Peer Grading in Gamified Blended Learning Environments

Jessica Ribeiro, Daniel Gonçalves, Miguel Mira da Silva
Dept. of Computer Science and Engineering
IST/Technical University of Lisbon
Lisbon, Portugal
jessica.ribeiro@tecnico.ulisboa.pt, daniel.goncalves@inesc-id.pt, mms@tecnico.ulisboa.pt

***Abstract*—Blended learning and gamification have become two main concepts used in higher education, in order to improve students motivation and engagement in course activities. To use gamification as a learning approach, it is necessary to apply game elements and principles into the educational context. Game elements, such as badges or experience points, can be used as tools to assess student's learning outcomes - the professor has to continuous reward student's work as they produce new learning content. However, this type of assessment procedure cannot scale properly if the number of students is high or if the learning outcomes are too complex to handle. One typical solution to solve this problem is using peer grading as an assessment procedure. In this paper is explored if peer grading can be used as a valid assessment procedure in gamified blended learning environments. A scalable and modular peer grading plug-in was developed and instantiated in a higher education course (with blended learning and gamification implemented as educational approaches) and evaluated through data collected from the course' peer grading activity and usability tests. The retrieved data were subsequently analyzed. The results revealed no significant difference between the students and professor assessment grades, and that the plug-in system was easy to use but, in some cases, users needed some learning experience to perform more complex actions.**

***Index Terms*—Blended learning, Gamification, Moodle, Peer grading, Plug-in.**

I. Introduction

ADDING innovative and creative uses of technology to improve teaching practices have generated new opportunities in higher education [1] [5]. With the wide availability and advances in technology, a new learning concept named e-learning has attracted the attention of professors, in order to supplement or augment the traditional instructor-led classroom courses. As both the traditional face-to-face learning and e-learning simultaneously offers strengths and suffers from limitations [2], the strengths of these two learning methodologies had been naturally combined into a new learning method named blended learning.

The goal of this learning method is to join the beast features of face-to-face classes with the best features of online learning in order to promote active, self-directed learning opportunities for students with added flexibility [3] [4]. While blended learning as been used successfully in higher education [5], by improving students learning, promoting a stronger sense of learning community and a stronger peer support network, it has been observed that students participation in course activities is not voluntary and active as was hoped for - students feel a lack of motivation and engagement and only participate in course activities mainly because it is required. To better engage the students in learning activities, gamification is been seen as a solution.

Blended learning and gamification can be incorporated together with the aim of improving the overall engagement of students and providing impetus to active learning [6].

The use of gamification to assess students learning outcomes leads to a continuous assessment, i.e. as students continuously produce learning outcomes, professors will have to continuously assess them, by rewarding these student's learning outcomes. This scenario is similar to what happens in games as, for example, when a player completes a mission, he earns a badge, or when the player kills the boss he levels up and gain experience points.

In blended learning, this situation translates into a large number of moments when the professor must act. As the number of contributions from students increases, the traditional and gamified assessment procedures used to evaluate students learning outcomes become not suitable - these procedures suffers from scalability problems.

A typical solution to this scalability problem is peer grading. Peer grading has been shown in the literature to be a valid and reliable method to assess students in e-learning education [7] [8]. It offers the possibility to scale the grading of complex assignments in courses with lots of students [9]. The impact of peer assessment on gamified blended learning is promising, but its efficacy seems to depend on many factors including students attitudes, familiarity with the assessing criteria, the type of skill being assessed, and the possible presence of bias such as gender and friendship [11].

In this paper, we intend to study if peer grading can be effectively and unbiasedly used in a gamified blended learning context. To address the reliability validity of peer grading in gamified blended learning contexts, a scalable and modular peer grading plug-in was developed and instantiated in a higher education course, in which gamification and blended learning are used as educational approaches. We carefully analyzed the resulting empirical data, retrieved from the course peer grading activity and usability tests, which show that the plug-in was well accepted by the users, although to perform more complex action it is necessary some learning experience, and that the professor and student's peer grades were identical, which is a good indicator that peer grading can be used as an assessment

procedure in gamified blended learning environments.

We finalize this paper by discussing the lessons learned from this experiment and the improvements that can be done to improve the developed plug-in.

II. Blended Learning

Blended learning (also known as hybrid learning) has been an emerging trend in higher education. It can be defined as "a design approach whereby both face-to-face and online learning are made better by the presence of each other" [3], involving a combination of co-present interactions and technologically-mediated interactions between students, teachers and learning resources.

Qualitative research studies have shown that blended learning can improve learning outcomes and that students learn more in blended courses than they do in comparable traditional learning environments [3] [4] [12]. Professors responsible for the blended courses reported that students wrote better papers, performed better in exams, produced higher quality projects, and were capable of more meaningful discussions on course material [3] [4] [12]. These research studies also suggested that student's interaction and overall satisfaction have been shown to improve. Reasons for this high level of satisfaction included the increased student engagement in learning, enhanced interaction with students and flexibility of teaching [3] [4].

Blended learning can be achieved through Virtual Learning Environments, where students can take part in active and creative learning with other students through simulations, remote control of real-world tools and devices, online master classes, or collaboration with other schools or organizations.

III. Virtual Learning Environments

A Virtual Learning Environment (VLE) is a web-based communication platform, which allows students to access different learning tools such as course content, professor assistance, document sharing systems, discussion boards and learning resources, without limitations of time and place [13]. Individual instructors incorporate this web technology into their courses to complement traditional face-to-face activities [13]. Virtual Learning Environments are also known as Course Management Systems (CMS) and Learning Management Systems (LMS). These systems facilitate information sharing and communication between participants in a course: allows professors to distribute information to students, to produce course content material, to prepare assignments and tests, to manage distance classes, and enables collaborative learning in the forums [13].

One of the VLEs used in education is Modular Object Oriented Developmental Learning Environment (Moodle)[1], a free open-source LMS.

A. Moodle

Moodle is an open source VLE that helps professors create online learning communities effectively [13]. Moodle was created by Martin Dougiamas to support a style of learning named Social Constructionist Pedagogy. This social constructionist theory believes that "people learn best when they interact with the learning material, construct new material for others, and interact with other students about the material" [14].

