

A teaching proposal for the software measurement process using gamification: an experimental study

Lennon Sales Furtado

*Graduate Program in Computer Science (PPGCC), Institute of
Exact and Natural Sciences (ICEN)
Federal University of Pará (UFPA)
Federal Institute of Pará (IFPA)
Belém – PA, Brazil
lennonsfurtado@gmail.com*

Sandro Ronaldo Bezerra Oliveira

*Graduate Program in Computer Science (PPGCC), Institute of
Exact and Natural Sciences (ICEN)
Federal University of Pará (UFPA)
Belém – PA, Brazil
srbo@ufpa.br*

Abstract—This Research to Practice Full Paper presents a proposal to teach the software measurement process from a gamification class that makes use of games elements 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 a Federal University 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. Finally, the gamification had a positive effect on those involved when it is not only the use of game elements in a non-gaming context, but it is the intelligent use of this concept. In conclusion, the proposal helped in the teaching process of 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 aims to collect, store, analyze and report 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].

The first hypothesis for the understanding of this problem lies in the teaching of this subject [5][6][7], in that it is little explored in the syllabus of undergraduate students, thus relegated to the background, and little incentive 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 and serious games [9].

The definition of Gamification consists of the use of games 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 games mechanics and elements, making the task or subject more attractive [11].

And According to [22], Serious Games are a challenge with a set of rules aimed at training or teaching in a playful way. One of the distinguishing characteristics of this type of game is that it is oriented in training and teaching, that is, its focus is not on entertainment, even though the fun is usually part of the user experience. The main advantages of using serious games are [23]: to rescue the pleasure in learning; create an environment where the student constructs his knowledge in an active way; develop concepts that are difficult to understand in a playful way; to enable the student in making and evaluating decisions; promote socialization among students; and allow the teacher to diagnose learning difficulties of the content taught.

Therefore, this work aims to present a gamified approach combined with a serious game to teaching the subject of software measurement. And with the findings, improve the present methodology and use these inputs as a basis for the development of a tool to support the teaching of software measurement.

For more, two experiments were carried out with a total of 22 students from the Graduate Program in Computer Science at a Federal University in 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. 2 theoretical classes, 2 practical classes, 2 practical evaluations, 1 theoretical evaluation and 1 feedback class composed the approach. In addition, these classes used gamification as a tool to promote discussion in the classroom, and in addition to this, the last component of this approach was the use of a serious game to solidify the student's theoretical knowledge.

In addition to this introductory section, which covers aspects that justify and identify the question about the study, this paper is structured as follows: Section 2 presents and discusses some related works, Section 3 presents the adopted gamification design, in Section 4 we have the explanation of the experiment carried out, in Section 5 the results will be presented, in Section 6 the discussions about the results obtained are presented, Section 7 presents the threats to the validity of this work, and, finally, Section 8 will report some final considerations.

II. RELATED WORKS

A search for related work was done in systematic reviews that addressed of teaching the software process improvement and software measurement using gamification or serious games. Among the reviews analyzed are the authors of [11][12]. In order to verify any paper not covered by these revisions, another search was made, but the search engines of the main indexes of scientific content, as IEEE, ACM DLL, Science Direct, Scopus and Web of science were used, which the authors' institution have access. In return, a list of related papers has been obtained. With this, it was possible to realize that few works cover the process of software measurement. The following are the most relevant papers on software measurement found.

The most relevant for this study was the Wangenheim, Thiry and Kochanski's work [13]. This is a serious game for teaching software measurement, using the Goal-Question-Metrics (GQM) paradigm, which reports an experimental study conducted with 15 students to validate the proposed serious game. This study is the only one that has focused exclusively on the software measurement process and its steps (Collect, Store, Analyze and Report). Due to its importance, this work was used as a teaching tool during the class.

The article [24] uses the Gaimware framework to propose a gamification to assist in the adoption of a Gamified Software Process Improvement program. First, the authors use Bartle's test [21] to identify the profile of users and adjust gamification to them. Then, they present a gamification proposal that uses a spreadsheet as a tool to record user points and levels in an environment that alternates between competitive and cooperative. In addition, the authors conducted an experiment with 6 users of a company. In which it was possible to perceive an increase in the team's motivation when adopting the software improvement processes, however, it was not noticeable a significant increase, according to the authors, of the team's intrinsic motivation. Another benefit was the gain in the performance of process improvement, however this was not a lasting benefit due to external factors. As metrics are used, ratio of project activities correctly identified without needing to be broken down, ratio of project activities having a variation against the estimation time higher than 10% and ratio of project activities of actual and estimated time effort. And as a validation method, the researchers applied the IMI before and after the implementation of the gamification initiative.

