

A gamified peer assessment model for on-line learning environments: An experiment with MeuTutor

Thyago Tenório¹, Ig Ibert Bittencourt²

¹Universidade Federal de Alagoas (UFAL) – Campus Arapiraca/Pólo Penedo
Av. Beira Rio – 57.200-000 – Penedo – AL – Brasil

²Instituto de Computação (IC) – Universidade Federal de Alagoas (UFAL)
Av. Lourival Melo Mota, s/n – 57.072-900 – Maceió – AL – Brasil

thyago.oliveira@penedo.ufal.br, ig.ibert@ic.ufal.br

Abstract. *Though the on-line learning environments provide scalable and automatic ways to present a content, forums and evaluate the progress of the students, they are still limited in their ability to evaluate complex activities such as correction of essays. Facing this issue, Peer Assessment (PA) offers a powerful solution. However, a major problem found when using PA is the presence of inadequate behaviours from students, which affect the learning and evaluation system. In this sense, this paper proposes a gamified peer assessment model, where the gamification elements improvement the motivational aspect of the students. In the experiments conducted, it was concluded that there were no statistical variations between the grades obtained compared to the experts, the correction time was reduced and the cost involved decreased 72.4%. The gamification encouraged students to use the platform: increased access (+64.28%), essays performed (+10.53%) and essays corrected (+20%).*

1. Introduction

The recent increase in the popularity of online courses became more accessible for anyone, with an internet connection, to enroll freely in university level courses [Piech et al. 2013]. However, while new web technologies allow scalable ways to deliver video lecture content, to implement social forums and to track student progress, we remain limited in our ability to evaluate and give feedback for complex and often open-ended student evaluations, such as mathematical proofs, design problems and essays [Piech et al. 2013]. This happens because this kind of activity requires manual and individually correction by teacher, which leads to an excessive workload. The further students in the environment, more activities will be done and the correction by teachers would quickly become infeasible.

Facing this issue, Peer Assessment (PA) offers a promising solution to scale the grading of complex activities in courses with tens or even hundreds of thousands of students [Piech et al. 2013]. PA stimulates students to reflect, to discuss and to collaborate in their learning process [Topping 1998]. Peer Assessment is a process by which students or their peers attach grades based on predefined benchmarks by professor [Sadler and Good 2006]. PA is recommended because reduces the teacher's workload [Rubin and Turner 2012], increases learning outcome [Murakami et al. 2012].

The effectiveness and quality of an assessment depend of their incorporation into of the learning process [Schuwirth 2004]. Peer Assessment is a collaborative evaluation

process, where the quality of the final results depends of good evaluations of their reviewers. However, a major problem found when using Peer Assessment is the presence of inadequate behaviours from students, which affect the learning and assessment system. Often, the students are not motivated to collaborate with the reviews of the activities. The emergence of this type of behaviour occurs by various cognitive and emotional factors, such as boredom, lack of motivation and the need to get results quickly [Kapp 2012]. To resolve this issue, it is necessary to use techniques that have the ability to “engage” positively the emotional/cognitive state of the student.

The paper proposes a gamified peer assessment model, where the gamification elements improvement the motivational aspect of the students. Gamification is the use of mechanical, ideas and aesthetics games (context, fast feedback, competition, stages, achievements, points, among others), to engage people, to motivate actions, to promote learning and to solve problems [Kapp 2012]. This term is commonly used to express the use of game elements (storyline, score, levels, quests, badges and rankings) in environments that are not games (educational environment, for example) to motivate or influence people to perform a certain activity. The combined use of peer assessment techniques with gamification become the process most powerful and complete, avoiding and/or decreasing the presence of inappropriate behaviour by students.

2. Peer Assessment Model

This section aims to describe the proposed model. The Figure 1 shows the model’s steps. The first step is the creation of the discursive activities and definition of the evaluation form containing the criteria on which the activities will be evaluated. The second step starts when the activity is released to student. They will answer it and will submit their responses to the system.