Moodle has a set of module activities that supports the social constructionist method [14]. These activities include: (1) the creation of static course material (such as text pages, web pages or links to web content); (2) the creation of interactive material (such as assignments, single questions, lessons or quizzes); and (3) interaction between students (in chats, forums, wikis and workshops) [15].

VLEs, such as Moodle, can integrate gamification techniques to greatly increase students engagement, motivation and focus through a more immersive experience. Gamification in Moodle provides an opportunity to give more visual storytelling and impactful feedback, among with other visual cues, resulting in better learning retention [15].

IV. Gamification

Gamification is becoming a popular design strategy for engaging users in non-game contexts and can be defined as the use of game elements and game design techniques [19] that can be applied to non-game applications and processes, in order to encourage people to adopt them or to influence how they are used. Before the term existed, many designers and researchers were already exploring the role of play and fun in computer applications. As user experience became a more solid field in the 90s and 2000s, there was further work and research in this area, with people considering the role of fun and play in user experiences. One of the reasons gamification is becoming so popular is because it has been shown to engage users [16] [18], by rewarding who accomplish desired tasks. Rewards such as badges and experience points are used to elevate status by showcasing the talents, expertise, and accomplishments of users. A popular construct investigated by empirical gamification studies has been behavior change, by promoting new habits among a population to produce desirable outcomes. The results of these studies have found that adding game elements can lead to a change in behavior [19]. In order to build an effective gamified experience, it is important to understand what motivates people to play.

A. Motivation

Motivation is a complex concept that must be well understood in order to lead people to adopt certain behaviors and change practices [18]. According to Werbach and Hunter [19], "motivation is at the heart of sustained behavior change, and games are among the most powerful motivational tools". Motivation involves an interaction between a person and a task, in a particular situation at a particular time. The foundations of human motivation have been a field of study for years. Behaviorism was the first dominant theory and, it suggested that extrinsic motivation was the way to encourage to do things [19]. Rewards and punishment, systematically applied, would condition and reinforce behaviors in the anticipation of further rewards or punishments. Against this behaviorist theory, are a variety of other cognitivist theories that try to explain what is

[1]https://moodle.org

happening in the human brain such as the Self-Determination Theory (SDT). The SDT was proposed by Edward Deci et al. [21] focus on intrinsic motivation to drive human behaviors. It complements the internal desire of an individual with on how the environment should support their grown and development otherwise, the internal motivators will vanish [21].

B. Game Elements

Gamification requires game elements to be applied in a non-game context. Adding game design elements to non-game contexts may encourage desired user behavior and motivate users [19] [15]. These elements do not need to be tangible (like pieces of a board); intangible elements such as rules or score points are also important to understand and define a game.

Game elements can be viewed as a pyramid hierarchy, belonging to one of three categories, with different levels of abstraction. As referred by Werbach and Hunter [19], *Dynamics* are elements at the higher level - it is the category of game elements that corresponds to the overall view of the system and need to be carefully considered and managed but do no enter directly into the game experience. In the middle of the pyramid, *Mechanics* "are the basic processes that drive the action forward and generate player engagement" [19]. Each mechanics can be used to achieve one or more game dynamics. At the base of the pyramid, we have *Components* that are the concrete forms that dynamics and mechanics can take (such as badges, points, levels, etc.). One type of elements that are not represented in the pyramid is *Aesthetics*, which corresponds to what makes the game fun. The aesthetic element is present in the Mechanics, Dynamics, Aesthetics (MDA) framework proposed by Robin Hunicke et al. [20].

Game components, mechanics, dynamics and aesthetic have to be combined correctly to provide experiences that are engaging to the player, appealing to different kinds of fun. Different gamification elements might appeal to different types of people, which is similar to video games where different types of games can appeal to different types of players.

C. Gamification in Educational Contexts

The education has been one of the areas identified with the greatest potential to apply gamification [16] [15].

Gamification was used by Lee Sheldon, a professor at Indiana University, in the context of education, with great success [17]. Basically, he abandoned the normal grading mechanism and implemented an Experience Points (XP) system among other game mechanics [17]. Lee Sheldon created a model which he called the *Multiplayer Classroom*. In this model, students of a course become players, starting at level 1 with 0 XP. Then, they were given quests, raids, and missions that had a counterpart in the traditional educational terminology, like quizzes, tests, and assignments. In the end of the semester, by replacing the traditional grading system with experience points and leveling up, the amount of XP of each student determines his/her final grade. *Multiplayer Classroom* model was one of the first experiences in integrating game design principles into instructional design and it delivered positive results [17].

There are some technological online examples in which gamification is used for educational purposes:

- **Codeacademy**[2]**:** is an online interactive platform which teaches students to code in several programming languages. Gamification elements on Codecademy are currently focused mainly on supporting the individual learning process;
- **Khan Academy**[3]**:** is an online service where students watch online videos and perform exercises to learn about several topics. Their progress is tracked with points and badges;
- **Stack Overflow**[4]**:** is a questions-answers site for programmers to ask and answer questions about coding. Gamification is used with the intent is to develop a community that is rewarded for its contributions through upvotes and badges.

In gamified blended learning, as it happens in traditional learning, the professor is responsible for assessing students. To complement this assessment method, peer grading can be considered. Instead of just being the professor to reward student's learning outcomes, students can also reward the work of their peers, encouraging the collaboration/cooperation between students/players, which is also a main concept in gamification.

V. Peer Grading

Peer grading, also known as peer assessment or peer review, is defined as "specific judgments of ratings made by pupils about their achievement, often in relation to teacher-designed categories" [22]. In other words, it is a form of participatory assessment where students grade and/or provide feedback on the learning outcomes of other students (i.e. their peers). In the peer grading process, students assess their peers' work and provide feedback - feedback is considered to be an integral part of the learning process [23]. By providing constructive feedback, students understand and appreciate other students perspectives, and are provided with an opportunity to reflect on their peers work as well on their own.

Several research studies reported a high correlation between student and professor grading results. As example, Falchikov et al. [24] conducted a meta-analysis of 56 studies about peer grading in face-to-face learning, and found out a significant correlation between student-assigned and professor-assigned scores; Bouzidi et al. [7] investigated peer grading in an online context and found its validity to be high in this context.

