

GradeMe – Studying peer-grading in in-person learning

Madalena Assembleia
madalena.assembleia@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

October 2019

Abstract

The use of peer grading in educational environments has been studied for several years and the results showed it is possible to obtain similar grades between students and professors, making the assessment reliable and also giving the possibility to increase the scale of given courses as this assessment method alleviates some of the grading burden from professors, allowing for more courses that use blended learning with gamification capacities. Current research, however, has not investigated if social relationships between students affect grading. Our work focuses on whether social relationships between higher education students and their sociometric status affect the accuracy and quality of peer grading in a gamified blended learning environment. We used an improved version of an existing plug-in, meant to bring peer grading capacities to the Moodle platform. This new version includes features that allow the collection of information regarding the student's relationships with their peers. This plug-in was then used in the Multimedia Content Production course and, at the end of the semester, the data was retrieved and analyzed. We did not find any indication of significant bias in the peer grading assignments from students which had shown preference or displeasure towards a given colleague. Students from different sociometric status groups showed similar matching grades with the professors overall, though certain groups showed more inclination to provide deflated grades and other groups to inflate. In short, we conclude that, in this context, peer grading can be used as a legitimate, unbiased form of grading.

Keywords: Peer grading, Blended Learning, Virtual Learning Environments, Gamification, Moodle.

1. Introduction

Blended Learning, has shown to be a successful procedure in higher education courses, displaying improvements in students' learning and an overall stronger sense of community (Tsai and Tsung, 2007). It provides a way for students to interact directly between themselves and their professors, adding virtual learning environments (VLEs) as a classroom expansion to provide to many students a great educational experience. Furthermore, professors have been including gamification features in blended learning environments to increase student engagement and motivate them to participate in class activities.

However, one of the main challenges in this approach is student evaluation as these course usually have a large number of enrolled students, making it unfeasible and costly for professors to grade all assignments delivered by students in time. Thus, it is essential to come up with an alternative solution to traditional student evaluation.

When dealing with more open-ended exercises, manual assessment is necessary. In most cases, this kind of exercise is the most important for student learning and outcome of the course. Peer

grading – evaluation of student work done by other student(s) – is a promising solution to help the problem of evaluation in large scale.

Previous work has shown that peer grading can be a valid and reliable assessment method, from primary schools to higher education (Topping, 2009; Tseng and Tsai, 2007; Taylor, Ryan & Pearce, 2015). It offers the chance to scale the grading of more complex assignments. However, in blended learning environments, students already know each other and have created social relationships. This leads to the possibility of bias from certain students towards others, an issue that has been raised before (Taylor, Ryan & Pearce, 2015), though there are no studies about how social relationships between students influence peer grading.

1.1. Research Problem

The use of blended learning environments, with the use of VLEs such as Moodle, has shown to increase student engagement and motivation. Including gamification systems that students are already familiarized with, helps increasing the participation in the course. The more participation and work submissions increase due to these systems, the harder is to scale all work with the current tra-

ditional assessment procedures.

The integration of peer assessment seems to be the best solution for resolving this problem. Peer grading can lead to the decrease of professors' workload, sharing of information between students that is mutually beneficial and furthers involves students in the course as it is a continuous procedure. Still, peer grading requires students to grade peers which they have an established social relationship with, possibly influencing the result of their peer assessment and induce bias in the final grade of students' work. Therefore, the aim of this work is to investigate if peer grading can be a valid and unbiased assessment method in a gamified blended learning environment.

To answer this question, we will be using a previously developed peer grading plug-in to be used in a higher education course that is taught using blended learning through the Moodle platform. While the plug-in already provides to ability for students to peer grade their colleagues, some features will be included to measure social relationships between students.

1.2. Contributions

State of the art. To develop this study, we started by gathering and reviewing several research documents related to the current offerings and problems of peer grading in various learning environments.

Development of Peer Grading Plugin. A plug-in was already developed to allow peer grading in a gamified learning environment, however, it contained had some issues, which were fixed. A training system was also implemented where students could consult the requisites each submission had plus some examples of previous submissions and their respective grading. A peer nomination and peer ranking system was also implemented.

Study and analysis of results. The peer grading plug-in was instantiated it in a higher gamified blended learning course. The user activity was retrieved and analyzed in order to assess the applicability of the developed peer grading solution. By compiling and analyzing the peer grades given, together with the results of the peer nominations and peer ranking, it was possible to verify that no significant bias was present - good indicators that peer grading can be used as an assessment procedure in gamified blended learning environments.

