A Teaching Proposal for the Software Measurement Process using Gamification: A Experimental Study

Lennon Sales Furtado
Federal Institstute of Pará (IFPA) and Graduate Program in Computer Science (PPGCC) - Federal University of Pará (UFPA)
Paragominas and Belém, Pará, Brazil
lennonsfurtado@gmail.com

Sandro Ronaldo Bezerra Oliveira
Graduate Program in Computer Science (PPGCC) - Federal University of Pará (UFPA)
Belém, Pará, Brazil
srbo@ufpa.br

***Abstract*—This paper presents a proposal to teach the software measurement process from a class of gamification that makes use of gamification and serious game. This class aims to encourage interaction in the classroom and thereby foster interest in the process. In addition, the proposal was applied in two classes with a total of 22 students from the Federal University of Pará (UFPA) in Brazil. As a result, one of the experiments yielded a level of 80% of the evaluations with positive criteria for the class using the gamification. In conclusion, the proposal helped in the process of teaching software measurement from the creation of a competitive and collaborative environment with the core in classroom interactions.**

Keywords-software measurement; teaching methodology; computer course; gamification.

I. INTRODUCTION

The software measurement process has the objective of collecting, storing, analyzing and reporting the data related to the products developed, as well as the processes implemented in the specific organization, in order to support the organizational objectives of the same one [1]. This process is highly relevant in the Software Process Improvement Program (SPIP); however, the software industry has been hesitant to apply efficient measurement programs [2][3]. This is due to the fact that many software managers and professionals, including academics in software engineering and computer science, are unaware of this topic [4].

This assumption is directly related to the measurement process being generally considered difficult domain and time consuming [5][6]. The first hypothesis for the understanding of this problem lies in the teaching of this subject [7], in that it is little explored in the curriculum of undergraduate students, thus relegated to the background, and little incentivated among the students in the learning of this practice. Another implication is the absence of guidelines for the practice of measurement [6][8].

In general, every measurement program has as a determinant for the success of the human factor, since it is not properly motivated and committed to the measurement program, it is unlikely that such a program will achieve the desired results, and therefore the control of software metrics to assist in decision making. Among the alternatives used to maintain the motivation and commitment of the people involved in a SPIP, one of them is the use of the gamification concept [9].

The definition of Gamification consists of the use of game elements and game design techniques outside the context of games [10]. This approach seeks to improve the engagement, motivation and performance of a user in the execution or learning of some task or subject by incorporating mechanics and game elements, making the task or subject more attractive [11].

Therefore, this work aims to present a gamified approach to teaching the subject of software measurement. For more, two experiments were carried out with a total of 22 students from the Graduate Program in Computer Science at Federal University of Pará, from Brazil, in the Special Topics in Software Engineering: Software Processes Technology. The first experiment was carried out in the second half of 2017 and the second experiment was conducted in the second half of 2018. As a way of evaluating the approach, a questionnaire was used where the students applied different criteria to evaluate the modules of the proposal, which used gamification and serious games as tools to promote discussion in the classroom to solidify the student's theoretical knowledge.

II. THE GAMIFICATION

In this section, the Game Design Document is displayed.

A. Overview

The components presented by Werbach and Hunter [10] framework that guided the making of the gamification proposal for this approach are:

- Define Business Objective: to create a competitive and collaborative environment that fosters communication in the classroom;
- Outline Target behaviors: arrive early, make the class or extra classes activities, attend classes, provide suggestions for the experiment, participate in the classroom and ask to keep the topic of interest in discussion;
- Describe the players: students of the subject of Special Topics in Software Engineering: Software Process Technology;
- Design Activities Cycles: engagement loop, from the beginning and end of each class the ranking is presented to encourage the competition;
- Fun: the fun is exclusively addressed in the competition between the players;
- Appropriate Tools: the use of a spreadsheet to record and track players' progress.

B. The Rules

The students were motivated to interact from the point's bonus. These points were awarded in different dimensions of participation, as can be seen in Table I. And the players (students) by the behaviors presented and explained in Table II.