Traditional assessment practices (such as professors grading individual students assignments) are impossible to do at an online student scale. To overcome this challenge, peer grading assessment is used, scaling the evaluation process to the size of even the largest classes [7].

Although the scaling properties of peer grading are attractive, there are several challenges in making it work effectively. In a study conducted by Kulkarni et al. [8], several students complained in class forums about being unable to complete

[2]https://www.codecademy.com
[3]https://www.khanacademy.org
[4]http://stackoverflow.com

peer assessments in time, and students who received an unfair assessment have lost motivation. In his study, Kulkarni et al. [8] also notice that students are generally satisfied with their overall grades, but are frustrated by inaccurate feedback from some peers. Despite all of this challenges, peer grading can have positive effects on students. Kulkarni et al. [8] also found out that peer grading can be an extremely effective learning tool; a higher percentage of students indicated that they learned more from grading the work of others students than from professor grading their work. Conducted studies also have shown that peer grading is an effective learning strategy to be used in higher education [24] [25] [26].

Making peer grading an integral part of the evaluation process in a course, not only encourages learners and professors to regard assessment as a shared responsibility, but it can also be applied to change the traditional one-way professor-centered classes to a more learner-centered one. It is believed that peer grading encourages reflective learning through observing others performances [26].

To assess learning outcomes produced by students in gamified blended learning contexts, teachers workload can be alleviated by extending the assessment procedure so that students are part of it through peer grading.

A. Peer Grading in MOOC's

Peer grading is an evaluation method mostly used in online contexts (e-learning) [9] [10] particularly in Massive Online Open Courses (MOOCs). MOOCs provide opportunities to teach to a massive and diverse student population through learning platforms such as Coursera[5], EdX[6], and Iversity[7].

Across all these three platforms, peer grading is implemented as a workflow consisting of several steps that must be completed, one after the other. The main parts of the peer grading workflow on these platforms are a submission step, followed by a peer evaluation step, and finally a result step.

In these platforms, due to different backgrounds and knowledge of participants, their eligibility and grading accuracy is doubted and peer grading itself is sometimes challenged. Therefore, it depends on the teaching teams to provide detailed rubrics to ensure the success of peer grading and to keep assessments comparable, consistent and fair [10]. Participants subjectiveness based on their education, culture, and knowledge of the assignment topic can influence to a certain degree the way a participant grades. Bias can be partially counteracted with anonymity, multiple reviews per peer and training. Another factor linked to the problem of participant eligibility are bad reviews caused by laziness, dishonesty, retaliation, competition or malice. These are always a problem for peer grading but have found to be a bigger problem in online learning due to increased anonymity and a decreases feeling of community affiliation [10].

B. Peer Grading in Moodle

In a blended learning context, Moodle is a chosen platform to online learning, since it provides to professors a set of plug-ins to customize their courses, according to their teaching methods and to the learning content of the courses. To assess students learning outcomes within a course, Moodle provides three distinct plug-ins: *Forum*, *Workshop*, and *Assignment*.

Both *Forum* as *Workshop* plug-ins allows students to evaluate their peers work; however, both present several limitations.

In the *Forum* plug-in, students can grade other students by using a numerical or a non-numerical scale but are not able to provide written feedback. Another significant problem is that all students must grade each other students submissions, i.e. it is not possible to select only a small group of peers work to be graded by a student. It is also important to notice that the identity of the peers who grade their colleague's work cannot be anonymous (which can bias the peer grade).

In the *Workshop* plug-in, as opposed to the *Forum* plug-in, it is possible to select a group of peers work to be assessed. However, these works are randomly assigned to peers, which can lead to a main problem: in a hypothetical scenario, if the peers are friends to each other, and if they belong to the same group project, they can beneficiate each other work' grades, which can bias the peer grading activity.

Another main problem, common to *Forum* and *Workshop* plug-ins, are the dropouts or disinterested students. If a dropout/disinterested student doesnt assess peers work, it can affect the peer in question, and it will continually be selected randomly by the Moodle plug-in, continually affecting other peers works.

In the *Assignment* plug-in only the professor can assess students' works. Thus it cannot be used to perform peer grading in a course.

As none of these Moodle plug-ins provides a reliable peer grading activity that can be implemented in gamified blended learning environments, a peer grading Moodle plug-in was developed.

VI. Peer Grading Plug-in

To verify if peer grading is a reliable and valid assessment procedure in gamified blended learning contexts, a scalable and modular plug-in was designed, incorporating best practices of peer grading. As a first step, it was identified a set of features that this system must meet so that peer grading can be considered a valid assessment procedure to be used in a gamified blended learning context. These features are as follows:

- Anonymity in the grading process - it must be possible to hide students identity in order to avoid bias in the peer grading activity (such as gender or friendship);
- A distribution mechanism of students work submissions by their peers - it is important to implement a distribution mechanism in which students' works are assigned to peers in an equitative and balanced way;
- A manage conflicts mechanism - to manage conflicts that may appear during the peer grading activity, such as dropouts students or students from a same group project;
- An outliers detection mechanism - professors should detect anomalies in the peer grading activity, such as outliers;

[5]https://www.coursera.org
[6]https://www.edx.org
[7]https://iversity.org

A. Technologies Used

The developed plug-in is composed by a client and a server-side. The client-side, which is composed by the pages on the user side of the application, was implemented using web technologies such as HTML, CSS, and AJAX (Asynchronous Javascript and XML). The server-side encapsulates the plug-in logic. This developed plug-in is a typical web application, following the client-server architecture which, in this case, the web browser is the client application. On the server-side, is used the PHP language for dynamically generating the pages that are presented to the user, as well as the application logic and the scripts to interact with the database.

B. Plug-in Modules

In Moodle, there is two type of modules: the *Block* module and the *Activity* module. *Block* modules are only used to display information, and the *Activity* modules are interactive tools used to engage students in learning and to assess their progress. To develop the peer grading plug-in, it was necessary to create two modules: an *Activity* module, named *PeerForum*, in which the peer grading activity takes place, and a *Block* module, named *PeerBlock* in which is possible to manage the peer grading activity that occurs in the *Activity* module. A Moodle course front page containing both these module is displayed in Figure 1.