2. Related Work

In this chapter, we present the results from searching literature around peer grading and its actual state. First, the concept of peer grading is presented in more detail. Blended Learning and how it has been applied through virtual learning environments, such as Moodle, will be then explained. Learning environments that use this platform can

also include gamification elements, which work as a mechanism to enhance the learning experience and create a community environment where students feel more motivated to engage and can learn from each other.

2.1. Peer Grading

Peer grading or peer assessment is defined as "an arrangement for students to considerate and specify the level, value or quantity of a product or performance of other equivalent students" (Topping, 2009). Peer grading has been implemented in several subjects and it can be applied in all types of work. It allows students to learn through its peers' work and their problems, as well as provide feedback in bigger quantities and immediacy than feedback from professors, which could be later used to generate future problems or serve as exemplification (Gehring, 2001).

In order to use peer grading as a reliable evaluation method, it would be ideal if the grade awarded by the student was the same as the professor grade, which can be easy to achieve when the questions are in the form of multiple choice or short direct answers.

Still, the most prevailing goal of peer assessment is providing feedback to the learners. Feedback that is thoughtfully and constructive, can help reduce errors and enhance the learning experience. By doing this, both sides gain from doing this type of activity. Feedback from students is available in bigger quantities and immediacy than feedback from professors, which could compensate for the lesser experience and knowledge of assessment from students. When students believe that their peer grade feedback accounts for their own final grade, it improves feedback quality provided by the reviewers and leads to a higher consistency on ratings.

Peer assessment has also showed to be more reliable when done with training, exemplification, professor assistance and monitoring and leads to an overall increase in reliability and validity of the grading. It has also been shown that a student with lesser knowledge than its professor, but more time to evaluate can produce a similar evaluation (Topping, 2009).

The use of peer grading also makes the assessment process more understandable for students, since instead of relying solely on professor judgement or objective test scores, becomes the reaching of consensus between individuals (Taylor, Ryan & Pearce, 2015). In higher education environments, it can serve to help students in adjusting to university, as it provides a way for students from all kinds of cultures and educational backgrounds to communicate and have some type of assistance.

Peer assessment provides a way for students

to plan their own learning, identify what are their strengths and weakness, understand in which parts need to be revised, develop metacognitive and professional transferable skills (Gehringer, 2001) and enhance their reflective thinking and solving abilities during the learning experience and finally increase students' interpersonal relationships in the classroom (Tsai and Tsung, 2007).

Peer grading still can have some drawbacks. Students are used to the idea of having a professional be responsible for their evaluation and can dislike or even reject the idea, as they see evaluation as a responsibility of the professor and not their own (Topping, 2009). The task of having to evaluate another student, which also happens to be a friend, can lead to lesser exact grade, as it provokes a conflict of loyalty. Factors such as reward difference between professor and student, the idea from some students that they lack the necessary skills, problems related to validity and reliability, bias and fairness can lead to inferior seriousness and motivation from the student (Taylor, Ryan & Pearce, 2015). Finally, it is worth mentioning that some for these studies found a certain bias in the grades of lower level students, which tend to under grade their higher-level peers (Topping, 2009).

In terms of concerns from the staff, the difficulty around how to prepare the curriculum so it includes peer review that leads not only to student learning and professional development but also to interest in participation (Taylor, Ryan & Pearce, 2015). Professor workload may also increase greatly, especially in large class, if the peer assessment is done inefficiently. In an effort to solve some of these drawbacks, several approaches were taken, and Web-based platforms started to include special features.

Overall, peer grading looks a promising alternative to professor-only evaluation in education, especially in higher education settings. In blended learning environments, peer grading gathers all advantages of peer grading achieved in traditional environments and fully online environments and helps removing the biggest challenges for professors and students when using this approach.

2.2. Blended Learning

Blended Learning is defined as a combination between face-to-face instruction with computer-mediated instructions. Blended learning offers a way to transcend the conventional paradigm of classroom. It maintains the possibility of direct communication between professor and students and removes the limitations of traditional learning, by adding features available in fully online courses, and combines it in the most complete and engaging learning experience for students, especially in

a time where online presence is very strong. As technology advances, in the future, blended learning systems will distinguish themselves according to how they allow/create blending (Garrison & Vaughan, 2008), especially with the growth of virtual learning environments.

2.2.1. Virtual Learning Environments

VLEs appear as a remote learning experience that offers an organized way of distributing learning materials in an electronic and more efficient manner, without insulating its users from each other or its professors (Chou & Liu, 2005). Moodle is an example of a VLE commonly used in fully online and blended courses, due to the enormous capacities it offers to educational settings.