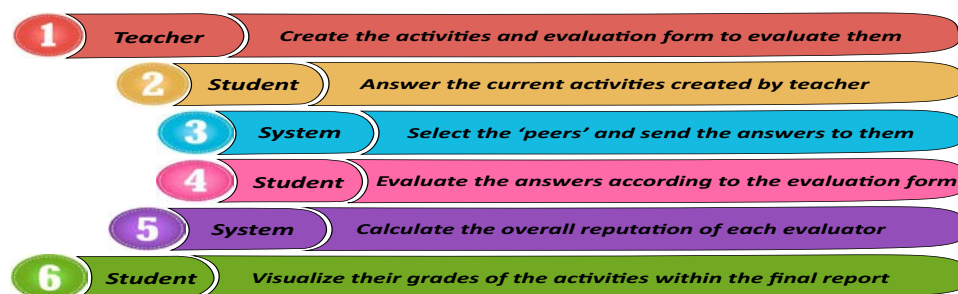


Figure 1. Model in details

In the third step, the system sends the submitted activities to others students correct them. Then, in the fourth step, the students, according to the distribution in the previous step, will evaluate the responses of their peers based on pre-defined evaluation form by the teacher (step 1), in other words, in this step the students will evaluate their peers and will send the grades to the system. The deadline for the correction of the activities in the system is called evaluation period. To assign his grade, the student will receive the responses of a particular student and the evaluation form with pre-defined criteria. He must analyse the activity from the point of view of the criteria and then simply fill out the form with their grades.

In order to have greater confidence in the final grade, the fifth step evaluates the overall reputation and the level of competence of each evaluator for each criterion. In addition to the general reputation in the system, it is necessary to calculate the reliability in the grades of the users of the system. So, the system calculates the reliability of each assessment, the competence's level of evaluator in each criteria and their reliability and can then calculate the final result as a weighted average. Finally, the sixth step consists in generate a report by system with the grades of each criteria as well as the final result of the activity, which to be presented to the student. It is possible include comments and there is possible to changes on the grades by teacher before their presentation to the student.

The details of each step of the model can be found in Chapter 4 (section 4.1) of the dissertation. This entire process is kept in functioning through the use of gamification elements, which aims to increase the motivation of users. The main objective is to ensure students feel motivated to participate on both performing activities and correcting them. In this sense, we defined the following target behaviours: To encourage students to participate more often in the activities; To challenge students to get badges; To provide medals to students who participated of the activities; To provide badges or points to students that answered/corrected activities; To promote a ranking parametrized for students; To encourage competition with rewards to the winners.

3. Controlled Experiments and Results

We applied our Peer Assessment model in a on-line learning environment called MeuTutor®. The educational environment MeuTutor is an intelligent tutoring system, which aims to monitor the learning of students in a personalized way, ensuring quality in teaching and improving the performance. The chosen version was the MeuTutor-ENEM, which aims to help high school students to prepare for the National High School Exam (ENEM). Thus, the environment offers courses related to high school subjects like Portuguese, mathematics, among others. Our goal was evaluate the effectiveness of to use our proposed model in the context of correction of essays.

We analysed two fundamental aspects: 1- Effectiveness of using Peer Assessment - It involves questions like **“How can we evaluate the quality of the corrections made by the students?”** and especially **“Can these assessments be compared with expert’s assessments?”**. We also inquired about **“Does the model applied in this context really reducing the cost and overwork of the teacher? ”**; 2- Gamification impact - It involves questions like **“What is the influence of gamification in the proposed model?”** and **“Does the gamification really influence and motivate students within the peer assessment process in this competitive environment?”**

Thus, two experiments were performed at separate. The first evaluates the effectiveness of the peer assessment model and the second evaluates the influence of gamification into model. For the first, we selected about 30 students who had some essays done in the MeuTutor environment. The correction process was done using the proposed model and the traditional way (where the experts corrected the essays). The variables analysed in experiment planning were: grade (N) (from 0 to 1000); Time for correction (T) (in minutes); Cost per student per correction (C) (em R\$).