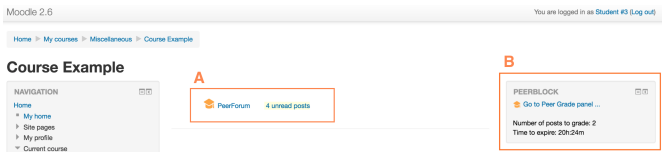

Fig. 1: Course front page with *PeerForum* (A) and *PeerBlock* (B) modules

The *PeerForum* activity module, has a structure similar to the *Forum* module. The *Forum* module is a Moodle activity in which students and professors can exchange ideas by submitting posts in a forum discussion. These *Forum*' posts can be graded by professors by assigning rates. For developing the peer grading plug-in activity module, was decided to reuse the source code of the *Forum* activity, instead of developing a *Forum* extension with the peer grade feature - this decision was made because creating new code in an existent module could generate conflicts on the module source code (provided by Moodle developers) as well in the database.

C. Peer Grading Activity

To start a peer grading activity, an assignment post discussion must be created by the professor in *PeerForum*. A student, to submit his work, has to reply to the assignment post discussion (in Figure 3 is displayed a student post). When the student' post is submitted, the process of assigning peer graders to assess this post begins - an assigning peer grades algorithm was developed for this purpose, which assigns peers in a balanced way. The peers selected to peer grade this post have to assess it by submitting a peer grade. Through this process, the professor is also able to assign or remove peer graders from a *PeerForum* student submission (after the assigning peer graders algorithm is completed) and to manage all this peer grading activity. An interaction diagram was created to demonstrate this peer grading activity, and can be seen in Figure 2.

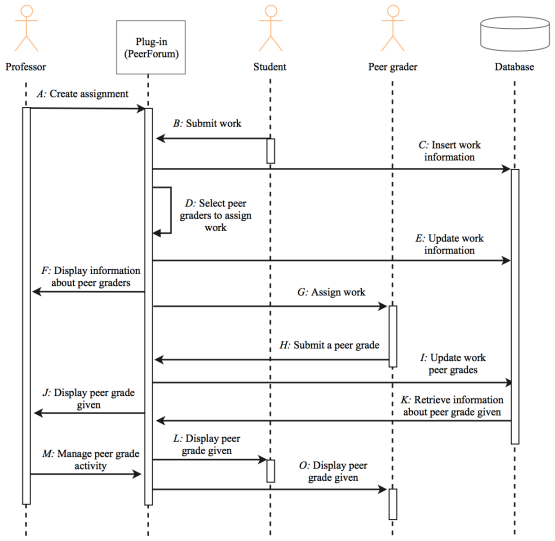

Fig. 2: Interaction diagram of peer grading activity

The developed assigning peer's algorithm follows this logic: to assign peer grades to a student post in *PeerForum*, first, a list of all students enrolled in the same course for which the post was submitted is obtained. Then, from this set of students, is verified which students dont have any conflicts with the post author. If there is no conflict between the students, the algorithm proceeds, otherwise the process to assign the student in question ends and he is not assigned to peer grade the post. Next is verified if the student is blocked to peer grade any post in the course. If he is blocked the process ends, otherwise it is verified if the post submitted was already peer graded by him. If the post is not assigned to the selected student, he becomes a peer grader of the post. If the post is already assigned to the student or if the student had already given a peer grade to the post, the algorithm ends. Otherwise, if none of these conditions is verified, the post can be assigned to the student to be peer graded by him.

The faculty staff can manage post 's peer grading activity by assigning or removing peer graders manually. Inside the student post, there is an area (Figure 3 (B)) in which are displayed the students assigned to peer grade the post as well as two selection boxes: *Assign student* and *Remove student*. The *Assign student* lists all the students that can be assigned to peer grade the post, which means that it can only be selected students who are not already assigned to the post. The *Remove studen*t selection box lists the students that are assigned to peer grade the post because only the peers assigned to peer grade the post can be removed.

The area to give the peer grade appear inside the assigned post (Figure 3 (A)) and, in this area, there is a selection box, in which the peer grader has to choose the appropriate grade to give and a textbox in which the peer grader has to write a

feedback text.

In the *PeerForum*'s peer grading activity, a professor can also give a grade to a student post, by giving a rate. This feature was reused from the *Forum*'s module, however, some code changes have to be made in order to avoid conflicts between the existing Forums table in Moodle database, such as the creation of a new table to contain only the rate information of our peer grading plug-in.

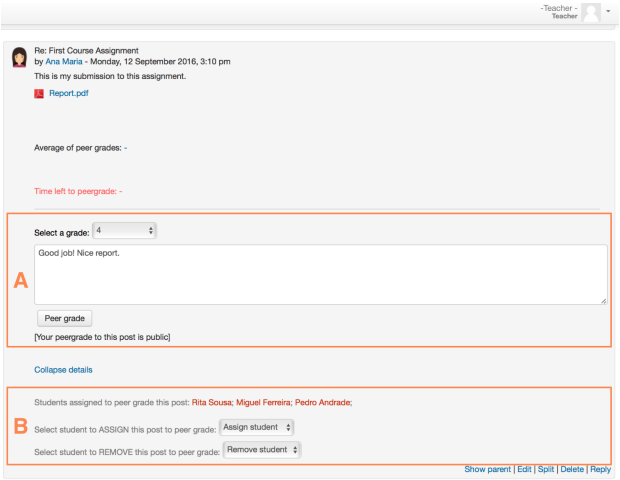

Fig. 3: Interface for peer grading a student's post

The students can see all the posts in a PeerForum and its respective final peer grades, which is displayed inside each post (Figure 4). However, the ability for a student to see the individual peer grades given to a post can be configured by the faculty staff in the PeerForum local settings (i.e. it can be configured for only the professor to see the peer grades given, instead of being public, for example). In order to avoid the users to be influenced by other grades and feedbacks already given when grading a peer post, they can only see the given peer grades after submitting his own peer grade. After peer grading the post, the user can edit it by changing the grade or the feedback he gave, in a time period of 30 minutes. After that time period passes, the peer grade cannot be changed or eliminated.