2.2.2. Moodle

The Modular Object Oriented Dynamic Learning Environment, known as MOODLE, is an open source course management system used by education systems to provide the addition of Web technology their courses (Cole & Foster, 2007). The platform can be used exclusively as an online course or as an enhancement to traditional learning.

The platform allows for an easy exchange of information between students dispersed geographically and can be accessed anytime. It is also easy to reconfigure so that professors can evaluate students in a wide range of ways, provide lots of complementary tools to support the learning and teaching process. Lastly, Moodle includes tools for controllable management of tasks according to the timetable of certain courses.

The use of a gamification methodology in Moodle is easily added through some tools which are mostly included in Moodle already such as avatar/profile picture, visualization of student progress, display quizzes results, levels, feedback, badges, etc.

This platform it is already used in the MCP course and offers the opportunity to get more concise feedback, ensuing in a better learning experience. Peer grading has been included in this course before (Ribeiro, 2016), through the use of a plug-in, in which several students were given assignments and graded their colleagues. This activity included a gamification component. For their participation, they would be awarded with a badge and extra XP, which were added as a bonus to students' final grade.

2.2.3. Gamification

Gamification is the adoption of game experiences and mechanics in non-game activities, such as leaderboards, badges, etc., leading to the simulation of experiences and practicing the acquisition of

skills. This idea is becoming a common feature in businesses and education (Johnson et al., 2013).

The increasing use of gamification in educational courses recognizes that games stimulate the productivity and creative inquiry between students (Johnson et al., 2013). It allows for entertainment that is awarded with knowledge and new skills that can be tracked.

Peer grading could be included in this context in various ways, for instance, using badges as a reward to their peers' work (O'Connor & McQuigge, 2013), offering a way for students to reward each other as well as enabling better communication and cooperation.

3. Sociometric Analysis

Sociometry is a measurement based on the degree of affiliation amongst a group of people. It is a powerful tool for assessing the hidden dynamics that create the structure of certain groups (Rostampoor-Vajari, 2012). Previous research conducted in terms of sociometric analysis showed that occupying a central position in the network of informal peer groups, having friends and having a certain sociometric status, that is, the degree to which someone is liked or disliked by their peers in a group, relates to unique behavioral reputations (Gest, Graham-Bermann & Hartup, 2001), which can lead to bias, whether it is intentional or not, in the assessment of another peers' work.

Classifying people into social status groups (Terry & Coie, 1991), can be done either through peer ratings – the assessment of relationships achieved through peers rating each of their peers in terms of much they like them or through peer nominations – each peer classifies a small number of peers in a group of who they like the most and who they like the least. The status a person belongs to can be determined simply from the totals of positive and negative nominations or the two totals can be combined and create a value of social impact and social presence. It is also possible to combine peer ratings with peer nominations by replacing the negative nominations and instead use 1 a rating in a 5-point rating scale.

To reach the objective of this work, it was required a system that could get the most truthful sociometric groups possible, even if meant leaving some students unclassified and should be one that does not need for all students to know each other, as the course where the study will occur is constituted by hundreds of students. While it is infeasible to use a typical rating-scale method, that could also lead to higher error levels, since the core of the population in these course are students from a relatively stable classroom, peer rankings could also provide useful extra information,

though peer nomination seems the most relevant and feasible method to verify whether relationships between students have an impact in the grading provided by students to their peers. Therefore, the Coie et al.(1982) method, hereafter described as the CDC system was used for this study.

It consists in a bi-dimensional system which uses peer nominations to group a given population into five main sociometric status: popular, rejected, controversial, neglected, average. In this system it is possible for a group of the population to be left unclassified. To calculate to which group each person from a given population belongs to, each absolute frequency of positive and negative nominations is standardized, creating an liked most (LM) and liked least (LL) value. With these values, it is possible to calculate the social preference (SP) and social impact (SI) of each person, as LM-LL and LM+LL respectively. Each sociometric status group is then formed according to their respective formulas based on the values computed previously.

For this study, students were asked directly how they feel towards certain students. The new version of the plug-in includes mandatory peer nominations previous to peer grading and peer ranking will be prompted as an optional inquiry available to do throughout the whole semester. The details in how these metrics will be presented to students are explained in more detail in Chapter 5.