TABLE I. BONUS MECHANICS

Dimension Name	Meaning	Points for Interaction
Attend Classrooms	0 - Missing; 1- Presence	1
Initiative	The first one to answer a question, or other situation in room	2
Suggestion	Any suggestion that points to an improvement in the experiment together with a solution	3
Participation	Any responses after the first respond, or other type of minor contribution.	1
Question	Ask and answer questions in class	2
Arrive early	Arrive in class before scheduled start of class	1
Class or extraclass activities	0 - Did not participate; 1 - Participated, but did not get 100% correct answers; 2- Participated and answered all questions correctly.	Up to 2 points per activity

TABLE II. PENALTY MECHANICS

Dimension Name	Meaning	Points for Interaction
Miss class without communicating	Missed class without communicating instructor.	-6
Missed class with warning	Missed class with previously communicated the instructor	-4
10 minutes of delay	-1 for every 10 minutes of delay, being able to arrive at the maximum of 40 minutes of delay, i.e., - 4 points per class.	Up to -4
Penalties	Any inconvenient activity during class time, such as using the cell phone, computer, tablet, etc.	-2

C. Improvements to the First Experiment

Some points of improvements from the first experiment to the second were identified, listed and applied, were the following: the inclusion of metrics used in agile methods; less complexity of the practical tasks; extra-class exercise list on metrics paradigm and inclusion of achievement mechanics by giving medals to students who reach certain milestones.

III. THE EXPERIMENT

A class monitor participated in the experiment as a judge, who is responsible for taking notes and punctuating students from the observation of their behavior in the classroom. In addition, 22 students participated in these experiments, 15 in the first experiment and 7 in the second experiment, from graduate program in computer science at Federal University of Pará in the subject of Special Topics in Software Engineering: Software Process Technology.

A. Pre-Questionnaire

In order to identify the class profile, a questionnaire based on the work [13] was applied. Such a questionnaire was not compulsory and its completeness added points in gamification. Thus, only 11 students from the first experiment and 6 from the second answered this questionnaire.

In general, the group showed an ignorance of the measurement process and its paradigms, even though 40% of the students entered the market and the other 60% as graduate students in computer science. All students had seen the subject of software measurement in a summarized way (as a topic within another subject) during their undergraduate degree.

B. Theoretical Class

The content of the software measurement subject was divided in 2 (two) days of expository classes, that is, 4 (four) 50-minute classes for the theoretical teaching of Software Measurement. All content was based on Chapters 1 and 2 of the book [14].

C. Practical Activities

On the third day of the class, a software measurement practice aligned with the expected results of the MPS.BR - Brazilian Software Process Improvement model [1] was made. Participating students performed a measurement activity that consisted of developing measures to attest to the physical and mental health of students in the classroom, using as tools a weight scale, a tape measure and blank papers for the development of questionnaires.

While on the fourth day, the serious game X-Med [13] was used as a simulator of the software measurement process. After its use, the game was analyzed in order to clarify the questions presented by it. In order to present the concept of software measurement in a dynamic and not frightening way, the Dojo Randori method [15] was adopted in practical activities, where a pilot and a co-pilot, wave every 7 minutes were replaced by someone from the audience, did the execution of the practice.

D. Practical Evaluation

On the fifth day a practical activity was conducted, where each pair of students planned metrics to improve the quality of the serious game X-Med. For this, the ISO 25010 norm was presented as a way to help students in the process of developing their metrics.

Therefore, on the sixth day the metrics were performed according to the measurement plan made in the previous class. The outputs of these classes were a worksheet with the metrics collected and a report with the analysis of the data collected. Both activities of the fifth and sixth days were done in pairs.

E. Theoretical Evaluation

On the seventh day a traditional evaluative method was applied, that is, a theoretical evaluation with 11 (eleven) questions about software measurement.

F. Post-Questionnaire

After the theoretical evaluation the students were informed that a questionnaire was available for them to answer. This post-questionnaire had as objective to identify the subjective impression of the students on the methodologies chosen for the teaching of software measurement.

G. Feedback Class

On the eighth day, the last day of the experiment, a class was held to discuss the questions posed in the post-questionnaire in order to discover the strengths and points of improvement of the classroom experiment for measuring software teaching.

IV. The Results Obtained

In this section will be presented the data collected by the questionnaires applied, together with the data collected during the feedback class.

A. Qualitative Analysis

Both the questionnaires and the feedback class reported similar results. The results of these approaches are presented below.

At the end of the theoretical evaluation, a questionnaire adapted from the work of Wangenheim, Thiry and Kochanski [13] was applied, with 26 questions for the students to evaluate the experiment. The questions of the questionnaire were divided into 5 major areas, being: the evaluation of theoretical classes; the evaluation of practical classes; the evaluation of practical evaluation; the gamification; and the evaluation of progress itself in relation to the subject of software measurement. Such a questionnaire was optional, and it counted only for gamification. Knowing this, only 10 students from the first experiment and 6 students from the second experiment answered. The most important question of this questionnaire will be presented in this section.