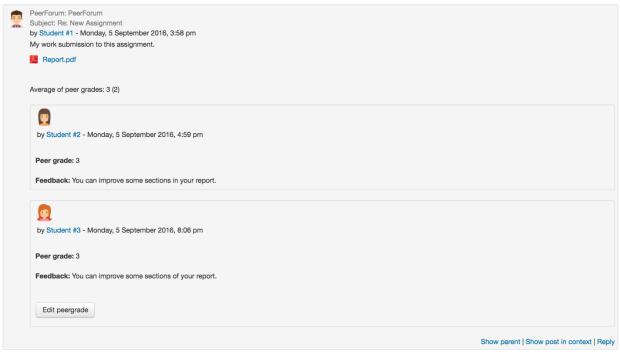

Fig. 4: Interface for displaying post's peer grades

D. Peer Grading Activity Management

To manage the peer grading activity inside a course was decided to develop a new *Block* module named *PeerBlock*, which can be accessed by professors and students. However, the interfaces, as well as the actions that can be performed on the *PeerBlock*, are significantly different for both these users. The *PeerBlock* acts like a management panel, where information can be visualized and actions can be performed to control the peer grading activity of a course. Is noteworthy that a course can have multiple *PeerForum*s, but can only have one *PeerBlock*, which is used as a management panel for all the *PeerForum*s peer grading activity.

The student management panel allows the student to have control on his own peer grade activity, which is taking place on a *PeerForum*. This panel is composed by three sections:

- **Posts to Peer Grade:** in this section is displayed all the posts assigned for the student to peer grade. The posts are shown as in the *PeerForum* with an area to select a grade and to write a feedback (Figure 3). After having been peer-graded, the post is removed from this section, and it is inserted into the section *Posts Peer Graded*;
- **Posts Peer Graded:** in this section is displayed all the posts the student had already peer graded as well as the peer grades given by the other students that had the same post assigned (Figure 4);
- **Posts Expired:** in this section is displayed all the posts that the student does not peer graded in a time period defined, i.e the deadline to give the peer grade has expired. This deadline can be configured by the faculty staff.

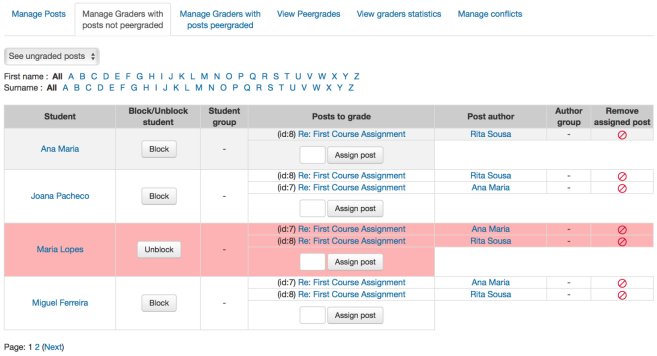

Fig. 5: Interface of *PeerBlock*'s panel

The professor management panel is much more complex than the student management panel, and it could be considered as the back-end of the peer grading plug-in. This panel is composed by six sections, which can be seen in Figure 5, and are as follows:

- **Manage Posts**: in this section is displayed all the existent *PeerForum*s posts of a course. The professor has a global view of which posts had already been peer graded or not, and can assign or remove peer graders to *PeerForum*'s posts;
- **Manage Graders with Posts not Peergraded**: in this section, the professor is able to manage the students who didnt had already peer graded posts assigned to them;
- **Manage Graders with Posts Peergraded:** as opposed to *Manage Graders with Posts not Peergraded* section, in

this section the professor is able to manage the students who had already peer graded posts assigned to them;
- **View Peergrades:** the professor can see all the peer grades given in all course' *PeerForums*;
- **View Graders Statistics:** in this section is displayed student's statistics regarding their peer grading activity in a course;
- **Manage Conflicts:** this section is used to manage conflicts, such as avoiding having students from the same group peer grading each other posts submissions in a *PeerForum* assignment.

An important feature developed for the peer grading plug-in was the outliers detection algorithm. In a peer grading activity can exist deviations in the peer grades given by the students to a *PeerForum* post. In this case we are in the presence of outliers - values that lies outside (are much smaller or greater than) most of the other values in a set of data. To identify the outliers, the plug-in can calculate the standard deviation of all the assigned grades or can detect obtuse grades. These outliers can be seen in the tab section *View Peergrades* of the *PeerBlock* panel. In this section, grades are categorized by colors: if red it is an outlier; if yellow it is a warning; and if green is the mean of all grades (a non-outlier).

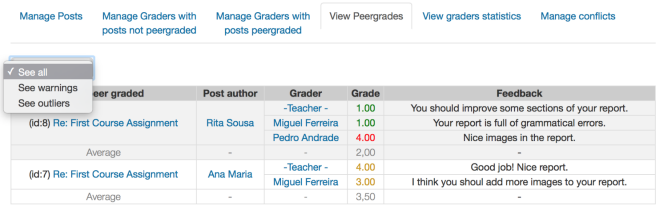

Fig. 6: Interface of *PeerBlock*'s panel in *View Peer Grades* section

VII. Usability Evaluation

To evaluate the developed peer grading plug-in, usability tests were conducted. The number of participants who underwent through experimental evaluation was 22 in total, of which 19 were students (9 undergraduate students and 10 graduate students), with ages ranging from 20 to 30, and 3 were professors (2 university professors and 1 secondary teacher), with ages ranging from 41 to 55. From the 22 users, 8 were female and 14 were male. Only 5% of the users never used the Moodle platform, and 23% didnt know the peer grading concept. From the 3 professors, only 1 didnt ever peer grade a students work, and from the 19 students, 8 students (42%) didnt ever peer grade a colleague's work. Different tasks were given to professors (8 tasks) and students (5 tasks). After the usability tests, questionnaires were given to users. The questionnaire encompassed several areas, such as application usefulness, data quality, easiness to use, motivation level, among others.