The work [25] aims to encourage students to write higher quality codes, in the C programming course, through the use of gamification (Points, Leaderboard, instant feedback and competition) when punctuating the metrics: lines of code and cyclomatic complexity. Where two groups A and B. were set up with a total of 35 students, that used the gamified proposal for 3 weeks at different intervals, group A used it at the beginning of the course and B at the end. Where, a significant gain in the quality of the Code was attested in the participating groups. However, it was noticed that some students explored the proposal's vulnerabilities, by creating codes that achieved a high score by reducing its cyclomatic complexity, however without improving the quality of the code. The authors point out that the competitive environment caused this behavior in some students.

The purpose of the article [26] was to teach Cosmic Points for effort measurement through a serious game, called Back to Penelope, which had as its main game elements: Time and narrative. With this game, it proposed a means of measuring points functions using class diagram. And as a highlight, this proposal made use of graphics based on pixel art; and the use of the adventure genre; with sound effects and soundtrack, for the ambience of the player. A total of 40 students participated, in which it was realized that the students' estimate was 17.6% higher than the real value. Which, in the authors' view, showed that the students estimated with an adequate accuracy, knowing that this is their first contact with the COSMIC method. Corroborating, as evidence for the efficiency of the use of serious games in teaching effort estimation.

In addition, a study was found that reports a teaching structure for the subject of software measurement [8], which introduces concepts of gamification into teaching and learning activities. This framework is based on Bloom's taxonomy on learning levels and adopts a constructivist approach. For this reason, the framework established by Villavicencio and Abran [8] is adequate to this work, since the involvement and commitment of students is incorporated in the learning process. This can be achieved by incorporating the concepts of games elements and mechanics that can be defined in the gamification area.

III. THE GAMIFICATION

This section presents some details of the proposed gamification desing for teaching of software measurement.

A. Overview

To foster engagement from the students in the subject of Special Topics in Software Engineering: Software Processes Technology, in the authors' institution, were introduced gamification practices, where students were scored for quantity and quality of the level of interaction in classroom. The making of the approach gamification proposal was based in the framework presented by Werbach and Hunter [10], which are condensed in the following six steps:

- Define Business Objective: to create a competitive and collaborative environment that fosters communication in the classroom, in order to broadly discuss the proposed software measurement topic,
- Outline Target Behaviors: arrive early, make the class or extra classes activities, attend classes, provide suggestions for the experiment, participate in the classroom and 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 to get bonus points. These points were awarded in different dimensions of participation, as can be seen in Table I. The players (students) only lose points when they commit one of the faults defined 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. The Awards

In case of a tie, the sum of the practical and theoretical evaluation scores was used as the tie-break criterion. In addition, students in the top three places were awarded as follows:

- First place: 1 extra point in the final grade and a development course for mobile devices,
- Second place: 1 extra point on the final grade,
- Third place: 0.5 extra point on the final grade.

D. The Software Tool

A spreadsheet editor was used to collect and store class data.

E. Improvements from the First Experiment

Some points of improvements from the first experiment to the second were identified, listed and applied, as can be seen in Table III.

TABLE III. IMPROVEMENT POINTS FROM THE FIRST TO THE SECOND EXPERIMENT

Improvement Point	Solution
The inclusion of metrics used in agile methods	We presented a section of metrics focused on Agile environments.
Less complexity of the task	Decrease practical lesson difficulty by reducing 10 to 7 tasks
More in depth in the GQIM paradigm (Goal, Question, Indicators, Metrics)	Extra-class list on the paradigm
The rules of gamification were not clear for some students	A slide presentation was made containing all the rules during the first class of the course and it was made available for future class inquiries
Exercises to be done at home counting points for gamification	Activities on GQIM and agile metrics followed this purpose
Inclusion the mechanics of achievement by giving medals to students who reach certain milestones	For this, a system of medals was created that allowed the student to follow his achievements and milestones in the course.