The data of these metrics are summarized in Table 1. We considered three aspects: Application of the traditional model (correction by expert) within the environment Meu-

Tutor (T1); Application of the proposed peer assessment model within the environment MeuTutor (T2); traditional model without the use of educational environment (T3).

Table 1. Summarization of the data of experiment 1

Treatment	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Metric grade (N)						
T1	80.0	490.0	660.0	615.6	750.0	880.0
T2	380.0	520.0	573.0	577.7	624.5	960.0
T3	400.0	460.0	540.0	536.0	590.0	680.0
T1 - T2	-400.00	-79.75	40.00	37.89	110.00	480.00
Metric Time (T)						
T1	153.0	382.8	600.0	679.3	889.5	1833.0
T2	240.0	337.5	472.5	501.7	588.8	900.0
T3	420.0	555.0	660.0	669.1	780.0	917.0
Metric Coust (C)						
T1	1.36	3.40	5.33	6.03	7.90	16.29
T2	1.68	1.68	1.68	1.68	1.68	1.68
T3	3.03	4.01	4.77	4.83	5.63	6.62

Through statistical tests presented in Chapter 5 (more specifically in Section 5.3.3), we had the following results: 1- There were no statistically significant variations between the grades obtained in both the proposed model and the traditional model. With this, it is possible to use the proposed model to replace the traditional model in the environment and get their benefits without compromising the final results of the evaluations; 2- The correction time using the proposed model showed faster than expert correction time; 3- There were enough statistical evidence to indicate that the cost involved of the proposed model showed lesser than the cost of the traditional model. Through of the creation of a regression model, it shows that the cost reduction is about 72.4%.

The second experiment aims to evaluate the motivational impact of the inclusion of gamification in the proposed model in this work. So, our research question was **“Does the use of gamification elements in the peer assessment model bring significant differences on the motivational aspect of the students involved?”**. The metrics used to evaluate the model of this experiment were: Amount of access (A), number of essays performed (RF) and amount of corrected essays (RC). The experiment was conducted in a private school with high school classes (about 100 students). Randomly, it was distributed to each student a group (1 or 2) representing the control group (GC) and experimental group (GT). Students in the control group (GC) performed the treatment T1 (without gamification) while students of the experimental group (GT) performed the treatment T2 (with gamification).

The Table 2 briefly present some of the main results obtained in the experiment, in a direct comparison between the model without gamification (T1) and with gamification (T2).

The Table 2 showed that there were two more registers in the group of the treatment T2 (+11.76 %). The number of pendent essays was four, regardless of treatment. This may have been caused by the short deadline for the realization and correction of essays, implying that some students failed to correct enough essays. By have more es-

Table 2. Summarization of data on the variables analyzed

Data observed	T1	T2	Difference	% Dif.
Number of student records	17	19	+2	+11.76%
Total Pending Essays	4	4	0	0%
Average grades	669.33	563.92	-105.41	-15.75%
Number of logins (A)	42	69	+27	+64.28%
Total Essays Made (RF)	19	21	+2	+10.53%
Total Corrected Essays (RC)	45	54	+9	+20.0%

says made and due to this amount be equal, we concluded that the percentage difference between grades is greater in gamification (-15.75%). Moreover, the data showed that, in T2, the amount of access was 64.28% higher, the number of essays performed was 10.53% higher and 20% more essays corrected if compared to T1, which indicates the strong influence of gamification on the platform.

4. Discussion

The paper proposes a gamified peer assessment model that uses gamification elements as a motivational aspect for students inside Peer Assessment process. Two experiments in MeuTutor educational environment showed satisfactory results (the grades of both models are equivalent, lower time and a cost about 72% lesser on proposed model). Moreover, gamification influenced positively on the overall context of the peer assessment model (64.28% more access, 10.53% more essays performed and about 20% more essays corrected).

To analyse these results, we will simulate with a real scenario of application. Suppose that a single school with about 100 students resolves to promote the creation of essays in their on-line environment. Suppose that going to be performed one essay per week per student, which leads to four essays monthly per student. Suppose an annual period of 10 months of work in this project. So, at the end we will have a total of 4000 essays performed (100 students x 4 essays/month x 10 months).