The processed results obtained from the usability evaluation performed by students can be seen in Figure 7 Figure 8 and in Figure 9, in which are respectively presented the student's success rate, the errors given and the execution time in completing tasks. The results obtained from professor's usability evaluation tests can be seen in Figure 10, Figure 11 and Figure 12, in which are respectively presented the professor's success rate, the errors given and the execution time in completing tasks.

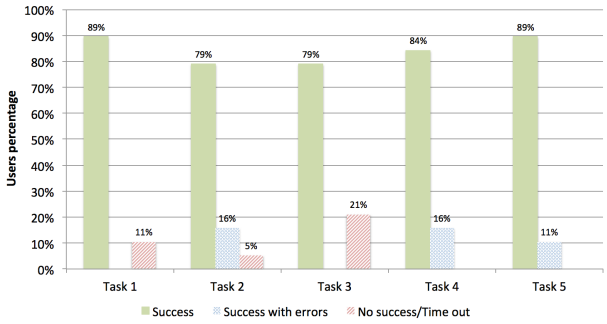

Fig. 7: Students' success rate in completing test tasks

Fig. 8: Students' errors per task

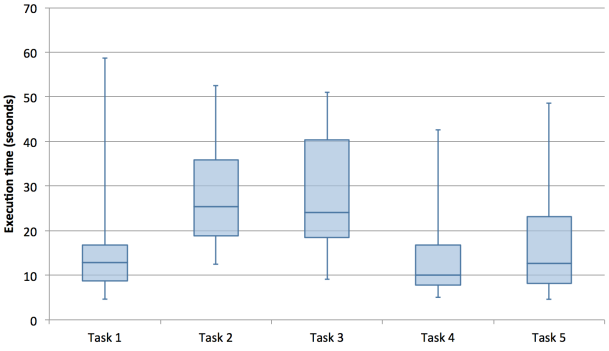

Fig. 9: Students tasks' execution time

Results show that the complexity of some actions performed by professors can involve some learning experience, instead of the actions that was performed by the students, which are much more intuitive to be executed. Despite some actions involving a certain complexity to be performed by users, in overall it has been observed that users were pleased with the simplicity of the interface and ease in executing several actions after some experimentation. The most difficult task to be performed by students (task 3) was related to editing a peer grade given to a post - results show an unsuccessful completion rate of 21% and a median execution time equals to 24,14 seconds. The easiest (task 4), with a success rate of 84% and a median time of 10,05 seconds, was related to seeing all the peer grades given in the *PeerBlock* panel.

The most difficult task to be performed by professors (task 7) with a null success rate (it was not successfully performed by any professor) was related to blocking a red outlier in the *PeerBlock* panel - to avoid a peer grade to affect the final student grade, peer grades can be blocked in the *PeerBlock* panel. The easiest task performed by professors with a 100% success rate, and with the lowest execution time and number of errors, was task 3 in which professors had to assess a student post in a *PeerForum*. In overall, the usability tests showed a good acceptance by users in performing the tasks.

To measure the system usability, we used a widely used measurement scale - the System Usability Scale (SUS) [27]. By compiling user scores, we obtain an average of 78.64, with a standard deviation of 15.69 and a median value of 81.25. In a scale from 0 to 100, the lowest user-value obtained was 25, and highest was 97.5. The majority of the individual user values were high, meaning that most of the users found the system very usable.

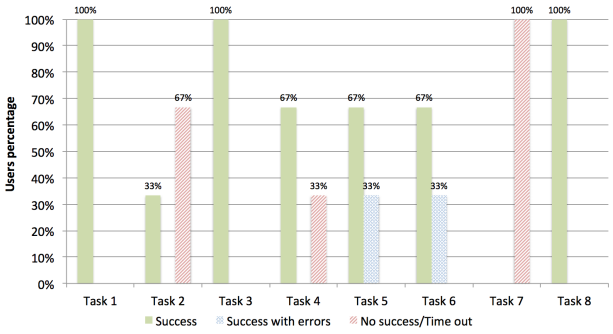

Fig. 10: Professors' success rate in completing test tasks

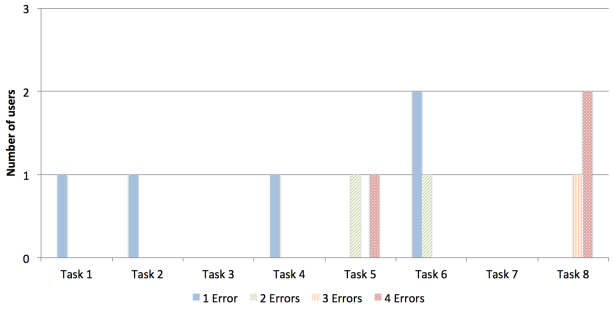

Fig. 11: Professors' errors per task

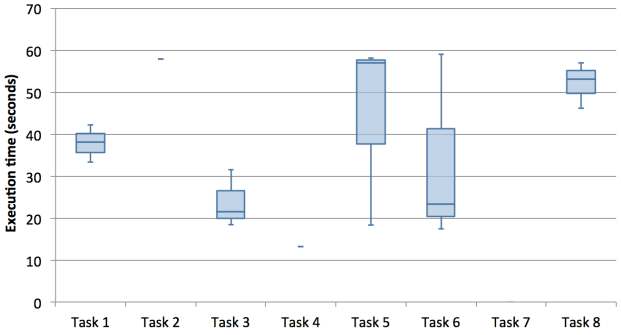

Fig. 12: Professors tasks' execution time

VIII. Peer Grading in a Gamified Blended Learning Environment

The demonstration of the developed plug-in was be performed in a MSc. Course named Multimedia Content Production (MCP). It is an annual semester-long course in Information Systems and Computer Engineering and is taught at Tecnico Lisboa university. This higher course uses gamification and blended learning as educational approaches. A preliminary version of the peer grading plug-in was deployed during the academic year of 2015/2016 in MCP course, for two months (8 weeks). In this initial study, to evaluate the developed peer grading plug-in, the peer grade was not used as an evaluation criterion. To motivate the students to voluntary use the developed plug-in, an extra 100XP in the final grade were given to those who participated in the peer grading activity. The course had 85 students enrolled, 16 female and 69 male, of which only 53 participate in our study voluntarily. The data related to grading scores, retrieved from the posts submitted in the MCP *PeerForum* assignment subjects, went under statistical analyses.