4. PeerForum Plugin

Peer grading has exhibited to be a valid assessment method in various learning environments. The objective of this work is to verify the validity and reliability of peer grading in a gamified blended learning environment and investigate the effect of social relationships on students' peer grading. To do this, we included a peer assessment option, through a plug-in.

This study required for two versions of the plug-in to be developed – 2.6 and 3.5 versions. Both versions have the same features, the 3.5 version corresponds to the most recent version, yet the 2.6 version was the one used during the study for compatibility reasons with the version of Moodle used in the higher education course.

In the following sections, a brief description of the current state of the PeerForum plug-in is given, followed by the description of the work done to fix the plugin known problems and development of new capabilities.

4.1. Peerforum Initial State

The PeerForum plugin offers peer grading capacities to Moodle (Ribeiro, 2016). It was created for students to evaluate their peers' submissions in Moodle, through the activity module PeerForum, and was responsible to hand out work

equally throughout the students. It also includes the PeerBlock module, responsible for the management of all activities. Both students and professors have access to this module, with different interfaces and actions available for each group. The students use the block to easily peer grade all posts given to them, check previous peer grades or expired posts.

The professor can visualize all information related to the peer grading activity. The professor was also responsible for creating new assignments, provide instructions and the respective peer grade criteria. After this, every enrolled student had the option to submit their work related to the assignment. Each submission was then given to a group of students automatically by the system.

Students assigned to peer grade have some time to provide a grade according to the rating scale defined by the professor, together with some feedback. The professor also rates and provides some feedback. After all parts provided a grade, the final grade is presented to the evaluated student.

However, some performance and usability issues hinder the plug-in from being used in its current state. Additionally, it did not include any way to gather and extract additional information regarding students' relationships and feelings towards each other. All these described issues were addressed and fixed for this newer version of the plugin. Every change made to the plug-in is described in the next section.

4.2. Peerforum Adjustments

In this section all corrections made to the previous version of the plug-in are addressed. Some of these issues were left reported and others were found while testing the system functionality and during its usage period.

In order to solve the unbalanced attribution of peer grades, the algorithm now prioritizes students which have participated less, i.e., the lesser number of posts peer graded and to peer grade, to be selected when a new post is submitted in the PeerForum. The student with the lowest resulting sum is then assigned to peer grade the post. This way students have the opportunity to make up for peer grades that they let expire and all have a similar number of peer grades done. Students who do not participate can be blocked by the professor as to avoid being continuously chosen.

To achieve an increase of the plug-in performance, some elements of the page were not loaded at the beginning and only when required through AJAX, allowing for less server requests each time a discussion page is opened.

Several user interface changes were made to improve system usability in both modules. In the Peerforum module, each post box had extra white

space removed and whenever a peer grading activity is considered finished, the message "The peer grading activity is now finished." is displayed in the post footer for both students and professor.

In the PeerBlock, the persistent block displayed on the top corner of the page now includes some information related to the peer grading activities taking place. In the PeerBlock itself, the existing tabs diving all content accessible now have shorter names to avoid overlapping, to give the interface a cleaner look.

The visualization of graded posts and ungraded posts by each student were in different tabs even though they shared several contents, therefore, the two sections were merged. The row "Feedback" now only displays an exert of its content in a link format, the remaining feedback is accessible by clicking on it to see it in full in another browser tab.

The graders' statistics tab now displays all information in the same page and the different rows can ordered.

Lastly, the "Manage Posts" tab was changed to display the information about posts that have not expired only, when it's loaded initially and filtering options were corrected to work even when changing page.

4.3. Peerforum New Features

To solve the research problem, the PeerBlock can now be configured to include a form where students nominate their favorite peers and least favorite peers. After peer grading a certain student, students can peer rank them too. The answers submitted by the students in either peer classification systems can be seen, and posteriorly edited if needed, in the professor's version of the Peer Block, in the tab "Students Relationships".

Training Pages. It was included the possibility for students to consult training pages related to the subject of post they had been given to peer grade, either through a link above the post they have to peer grade or in the discussions' page. These peer grade training pages explained the objective of the work, its requirements and examples of previous submissions, the rate given and some explanation of why it received a certain grade. In the professor version of PeerBlock, a "Training" tab is available in which all training pages can be accessed and configured.

Hidden Teachers' Reply Posts. To avoid students being influenced by the professor output on the posts, the professor rate and feedback could be configured to be temporarily hidden. This way, the professor can rate and provide feedback to posts whenever it decides to so and avoids accumulating submissions to grade because it had to wait for students to peer grade first.

