conway, R., Kember, D., Sivan, A., and Wu, M. (1993). Peer assessment of an individual's contribution to a group project. *Assessment & Evaluation in Higher Education*, 18(1):45–56.

[Cook et al., 2017] Cook, A. R., Hartman, M., Luo, N., Sng, J., Fong, N. P., Lim, W. Y., Mark, I., Chen, C., Wong, M. L., Rajaraman, N., et al. (2017). Using peer review to distribute group work marks equitably between medical students. *BMC Medical Education*, 17(1).

[Deeley, 2014] Deeley, S. J. (2014). Summative co-assessment: A deep learning approach to enhancing employability skills and attributes. *Active Learning in Higher Education*, 15(1):39–51.

[Devlin et al., 2016] Devlin, M., Marshall, L. F., and Phillips, C. (2016). Fair assessment of contribution and process in student team projects. In Whatley, J. and Nerantzi, C., editors, *Teaching with Team Projects in Higher Education*. Informing Science Press, Santa Rosa, CA, US.

[Ekblaw, 2017] Ekblaw, R. (2017). Effective use of group projects in online learning. In Kantola, J. I., Barath, T., Nazir, S., and Andre, T., editors, *Advances in Human Factors, Business Management, Training and Education*, pages 475–483. Springer.

[Gatfield, 1999] Gatfield, T. (1999). Examining student satisfaction with group projects and peer assessment. *Assessment & Evaluation in Higher Education*, 24(4):365–377.

[George, 1992] George, J. M. (1992). Extrinsic and intrinsic origins of perceived social loafing in organizations. *Academy of Management Journal*, 35(1):191–202.

[Gibbs, 1995] Gibbs, G. (1995). *Learning in teams: A tutor guide*. Oxford Centre for Staff and Learning Development.

[Gibbs, 2009] Gibbs, G. (2009). The assessment of group work: lessons from the literature. *Assessment Standards Knowledge Exchange*, pages 1–17.

[Gibson et al., 2015] Gibson, D., Ostashewski, N., Flintoff, K., Grant, S., and Knight, E. (2015). Digital badges in education. *Education and Information Technologies*, 20(2):403–410.

[Hamari, 2013] Hamari, J. (2013). Transforming homo economicus into homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service. *Electronic commerce research and applications*, 12(4):236–245.

[Hindle, 1993] Hindle, B. P. (1993). The 'project': putting student-controlled, small-group work and transferable skills at the core of a geography course. *Journal of Geography in Higher Education*, 17(1):11–20.

[Homola et al., 2017] Homola, M., Kubincová, Z., Kl'uka, J., Švolík, M., Darjanin, M., and Zapalač, P. (2017). Team workflow and peer review in a virtual learning environment. In *Information Technology Based Higher Education and Training, 2017 16th International Conference on*, pages 1–6. IEEE.

[Kench et al., 2009] Kench, P. L., Field, N., Agudera, M., and Gill, M. (2009). Peer assessment of individual contributions to a group project: Student perceptions. *Radiography*, 15(2):158–165.

[Koles et al., 2010] Koles, P. G., Stolfi, A., Borges, N. J., Nelson, S., and Parmelee, D. X. (2010). The impact of team-based learning on medical students' academic performance. *Academic Medicine*, 85(11):1739–1745.

[Ktoridou and Doukanai, 2016] Ktoridou, D. and Doukanai, E. (2016). *Teaching with Team Projects in Higher Education*. Informing Science Press, Santa Rosa, CA, US.

[Kubincová et al., 2016] Kubincová, Z., Homola, M., and Dropčová, V. (2016). Peer reviews in a web design course: Now students like them too. In *International Symposium on Emerging Technologies for Education*, pages 624–634. Springer.

[LaBeouf et al., 2016] LaBeouf, J. P., Griffith, J. C., and Roberts, D. L. (2016). Faculty and student issues with group work: What is problematic with college group assignments and why? *Journal of Education and Human Development*, 5(1):13–23.

[Lejk and Wyvill, 2001] Lejk, M. and Wyvill, M. (2001). Peer assessment of contributions to a group project: a comparison of holistic and category-based approaches. *Assessment & Evaluation in Higher Education*, 26(1):61–72.

[Lejk and Wyvill, 2002] Lejk, M. and Wyvill, M. (2002). Peer assessment of contributions to a group project: student attitudes to holistic and category-based approaches. *Assessment & Evaluation in Higher Education*, 27(6):569–577.

[Levine, 2008] Levine, R. E. (2008). Peer evaluation in team-based learning. In *Team-Based Learning for Health Professions Education: A Guide to Using Small Groups to Improve Learning*, pages 103–116. Stylus.