Given the importance of the peer grading activity and its impacts in a gamified blended learning environment, main research questions were formally stated and answered (after carrying all the required calculations) as followed:

1) **How similar are the professor's rates and the peer grades given to posts assignments?:** to answer this question, it was calculated the similarity between professor's rates and students peer grades, using the paired-sample t-test method (Table I). The value obtained ($P-value$ = 0.874) indicates that there is no significant difference between these two grades. The mean scores of peer and professor were also quite close to each other;
2) **How similar are the grades given to posts assignments?:** although the means of these two grading scores are similar, there can exist little agreement on individual grades. It was necessary to measure how well peer grades match professor's rates by calculating its concordance, using the interclass correlation coefficient (ICC) model. The value obtained (ICC = 0.59) indicates that exists a moderate correlation between this two type of grades. It was also calculated the variance among peer grades, which value obtained (ICC = 0.83) indicates that the peer grading scores given between students vary little. By comparing these two calculation scores, it is possible to conclude that exists much more disagreement between the grades given by students and professors, than the ones only given by the students;
3) **Can peer grading provide a reliable and valid assessment method in a blended learning environment?:** to answer this last research question, it was calculated the similarity between the final peer grading scores and the final professor rating scores, using the Pearson product-moment correlation coefficient method. The value obtained (r = 0.33) means that there is a medium correlation between the rates and the peer grades given. Although this value seems low, it could result from the fact that the grading scale used in the

MCP course was very small/compressed (from 1 to 4). To better understand the similarity between the final peer and professor grading, we analyzed more in detail each grade given. In 101 of student *PeerForum* posts submissions graded by the professor, only 65 peer grades were identical to professor's rate scores; this results can be seen in Figure 14. It is possible to observe that in most cases there was an agreement between professor and peer grades (65 equal grades), and in the remaining cases (36 different grades) the difference between the professor and peer grades was only by one value. There were 4 exceptional cases where the professor gave a 4-rate and the average of peers grades was equal to 2. These situations happened due to peers assigning a 0-score to posts and by giving bad reviews. By analysing this graph, is possible to conclude that the average of peer grades given in *PeerForum*' posts varies no more than 0.5 value from the professor's rates (in Figure 13 is possible to observe the mean of professor and peer grades per post assignment subject - in MCP course, each *PeerForum* corresponded to an assignment subject).

	Paired Differences							
				95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
PeerGrade - Rate	,00990	,62442	,06213	-,11337	,13317	,159	100	,874

TABLE I: Paired-samples t-test for professor's rates and peer grades

There are situations in which professor and peer grades given to students posts are significantly differents (such as the 4 cases identified above), which results from the fact that peers are sometimes unfair in assigning grades to their colleagues works. Thus the expected bias when grading colleague's work assignments does not occur. Instead, students seem to give grades similar to professor rates, as it is possible to observe in the results obtained.

The validity of several assessment metrics, such as the mode, maximum, minimum, count, and summation of peer grades were also determined, by their similarity to professor-assigned rates, using the ICC method. The higher ICC obtained was between professor's rates and average of peer grades (r = 0.59). This result indicates that the average of peer grades is the assessment metric with less difference to professor's rates, i.e. it is the metric that best fits in this learning environment. The remaining metrics (Mode, Min, Count, Sum) have almost no correlation with professor's rates.

IX. Discussion

The results obtained by analyzing the data collected regarding the student peer grading activity in MCP course, using our plug-in, suggests that, in general, the joint efforts of multiple peer graders can produce fairly consistent grading results in a gamified blended environment. These preliminary results suggest that peer grades were very similar to professor's rates, although there is little variance between them. In most of the cases, the difference between the rates and the peer grades are equal to 1 (i.e. when peers tend to give 3, the professor's rates with 4) and in general, the rate given by professor is superior to the peer grades given by students. It was possible to notice situations in which professor and peer grades are significantly differents, which resulted from an unfair assessment by students to their colleagues. Thus, the expected bias (resulted from gender or friendship) doesnt occur. Instead, students tend to give similar grades to professors rates. In this study, we could also verify that there is more disagreement between professor's rates and peer grades, than amongst peers grades.

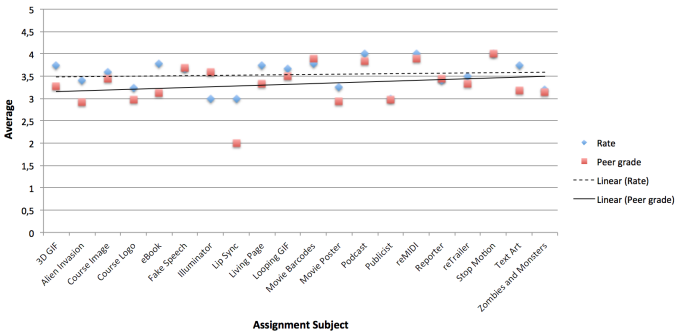

Fig. 13: Average of ratings and peer grades in different subject's assignments

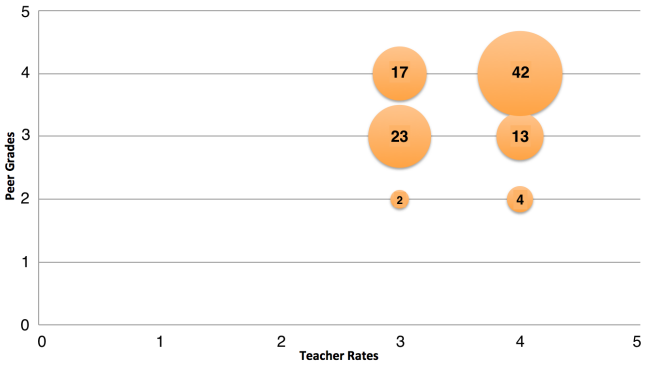

Fig. 14: Accuracy of peer grading

The performed usability tests allowed us to retrieve important metrics related to the plug-in usability. Being in a time period in which user experience on using system applications is in focus, we notice that our developed system was well accepted by the users. It was noted that all the users feel confident in using the plug-in system and were able to use it in educational environments where gamification and blended learning are implemented. The complexity of some actions that must be performed by the professor can involve some learning experience, instead of the actions that must be performed by the students, which are much more intuitive. Despite some complexity in performing certain actions, in overall it has been observed that users were pleased with the simplicity of the interface and ease in executing several actions after some experimentation.