Automatic Assignment of Improvement Posts. This new version of the plugin also allows for students replies to their initial reply to have automatically assigned the same set of peer graders from the first reply, as these cases are an improved version of the initial submission.

Advanced Topic Attribution Algorithm. The peer grading attribution algorithm also introduced the possibility to divide students throughout the different subjects available. This distribution can be done either by randomly dividing all students equally throughout all subjects. Another possibility is offered. For each available subject, the professor can decide whether the subject should be considered a "Specified" topic, that is, one that would require to designate a given number of students and only they can peer grade this topic submissions or a "Randomized" topic, one that all students left could be potential peer graders. By default, all created discussion topics are considered "Randomized" topics and any students can have assigned all posts.

Email and Internal Notification of Peer Grading Assignments. For a student to be aware that they have peer grades assigned, all students assigned to peer grade any given post now receive a message in Moodle from the professor, with the message "You have a new post to peer grade".

Data Exportation. Data related to all peer grading activity and collection of social relationships was made exportable through links appearing in the various tabs of the PeerBlock, which would generate a .csv file containing the information related to the PeerBlock tab it was displaying.

5. Results

The newer version of the plugin was developed for gathering information to study if social relationships between students influence students' ability to provide reliable peer grading. To solve this research problem, the plugin was included in a gamified blended learning environment, specifically the Multimedia Content Production (MCP) course, a blended learning course where students could assess each other's submissions, along with the course professors and share their social preferences. The online component was provided through the Moodle platform, which includes a system that is able to provide gamification features.

The section where the students submit work to later be evaluated is called the Skill Tree. This component is done through the Forums during the semester. The evaluation process is usually done by the professor, who has the responsibility of going through all new submissions to rate them and give feedback. Students can resubmit their work to improve the initial grade.

5.1. Experimental Protocol

The first version of the new plug-in was used during the final nine weeks (out of 13) of the second semester of the academic year of 2018/2019, in the Multimedia Content Production course. Students were given an extra XP if they participated in the peer grading activity as an incentive. All grades provided by the students were not accounted for the students' final grades in their posts.

5.2. Peerforum Plugin

The peer grading plug-in was introduced at the end of March. At this time, students had already been submitting their work assignments in the traditional MCP course forum, therefore the PeerForum created only included subjects corresponding to level three and four of the Skill Tree.

Students had five days to peer grade the post they received, and all peer grades given were anonymous. During this time professors could also rate and give their feedback on submissions. To avoid students (and professor) being influenced by the other responses submitted already, all rates and feedback given were temporary hidden until the peer grading activity was considered ended.

5.3. Research Questions

Students who had nominated peers and during the course performed peer grading in the same people were checked for biased grading. Peer rankings were also observed. These results were used to answer the following research questions:

1. Did the peers who nominated or ranked other peers favorably showed some sign of bias when evaluating their works?
2. Did the peers who nominated or ranked unfavorably showed some sign of bias when evaluating their works?
3. Is peer grading a valid and reliable evaluation method in a higher education gamified blended learning environment?

5.4. Data Overview

From the 112 students enrolled in this course, 97 of these participated at least once in the peer grading activity (86,6%). Throughout the seven discussion topics created on the PeerForum, a total of 180 posts were submitted, of which 96 were graded (53,3%). From these 96 graded posts, resulted 235 peer grades. Students who participated were assigned five posts on average and graded 68,1% of their assigned posts.

To be able to peer grade their colleagues, students had to first answer to a questionnaire where they would nominate the four peers they liked most and liked least. This questionnaire had a total of

101 answers and provided a total of 846 nominations. With this information, it was possible to apply the CDC system (Terry and Coie, 1991), which only considers peer nominations and organizes all students into the different sociometric status groups described in Chapter 3. Students were split into different five sociometric groups: popular, rejected, neglected, controversial and average. Students who did not fit any of these five groups' criteria were left unclassified - other. The results obtained are displayed below:

Popular	12
Controversial	12
Rejected	3
Neglected	19
Average	25
Other	43

Table 1: Sociometric status results of all nominated students

After grading an assigned post, students had the option to peer rank their colleague(s) from one (dislike a lot) to five (like a lot) and an extra option, zero, was provided in case students did not know who the peer they graded was. In the peer ranking activity, 80 students ranked at least one student, providing a total of 180 rankings to 50 different students, with some of these rankings coinciding with students they had previously nominated too.

5.5. Results

All gathered data regarding peer grading will be studied, firstly observing how students performed compared to each other, followed then by the observation of student peer grades correlation and agreement to the ideal case, the professor, and correlating the previous results with the sociometric information.