IV. THE EXPERIMENT

The teaching of software measurement experiment had a total of 8 (eight) days of classes: 2 (two) days of theoretical classes, 2 (two) days of practical classes, 1 (one) day for the theoretical evaluation, 2 (two) days for the practical evaluation, and on the last day a class was held to identify the students' subjective impressions about the teaching methodology and, with this, to highlight strengths and opportunities for improvement regarding the class.

In addition, a class monitor participated as a judge, who was responsible for taking notes and assign points to the students from the observation of their behavior in the classroom. In summary, the experiment had the following structure: Pre-Questionnaire, Theoretical Classes, Practical Activities, Practical Evaluation, Theoretical Evaluation, Post-Questionnaire and Feedback Class. In addition, 22 students participated in these experiments, 15 in the first experiment and 7 in the second experiment, from undergraduate program in computer science at author's University 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 in [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.

It was asked if the students had seen some kind of measurement paradigm during their academic life, among the options listed were: GQM (Goal-Question-Metrics), PSM (Practical-Software-Measurement), Function Points Analysis, GQIM (Goal-Question-Indicators-Metrics) and others. As a result, only two students known all the paradigms, and all the others had no knowledge of any paradigm. While another question addressed whether the students had already participated in a measurement program, and only 3 students have already had contact with such a program. Finally, the students' level of knowledge about the measurement process was asked and 18.2% stated that they had no knowledge about the process, 72.7% said they had low knowledge on the subject and 9.1% indicated that had average knowledge in such a process.

This scenario seems to be representative of how much the software measurement process is neglected in information technology courses.

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 [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 X-Med serious game [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 and after that they were replaced by someone from the audience for 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 X-Med serious game. 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 spreadsheet 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. The eleventh question was exclusively for gamification and did not count test points, acting as a multiplier, where the student

responding correctly earned 15% bonus points from the total points earned in gamification. If the student answered the question incorrectly, the student lost 10% of their total gamification points.

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 teaching of software measurement.

V. THE RESULTS OBTAINED

In this section, the quantitative and qualitative results obtained from the execution of the experiments will be presented and analyzed.

A. Qualitative Analysis

Both the questionnaires and the feedback class reported similar results. The results of these approaches are presented below in Tables IV, V, VI, VII and VIII.

TABLE IV. EXPERIMENT 1 - QUALITATIVE EVALUATION

Experiment 1	Poor (%)			Regular (%)			Good (%)			Excellent (%)		
Content	0	0	0	0	0	0	60	50	50	40	50	50
Sequence of topics	0	0	0	10	0	0	70	60	60	20	40	40
Difficulty Level	0	0	0	40	20	30	40	50	40	20	30	30
Duration	10	30	20	40	30	30	20	20	20	30	20	30
Teaching method	0	0	0	0	0	0	50	50	50	50	50	50
Exercises	0	0	0	10	10	10	50	40	60	40	50	30
Question	1.1	2.1	3.1	1.1	2.1	3.1	1.1	2.1	3.1	1.1	2.1	3.1

TABLE V. EXPERIMENT 2 - QUALITATIVE EVALUATION

Experiment 2	Poor (%)			Regular (%)			Good (%)			Excellent (%)		
Content	0	33	50	50	50	17	33	17	33	17	0	0
Sequence of topics	0	17	33	83	67	50	0	17	17	17	0	0
Difficulty Level	0	33	17	50	50	67	33	17	17	17	0	0
Duration	33	50	33	33	17	33	33	33	33	0	0	0
Teaching method	33	33	33	50	33	33	17	33	33	0	0	0
Exercises	50	33	50	17	50	33	33	17	17	0	0	0
Question	1.1	2.1	3.1	1.1	2.1	3.1	1.1	2.1	3.1	1.1	2.1	3.1

TABLE VI. EXPERIMENT 1 AND 2 - ACCEPTANCE OF THE CLASSES

Experiment 1	Q1.2	Q2.2	Q3.2	Experiment 2	Q1.2	Q2.2	Q3.2
Excellent	60%	70%	60%	Excellent	0%	0%	0%
Good	40%	30%	40%	Good	33%	17%	33%
Regular	0%	0%	0%	Regular	50%	83%	33%
Poor	0%	0%	0%	Poor	17%	0%	33%