However, it was also possible to notice that there is still plenty of work that can be done to improve this system, taking

into account the several comments and suggestions given by users, collected throughout the usability tests and final questionnaires. These improvements are related to a couple of small corrections in the plug-in interface and in mechanisms to facilitate the management of the peer grading activity.

In the overall, this study shows us that the peer grading can be used as a valid assessment procedure in a gamified blended learning environment. In addition, no bias presence was detected, since the grades given by students and professors were very similar.

X. Conclusion

Blended learning, as well as gamification, have been emerging trends in higher education. These two concepts together create an educational environment where classroom face-to-face learning experiences are integrated with online learning experiences, complemented with game elements and mechanics to motivate students. Despite all of the benefits of combining these two learning concepts in an educational context, theres still the challenge of the assessment procedure scalability. To solve this problem, we developed a new Moodle peer grading plug-in. In this paper, we studied the extent to which peer grading works as an assessment procedure in courses that have gamification and blended learning as educational approaches implemented, by instantiated it in a higher gamified blended learning course.

The obtained results lead us to believe that peer grading plug-in can make an integral part of evaluation procedures in gamified blended learning environments, encouraging learners and professors to regard assessment as a shared responsibility.

As future work, there are some aspects that can be improved in our peer grading plug-in. Regarding the complexity of performing certain actions, some mechanisms can be implemented in order to facilitate the user experience. To have more accurate results, more data can be collected by using the developed peer grading plug-in in other high education courses, by a larger period of time, and by using a grading scale with a bigger range. We have preliminary data about how well the system works, but as adoption increases, we can have more data about the system and its learning benefits.

With the continuous use of the peer grading plug-in, it is expected that this peer grading solution reduces professors workload (because grading each student work individually in the Moodle will be no longer necessary), engage students to participate in the grading process, and make the process of assessing and rewarding students fairer and equitable.

References

[1] N. Vaughan, *Perspectives on blended learning in higher education*, International Journal on ELearning, 6(1):81, 2007.

[2] W. Kim, *Towards a definition and methodology for blended learning*, The Proceedings of Workshop on Blended Learning, pages 18, 2007.

[3] C. Garnham and R. Kaleta, *Lessons learned from the hybrid course project*, Teaching with Technology Today, 8(6):17, 2002.

[4] J. F. Wang and M. Choy, *Blended learning for programming courses: A case study of outcome based teaching & learning*, Blended Learning, 30:3041, 2007.

[5] D. R. Garrison and H. Kanuka, *Blended learning: Uncovering its transformative potential in higher education*, Internet and Higher Education, 7:95105, 2014.

[6] M. Tan and K. F. Hew, *Incorporating meaningful gamification in a blended learning research methods class: Examining student learning, engagement, and affective outcomes*, Australasian Journal of Educational Technology, 2016, 32(5): 19-34.

[7] L. Bouzidi and A. Jaillet, *Can online peer assessment be trusted?*, Educational Technology & Society, 12(4):257268, 2009.

[8] C. Kulkarni et al. *Peer and self assessment in massive online classes*, Design thinking research. Springer International Publishing, pages 131-168, 2015.

[9] J. Swinglehurst and T. Greenhalgh, *Peer observation of teaching in the online environment: an action research approach*, Journal of Computer Assisted Learning, 24:383393, 2008.

[10] D. T. Staubitz and C. Meinel, *Improving the peer assessment experience on mooc platforms*, Proceedings of the Third (2016) ACM Conference on Learning@ Scale, ACM, pages 389398, 2016.

[11] M. Azarnoosh, *Peer assessment in an efl context: attitudes and friendship bias*, Language Testing in Asia, 3(1):1, 2013.

[12] R. Boyle and P. Pickard, *Using blended learning to improve student success rates in learning to program*, Journal of Educational Media, 28(2-3):78165, 2003.

[13] J. Cole and H. Foster, *Using Moodle - Teaching with the Popular Open Source Course Management System*, OReilly, 2005.

[14] W. Rice, *Moodle E-learning Course Development. A complete guide to succeful learning using Moodle*, Packt publishing, 2006.

[15] G. Kiryakova, N. Angelova and L. Yordanova, *Gamification in education*, Proceedings of 9th International Balkan Education and Science Conference, 2014.

[16] J. Lee and J. Hammer, *Gamification in education: What, how, why bother?*, Academic Exchange Quarterly, 15(2):15, 2011.

[17] L. Sheldon, *The Multiplayer Classroom: Designing Coursework As a Game*, Course Technology, 2011.

[18] B. Burke, *Gamify: How Gamification motivates people to do extraordinary things*, Routledge, 2014.

[19] K. Werbach and D. Hunter, *For the Win: How Game Thinking Can Revolutionize Your Business*, Packt Wharton Digital Press, 2012.

[20] M. Hunicke and R. Zubek, *MDA: A formal approach to game design and game research*, Work. Challenges Game AI, 2004.

[21] E. Deci and R. Ryan, *Handbook of Self-Determination Research*, Boydell & Brewer, 2004.

[22] P. M. Sadler and E. Good, *The impact of self- and peer-grading on student learning. Educational Assessment*, 11(1):131, 2006.

[23] S. K. Topping, E. Smith and A. Elliot, *Formative peer assessment of academic writing between postgraduate students*, Assessment & Evaluation in Higher Education, 25(2), 2000.

[24] N. Falchikov and J. Goldfinch, *Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks*, Review of Educational Research, 70(3), 2000.

[25] K. Topping, *Peer assessment between students in colleges and universities*, Review of educational Research, 68: 249276, 1998.

[26] H. Saito, *Efl classroom peer assessment: training effects on rating and commenting*, Language Testing, 25:553581, 2008.

[27] J. Brooke, *SUS-a quick and dirty usability scale*, In Usability evaluation in industry, pages 189194. 1996.