5.5.1. Peer Grading Between Students.

To measure the agreement among student graders assigned to grade the same submission, the concordance, also known as inter-reliability, was calculated. Using the statistical model Intraclass Correlation Coefficient case 1 [ICC(1)], the inter-rater agreement between the students randomly selected to rate the same post result of 0,567 suggests a significant agreement among students when peer grading the same post. The result from the concordance of each student individual peer grade to the final averaged peer grade score obtained was of 0,691, which shows that there is a strong agreement between the peer grades given and the average final result, meaning the average final result does not result from very differentiated peer grades.

For this study, the minimum number of students required to peer grade a post to consider its peer grading activity ended, and display the professor

rate and students peer grade was of three. To understand how the number of students that peer graded the posts affects the mean score, and what should be minimum number of students peer grading posts to achieve significant agreement between them, the ICC was also used to compare peer grade scores from posts with different number of peer grades submitted. The results showed moderate agreement (0,391) when four or more students peer grade a post.

The plug-in also included training pages for each skill to help students peer grading. These students, had an agreement with the professor in 51% of time. In the remaining 49%, in most of the times the disagreement of grade between student and professor came from the fact that students did not verify if the submission contained all of the additional required contents, as the grades ended up matching the professor after these submissions corrections were made.

Finally, there were seven instances in which there was two peer graders which had checked the training page corresponding to the topic which they had to peer grade in the same post. In these cases, in less than half the cases (three), there was consensus between the two. It is worth mentioning that it is not possible to know whether these students actually read carefully these pages or just opened those and skimmed through them.

To understand if checking the training pages may have had any impact in the student's submissions, as some students checked them but did not peer grade, the scores received between students who did check the training pages before submitting work in the skill tree with the students who did not were compared. The results indicate that students who checked training display an increase in the average of results obtained (3,657), from the students who did not check the training pages (3,591).

5.5.2. Peer grading Between Students and Professor

Through inter-rater agreement, the results showed that, from the 96 posts graded by both professor and student(s), 55 peer grades matched the professor rates (57%), thus there was a medium agreement between professor and students. The cases that did not match are equally divided by lower (15) and higher (16) grades than the professor, though most cases the difference was of only one point. There were six exceptional cases where the difference was of two points. In all these cases, the students give a wrong scoring to the posts and did not verify if all the requirements for a submission were met before peer grading.

When comparing each individual grade given by a student to the professor grade, the percentage of peer grades that match the professor decreases to

53%, a good indicator that having more than one student peer grade the post can lead to a more accurate peer grade. Usually, the grades given by each student to the assignments tend to converge with the professor the more they participate; as the student gains more experience from peer grading. Comparing the overall mean scores given by professors and students, there seems to be a similar value between the two, 3,448 and 3,595 respectively, though students seem to be more conservative in their results, mostly rating work three or four, and give slightly higher scores than the professor.

Finally, to understand the validity of peer grading, that is, assuming the professor grade is the right grade, it was studied the similarity between the professor rate and peer grading scores through the Pearson product-moment correlation coefficient (r). When applying the statistical model, the result of 0,237 was achieved. The small value confirms that the students and professor tend to grade somewhat different scores, however this result is highly affected by the fact that the scale used for peer grading is considerably small (1-5), making the disagreement of one value of difference impact highly the validity results.

Looking at students' overall participation in the skill tree activity, students with the highest number of skills achieved in the skill tree usually give the same grade as the professor. Students who few submissions have more inconsistent peer grades. The correlation value obtained between students' participation in the skill tree and matching of peer grades given was of 0,515, corresponding to moderate strength, indicating that participating in the skill tree and being familiar with the grading process helps achieving better peer grades.

When looking at participation in the levels three and four of the skill tree only, the more skills students achieved, the higher match in scores between the student peer grades and professor. The correlation between students peer grades and professor rates was of 0,781, confirming that higher participation in the skill tree helps peer grade validity.