The most relevant for the area of evaluation of the theoretical classes were questions 1 and 2, which are:

- How do you evaluate the content, topic sequence, difficulty level, duration, teaching method, and exercises in theoretical classes?

Taking into account the first experiment, the content of the topics, the sequence of topics, the teaching method and the exercises had a level above 90% acceptance, mainly receiving good and excellent criteria, having only one criterion regular. Already the degree of difficulty, presented results between the regular and good criteria, and the duration of the classes was the only dimension presenting a poor criterion. Even so, the average class length criteria ranged from regular to good criteria.

For the second experiment, the content had a 50% acceptance from the evaluations of the good and excellent levels, the rest of the evaluated items stood out the regular criterion, representing class neutrality before the theoretical class. Only in the last item, exercises, which were evaluated mainly with the bad criterion, were being a total of 50% of the class, 18% of regular and 32% of the good criterion.

While for the evaluation of the practical class, the same questions were asked, and their results will be presented below.

In the first experiment, it can be observed that the good and excellent criteria were applied in a balanced way among the different topics evaluated. With the exception of the topic duration, which as in the theoretical class was the one that had the highest incidence of the criterion Bad applied in its evaluation. This same question for the second experiment had in the great majority the criterion applied, with the exception of duration that had a higher incidence of the bad criterion. That is, even with the improvements applied between the experiments to reduce the difficulty of this activity, it still presents a high difficulty index, where the students indicate that the duration is insufficient to solve this activity.

For the evaluation of gamification, the main question raised was the following:

- How do you evaluate the elements of games, game mechanics, difficulty level, game balancing, and gamified classroom entertainment?

For the first experiment, in relation to the other results, gamification was the only one that had a greater application of the criterion Excellent by the students, as well as had only one case where the criterion Bad was applied and this was applied to the balance of the proposal.

However, in the second experiment, the result was different, following the trend of the other areas of the experiment, which had a high neutral index in most of the evaluated items and the presence of the bad criterion with 66.7% in the entertainment item.

B. Quantitative Analysis

For quantitative analysis, the total score of the class was used in the interactions within the classroom, and the penalties were analyzed to be able to evaluate the behavior of the class before the teaching of software measurement from the use of gamification.

For the first experiment, in the dimension Delay in 10 minutes the class behaved as follows: the first day of school was the one with the highest incidence of this dimension, with 15 incidences. As of this day, there was a reduction in the incidence of undercurrent between no delay in the class 8 and 13 incidences in class 3. For the second experiment, there were a maximum of 9 incidences in class 6 and, unlike the first class; there were 2 students who they were late every day they went to class.

The penalties dimension had its highest incidence in the days of practical class, in the first as in the second experiment, for example the class 5 had 15 incidences of this dimension in the first experiment and in the second it had 2, it was also the highest incidence of this penalty in the second experiment. This can be understood because the class had access to their computers, some students used them to access non-class sites, such as social networks, and some students were using the cell phone after being pilots or co-pilots in the Dojo.

For the presence dimension, it was possible to perceive in the first experiment that on average 13 students, that is, 86% of the class was present in all classes and in the second experiment only 4 (57%) of the 7 that were present. With this, it was perceptive, in both experiments, that the 3 students who were at the top of the ranking did not miss any classes. In addition, the first two places arrived early in all classes in the first experiment and the second in 6

classes, indicating that the gamification serves as an extrinsic motivator to keep those in the lead focused.

The two most important dimensions were: participation and questions. Overall, students made significant progress in participations during the time of the experiment, indeed that the core of the experiment was to encourage mainly these two dimensions. The participation dimension was the one that had the highest occurrence incidence in most classes, with a total of 239 participations and with the average of approximately 30 participations per class for the first experiment. Already for the second, 183 in total and 22.8 on average per class.

In contrast, for the first experiment the question dimension had a total of 48 and the average of 6 incidences per class, and its apex was in class 1, where 22 questions were asked. However, most of the questions were basic and shallow, and in other classes this number was reducing. However, there was a great leap in the quality of questions asked in the classroom, this may be due to the reason students are still getting used to the subject of measurement, since 72% of the class stated that they had a shallow level of knowledge of this process. In the second experiment this dimension was inexpressive, having a total of 9 questions and 1,125 questions on average.