[Mello, 1993] Mello, J. A. (1993). Improving individual member accountability in small work group settings. *Journal of management Education*, 17(2):253–259.

[Michaelsen and Fink, 2004] Michaelsen, L. K. and Fink, L. D. (2004). Calculating peer evaluation scores. In *Team-Based Learning: A Transformative Use of Small Groups in College Teaching*, pages 241–248. Stylus.

[Michaelsen et al., 2004] Michaelsen, L. K., Knight, A. B., and Fink, L. D. (2004). *Team-based learning: A transformative use of small groups in college teaching*. Stylus.

[Moccozet et al., 2013] Moccozet, L., Tardy, C., Opprecht, W., and Léonard, M. (2013). Gamification-based assessment of group work. In *Interactive Collaborative Learning (ICL), 2013 International Conference on*, pages 171–179. IEEE.

[Murray, 2017] Murray, J. W. (2017). I hate/don't hate/still hate group projects! a tripartite ethical framework for enhancing student collaboration. *Cogent Education*, 4(1).

[Ohland et al., 2005] Ohland, M. W., Layton, R. A., Loughry, M. L., and Yuhasz, A. G. (2005). Effects of behavioral anchors on peer evaluation reliability. *Journal of Engineering Education*, 94(3):319–326.

[Orsmond et al., 2004] Orsmond, P., Merry, S., and Callaghan, A. (2004). Implementation of a formative assessment model incorporating peer and self-assessment. *Innovations in Education and Teaching International*, 41(3):273–290.

[Planas-Lladó et al., 2018] Planas-Lladó, A., Feliu, L., Castro, F., Fraguell, R. M., Arbat, G., Pujol, J., Suñol, J. J., and Daunis-i Estadella, P. (2018). Using peer assessment to evaluate teamwork from a multidisciplinary perspective. *Assessment & Evaluation in Higher Education*, 43(1):14–30.

[Raban and Litchfield, 2012] Raban, R. and Litchfield, A. (2012). Peer assessment in large group projects: forming professional attitudes in it students. In *Australian Conference on Science and Mathematics Education*.

[Rughiniş and Matei, 2013] Rughiniş, R. and Matei, S. (2013). Digital badges: Signposts and claims of achievement. In *International Conference on Human-Computer Interaction*, pages 84–88. Springer.

[Sridharan et al., 2019] Sridharan, B., Tai, J., and Boud, D. (2019). Does the use of summative peer assessment in collaborative group work inhibit good judgement? *Higher Education*, 77(5):853–870.

[Steensels et al., 2018] Steensels, C., Leemens, L., Buelens, H., Laga, E., Lecoutere, A., Laekeman, G., and Simoens, S. (2018). Peer assessment: A valuable tool to differentiate between student contributions to group work? *Pharmacy Education*, 6(2).

[Šuníková et al., 2018] Šuníková, D., Kubincová, Z., and Homola, M. (2018). A badge for reducing open answers in peer assessment. In *International Conference on Web-Based Learning*, pages 14–24. Springer.

[Šuníková et al., 2017] Šuníková, D., Kubincová, Z., Simionescu, S., and Popescu, E. (2017). Peer assessment augmented with digital badges: A first experience report. In *Information Technology Based Higher Education and Training, 2017 16th International Conference on*, pages 1–5. IEEE.

[Synnott, 2016] Synnott, K. (2016). Guides to reducing social loafing in group projects: Faculty development. *Journal of Higher Education Management*, 31(1):211–221.

[Tenório et al., 2016] Tenório, T., Bittencourt, I. I., Isotani, S., Pedro, A., and Ospina, P. (2016). A gamified peer assessment model for on-line learning environments in a competitive context. *Computers in Human Behavior*, 64:247–263.

[Šuníková et al., 2015] Šuníková, D., Kubincová, Z., and Navrátil, M. (2015). How to use open badges? *Information and Communication Technology in Education*, pages 222–230.

[Wilson et al., 2018] Wilson, L., Ho, S., and Brookes, R. H. (2018). Student perceptions of teamwork within assessment tasks in undergraduate science degrees. *Assessment & Evaluation in Higher Education*, 43(5):786–799.

[Wu et al., 2012] Wu, K., Davison, L., and Sheehan, A. H. (2012). Pharmacy students' perceptions of and attitudes towards peer assessment within a drug literature evaluation course. *American Journal of Pharmaceutical Education*, 76(4):62.

[Young and Henquinet, 2000] Young, C. B. and Henquinet, J. A. (2000). A conceptual framework for designing group projects. *Journal of Education for Business*, 76(1):56–60.