5.5.3. Peer grading and Sociometric Status Groups

Looking at the data in detail, it was found that 20 peer grades were given to students by peer graders which also had been nominated by them. These 20 peer graded students had previously received 10 positive nominations from their grader and other 10 had received a negative nomination. In 13 of those cases, the grades given by the students matched the professor grade. The remaining seven cases, all of them were by a point difference. Only in two situations, the discrepancy in professor

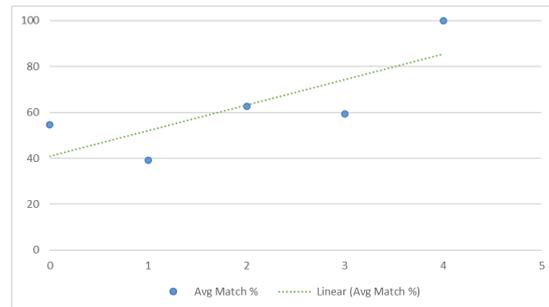


Figure 1: Relation between the number of skills of level 3 and 4 achieved and percentage of agreement between student and professors in grading

and student scores came from instances in which a negative nomination was given beforehand.

The other five cases came from students which had given a positive nomination to their colleague and in three of them, the students gave a higher grade than the professor. This indicates that it does not look like social relationships have an impact in student's capacity to peer grade a colleague fairly.

Using the ICC method again, it was calculated the coefficient of concordance between professors and students who peer graded posts of students which they peer nominated previously. The results show moderate concordance between professor and students, 0,617, especially when students had given a negative nomination (0,802). The opposite case happened in cases where students had given a positive nomination beforehand (-0,053). While this results may not convey it, there was actually some agreement between professor and student when looking at positive nominations, with almost half of the cases observing actually having matching scores.

In terms of peer rankings given, from the 180 answers provided by the students, only 86 (47,8%) were rankings from one to five, the 94 remaining being zero rankings and therefore not many conclusions can be gathered from them. In the rankings from one to five, observing the results, it was found almost identical scores between professor and students, and in the few cases where there was not, the student usually gave a higher score than the teacher, even if after ranked the student lowly.

The Controversial students were found to be the sociometric group which the peer grades agrees the least with the professor rate, by either deflating or inflating at least one grade, and 32% of them did both. Rejected students only have cases of grade inflation compared to the professor and most students had multiple cases. Neglected students were the best group, however, its group population is too small to make any conclusions.

To understand better these results, a more in

depth observation of data is made, firstly by observing how each group of students acted when peer grading their colleagues and then observing if there any sociometric group that tends to receive biased peer grades.

To examine if students belonging to a certain sociometric status were displaying some bias towards students from a certain sociometric group when grading (whether these group of students was its own or another), students who did not peer grade where removed from the initial sociometric results.

Each group and their respective students were analyzed in terms of peer grades provided. All groups, with the exception of the neglected group, which only had two students, had similar matching performances between professor and students, ranging from 49,05% (controversial students) to the highest, 54,31% (popular students). Using the ICC coefficient to compare how agreement between professor and students from the different groups changed, the results found are displayed in Table 5.8.

Popular	0,32
Controversial	0,246
Rejected	0,311
Neglected	1
Average	0,287
Other	0,342

Table 2: Comparison of ICC coefficient between professor rates and students peer grades

The results display low difference between the groups that the students are inserted into and it does not seem that their sociometric status affects their reliability of peer grading.

To see if the placement in these groups could have an influence in the grading that the 55 students which presented work received, it was again observed how the students in each groups had received peer grades that matched the professor and when it did not, if any of these groups had a tendency to have more inflated or deflated grades.

When looking at each group, the students' scores received show no indication of bias towards students of one certain group, with each group receiving a similar proportion of peer grades from two to four. The sociometric group which had the best peer grades, including several scores of five were the Rejected and the groups of Popular, Controversial and Average received identical amount of peer grades scored two and similar number of the remaining scores.

5.6. Discussion

The results presented after studying and analyzing the data regarding peer grading, reveal that students can produce moderately consistent peer

grades, especially in groups of four or more students. The scores of students and professor only differ significantly due to wrong grading from the students, who did not checked for the presence of all requirements. In most cases, the difference is only of one point, therefore, while professor and student concordance might have been low, students do seem to agree moderately with the professor. Students also have a significant tendency to agree more with each other than with the professor. The students who checked training seemed to agree more with the professor, a good indicator that more training, can lead to better peer grading results.

Looking at the different sociometric status groups formed through the CDC system, students who peer graded peers which they had nominated before, had some disagreement with the professor, though of one-point difference only, with the professor usually giving the higher grade. When a negative nomination was given before, there was almost complete agreement between students and professor regarding the scores given, so it is possible to conclude that there is not sign of significant bias in these situations.

It was possible to notice that students belonging to the Controversial group had the most frequent situations of disagreement with the professor, in most cases assigning a lower grade than the professor. Rejected students displayed the exact opposite behavior, being the worst performers in the course and usually inflating grades. However, when submitting work, the Rejected students were the one with the highest peer grades received and the remaining groups received mostly similar number of peer grade scores form across the scale. No sociometric group displayed significant bias when peer grading nor suffered significant bias when being peer graded.