V. Conclusion

In this paper, we present a granular approach to the teaching of the software measurement process in order to promote discussion about the topic in the classroom. One of the main limitations encountered was the lack of work related to the teaching of software measurement based on games, being one of the main challenges encountered in the development of this work to support best practices on the lessons learned from them. In addition, the following are listed some limitations and points of improvement of the experimental study conducted in this work:

- Flexibility of the schedule;
- Much information condensed in the same classroom;
- Few exercises and examples in the area of Information Technology;
- High complexity in practical tasks;
- Balancing the teaching approach by not allowing exploitation of its failures (exploitation);

As future work, this study will serve as an input to improve the gamified classroom and direct the development of a serious game for teaching the measurement process in the context of software projects. In addition, in order to validate the tools, new interactions of the experiment will be performed in order to compare and verify with the user if the game-based approach is considered appropriate in terms of content relevance, correctness, the difficulty degree and method of teaching.

Also, this is a work in progress. That is, the final results comparing the traditional means of teaching with the gamified has not yet been contemplated.

Acknowledgment

The authors would like to thank the students that participated in the experimental study and the Amazon Foundation for Studies and Research Support by granting a doctoral scholarship to the PPGCC / UFPA to carry out this research.

References

[1] SOFTEX, “MPS.BR - Brazilian Software Process Improvement General Software MPS Guide,” 2016. Available in: https://www.softex.br/wp-content/uploads/2016/04/MPS.BR_Guia_Geral_Software_2016-com-ISBN.pdf. Last cess: 2018-02-28.

[2] C. Dekkers, and P. Mcquaid, “The dangers of using software metrics to (mis) manage,” IT professional, IEEE, v. 4, n. 2, p. 24–30, 2002.

[3] M. Kasunic, The state of software measurement practice: results of 2006 survey. Technical Report, CMU/SEI, 2006.

[4] C. Jones, Applied software measurement: global analysis of productivity and quality. McGraw-Hill Education Group, 2008.

[5] J. Iversen and O. Ngwenyama, “Problems in measuring effectiveness in software process improvement: A longitudinal study of organizational change at danske data,” International Journal of Information Management, v. 26, n. 1, p. 30 – 43, 2006. ISSN 0268-4012, 2006.

[6] M. Díaz-Ley, F. García, and M. Piattini, “Implementing a software measurement program in small and medium enterprises: a suitable framework,” IET software, IET, v. 2, n. 5, p. 417–436, 2008.

[7] S. Löper and M. Zehle, Evaluation of software metrics in the design phase and their implication on case tools. Blekinge Institute of Technology, 2003.

[8] M. Villavicencio and A. Abran, “Towards the development of a framework for education in software measurement,” In: 2013 Joint Conference of the 23rd International Workshop on Software Measurement and the 8th International Conference on Software Process and Product Measurement. p. 113–119, 2013.

[9] E. Herranz, R. Colomo-Palacios, A. Seco, and M. Yilmaz, “Gamification as a disruptive factor in software process improvement initiatives,” J. UCS, v. 20, n. 6, p. 885–906, 2014.

[10] K. Werbach and D. Hunter, For the Win: How Game Thinking Can Revolutionize Your Business. Wharton Digital Press. ISBN 9781613630228, 2012.

[11] O. Pedreira, F. García, N. Brisaboa, and M. Piattini, “Gamification in software engineering – a systematic mapping,” Information and Software Technology, v. 57, p. 157 – 168, 2015. ISSN 0950-5849, 2015.

[12] M. Kosa, R. O’Connor, M. Yilmaz, and P. Clarke, P., “Software Engineering Education and Games: A Systematic Literature Review,” Journal of Universal Computer Science, 22, n. 12, p. 1558–1574, ISSN 0948-695X. 1st Gamification Workshop at the 22nd EuroAsiaSPI Conference, Ankara Univ, Ankara, TURKEY, 2016.

[13] C. Wangenheim, M. Thiry, and D. Kochanski, “Empirical evaluation of an educational game on software measurement. Empirical Software Engineering,” Kluwer Academic Publishers, Hingham, MA, USA, v. 14, n. 4, p. 418–452, ago. ISSN 1382-3256, 2009.

[14] A. Rocha, G. Souza, and M. Barcellos, Software Measurement and Statistical Process Control. MCTi: Série de Livros PBQP, 2012.

[15] R. Luz and A. Neto, “Using programming dojos for teaching-driven development testing,” In: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE), 2012.