Considering the number of essays and the average time obtained in our experiments, we have that an expert would take about 754 hours (4000 x 11.31min) to correct all essays in total. On the other hand, the simple use of the model without gamification would decrease this time to an average of 8.36min for essay, which would give us a total of 557 hours (4000 x 8.36min) spent with corrections (a gain of 197 hours, or 26.12%). The inclusion of gamification in the model would reduce the correction time for 6.68min. So, with the application of the complete model, the time total to correct the essays would be 445 hours (4000 x 6.68min), a reduction of 309 hours or 40.98% compared to the traditional model or reduction of 112 hours or 20.10% compared to the model without gamification.

Considering the same number of essays and analyzing the average of the costs involved in the experiments, we have that on average, an essay correction by a specialist would cost about R\$6.04. Thus, the total to correct all essays would be about R\$ 24.160 (4000 x 6.04). On the other hand, the simple use of the model without gamification decrease this cost to an average of R\$1.68 for essay, which would give us a total of R\$

6.720 (4000 x 1.68) (about R\$ 17.440 lesser or 72.18%). So, with the application of the complete model, the cost total to correct the essays would be R\$ 5.360 (4000 x 1.34), a reduction of R\$ 18.800 or 77,81% compared to the traditional model or reduction of R\$ 1.360 or 20.23% compared to the model without gamification. Now, imagine the gains if increase the number of schools, which means more students or increase the amount of essays performed by student.

5. Publications

The work presented here was reported in some publications in periodic and conferences directly and/or indirectly. Some publications were: 1- A gamified peer assessment model for on-line learning environments in a competitive context (accepted for publication in *Computers in Human Behavior - Qualis A2 - Impact Factor=2.694*); 2- "Avaliação por pares como ferramenta colaborativa na correção de redações: Um experimento com o ambiente educacional MeuTutor" (accepted for publication in *CSBC SBSC 2016 - Qualis B4*); 3- A gamified peer assessment model for on-line learning environments: An experiment with MeuTutor (*CSBC CTD 2016*); 4- Does Peer Assessment in On-line Learning Environments work? A systematic review of the literature (in analysis on *Computers in Human Behavior*); 5- Ontology-based Feature Modeling: An Empirical Study in Changing Scenarios (published on *Expert systems with application - Qualis A2 - Impact Factor=2.240*); 6- Java ontology integrated toolkit (same number 5); 7- A Semantic Tool to Assist Authors in the Instantiation of Software Product Lines for Intelligent Tutoring Systems Context (published on *IEEE Multidisciplinary Engineering Education Magazine*); Among other papers indirectly.

References

- Boud, D., Cohen, R., and Sampson, J. (1999). Peer learning and assessment. *Assessment & Evaluation in Higher Education*, 24(4):413–426.
- Kapp, K. M. (2012). *The gamification of learning and instruction: game-based methods and strategies for training and education*. John Wiley & Sons.
- Murakami, C., Valvona, C., and Broudy, D. (2012). Turning apathy into activeness in oral communication classes: Regular self-and peer-assessment in a tblt programme. *System*, 40(3):407–420.
- Piech, C., Huang, J., Chen, Z., Do, C., Ng, A., and Koller, D. (2013). Tuned models of peer assessment in moocs. *arXiv preprint arXiv:1307.2579*.
- Rubin, R. F. and Turner, T. (2012). Student performance on and attitudes toward peer assessments on advanced pharmacy practice experience assignments. *Currents in Pharmacy Teaching and Learning*, 4(2):113–121.
- Sadler, P. M. and Good, E. (2006). The impact of self-and peer-grading on student learning. *Educational assessment*, 11(1):1–31.
- Schuwirth, L. (2004). Optimising new modes of assessment: in search of qualities and standards. *Tijdschrift voor Medisch Onderwijs*, 5(23):250–251.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of educational Research*, 68(3):249–276.