6. Conclusions

Peer grading in blended learning environments has shown to be a reliable form of grading and helps solving scalability problems. However, the use of students which have developed social connections between each other, may result in biased peer grades, which could lead to unfairness in students' evaluations. To investigate this, a plugin was developed for Moodle, bringing peer grading capacities to the platform, together with features for retrieving information regarding students' connections. The plugin was deployed to the Multimedia Content Production course and the data collected revealed that students tend to have high agreement rates with each other, especially when at least three students peer grade a post, however, there was some disagreement between professor

and student, though of one-point difference in a five-point scale, and in most cases, both professor and student agreed that the work presented was good enough to be accepted. Students with training, appear to be more in accordance with the professor.

Looking at the results of peer nominations, no significant bias was found when observing the score given by students who peer graded colleagues they had nominated in advance, nor when they peer rated afterwards. When looking at peer grade results by dividing students in different sociometric status groups, it was also found that the agreement levels of peer grades and professor rates was similar throughout all groups, meaning no group seems to provide more biased grades. Finally, when observing the peer grades received by students from the different sociometric groups it was also not found any group suffering from biased peer grading, therefore concluding that the sociometric status of students or their previous social connections do not lead to biased grading, and students can be considered a reliable source of grading, especially if more practice to the task is done beforehand.

6.1. Future Work

Although the Peerforum plugin is now in a stable form, there are some aspects that can be improved regarding plugin performance as the system is still considerably slow loading the pages. The results showed that training does help students becomes better peer graders and achieve better scores in their submissions, therefore training pages should become mandatory in the beginning of the course and include some exercises to make sure that there some effort made when students consult those pages. In terms of the study itself, having the PeerForum plugin working since the beginning of the semester, can lead to more students participating and more chances for students to improve their peer grading capacities and increase their level of concordance with the professor, someone who has extensive training in the same type of work. The use of a increased scale can also help to decrease the difference in results of peer grading versus the professor rating.

References

- [1] B. J. Arnold. Gamification in education. *AS-BBS Proceedings*, 21(1):32, 2014.
- [2] S.-W. Chou and C.-H. Liu. Learning effectiveness in a web-based virtual learning environment: a learner control perspective. *Journal of computer assisted learning*, 21(1):65–76, 2005.
- [3] J. D. Coie, K. A. Dodge, and H. Coppotelli. Dimensions and types of social status: A cross-age perspective. *Developmental psychology*, 18(4):557, 1982.
- [4] J. Cole and H. Foster. *Using Moodle: Teaching with the popular open source course management system.* O'Reilly Media, Inc., 2007.
- [5] D. R. Garrison and N. D. Vaughan. *Blended learning in higher education: Framework, principles, and guidelines.* John Wiley & Sons, 2008.
- [6] E. F. Gehringer. Electronic peer review and peer grading in computer-science courses. *ACM SIGCSE Bulletin*, 33(1):139–143, 2001.
- [7] S. D. Gest, S. A. Graham-Bermann, and W. W. Hartup. Peer experience: Common and unique features of number of friendships, social network centrality, and sociometric status. *Social development*, 10(1):23–40, 2001.
- [8] E. A. O'Connor and A. McQuigge. Exploring badging for peer review, extended learning and evaluation, and reflective/critical feedback within an online graduate course. *Journal of Educational Technology Systems*, 42(2):87–105, 2013.
- [9] J. Ribeiro. Peer grading in gamified blended learning environments, 2016.
- [10] M. Rostampoor-Vajari. What is sociometry and how we can apply it in our life. *Advances in Asian Social Science*, 2(4):570–573, 2012.
- [11] S. Taylor, M. Ryan, and J. Pearce. Enhanced student learning in accounting utilising web-based technology, peer-review feedback and reflective practices: a learning community approach to assessment. *Higher Education Research & Development*, 34(6):1251–1269, 2015.
- [12] R. Terry and J. D. Coie. A comparison of methods for defining sociometric status among children. *Developmental Psychology*, 27(5):867, 1991.
- [13] K. J. Topping. Peer assessment. *Theory into practice*, 48(1):20–27, 2009.
- [14] S.-C. Tseng and C.-C. Tsai. On-line peer assessment and the role of the peer feedback: A study of high school computer course. *Computers & Education*, 49(4):1161–1174, 2007.