TABLE VII. EXPERIMENT 1 - GAMIFICATION EVALUATION

Experiment 1	Poor	Regular	Good	Excellent
Game Elements	0%	0%	30%	70%
Game Mechanics	0%	10%	40%	50%
Difficulty Level	0%	20%	40%	40%
Game Balance	10%	20%	30%	40%
Entertainment	0%	10%	30%	60%

TABLE VIII. EXPERIMENT 2 - GAMIFICATION EVALUATION

Experiment 2	Poor	Regular	Good	Excellent
Game Elements	17%	83%	0%	0%
Game Mechanics	50%	50%	0%	0%
Difficulty Level	17%	50%	33%	0%
Game Balance	33%	50%	17%	0%
Entertainment	67%	33%	0%	0%

At the end of the theoretical evaluation, a questionnaire adapted from the Wangenheim, Thiry and Kochanski's work [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 questions of this questionnaire will be presented in this section.

The most relevant for the area of evaluation of the theoretical classes were questions:

- Q1.1 How do you evaluate the content, topic sequence, difficulty level, duration, teaching method and exercises in theoretical classes?
- Q1.2 How much did you like the theoretical class?

For the first question, 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 regular criterion. 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.

Whereas in the second question it was reported that 60% of the students liked the theoretical class, while the other 40% liked very much during the first experiment. For the

second, these criteria fell, which 33.3% liked, 50% liked little and 16.7% did not like. While for the evaluation of the practical class, the same questions were asked (Q2.1 and Q2.2), and their results will be presented below.

For question 2.1, 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.

In addition, for question 2.2, 70% of the students evaluated these classes with the criterion Excellent and 30% with the criterion Good, in the first experiment. For the second, 83.3% liked little and 16.7% liked the practical activity, showing a great disparity between the experiments, which can be understood by the difference between the profiles of students who participated in them.

The evaluation of the practical evaluation method was done with the same questions raised in the evaluation of the practical class (Q3.1 and Q3.2). The results were similar to those already presented for the evaluation of the practical class, with the main difference being that only 60% of the students applied the criterion Excellent and the other 40% applied the criterion Good in question 3.2 of the first experiment. And for the second experiment, it was also similar to the results of the practical class, thus having a large portion of the class evaluating this stage in a neutral or poor way.

For the evaluation of gamification, the main questions raised were as follows:

- Q4.1 How do you evaluate the games elements and mechanics, difficulty level, game balancing and gamified classroom entertainment?
- Q4.2 How much did you enjoy the class?

For the first experiment, the first question 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, as can be see in Table VII. This consequently reflects in the second question, where it is pointed out that 70% of the students considered their experience with the class to be excellent, another 20% considered the Regular experience and the other 10% had a good experience.

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, as depicted in Table VIII. In addition, the second question indicated that 50% were neutral, 16.7% did not like it and 33.3% liked the gamification. As can be see in Table IX.

TABLE IX. EXPERIMENT 1 AND 2 - GAMIFICATION ACCEPTANCE

Experiment 1	Q4.2	Experiment 2	Q4.2
Excellent	70%	Excellent	0%
Good	10%	Good	50%
Regular	20%	Regular	33%
Poor	0%	Poor	17%

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. As penalties, the class had the following dimensions: Delay in 10 minutes (ATM), Foul with Warning (FCA), Missing Warning (FSA) and Penalties (PEN).

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.

As for the missing class with warning, there were a total of 4 incidences in the first experiment and 1 incidence in the second experiment. In the first experiment, there were no incidents of this dimension in the first experiment. Both dimensions were not observed correlation with the gamification, that is, the same influenced little or nothing.

The class was evaluated in the following positive classroom interactions: Presence, Initiative, Suggestions, Participation, Questions and Arrive Early. 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 perceptible, 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.

For the initiative dimension, there was a great contrast and a perceptible evolution in such behavior within the classroom. To illustrate, in class 1 students had 5 incidences of this dimension and 12 for the first and second experiment, respectively, while in class 3 there were a total of 45 incidences, in other words, 9 times higher than at the

beginning of the first experiment and 24, that is, double for the second experiment. Notably, this dimension was maintained with an average of 19 incidences throughout the first experiment and 10 times the second. As for the suggestions, these had an average of 3.75 and 3.62 incidences per class for each experiment. Thus, the lowest average of incidences between the other interactions.

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.

As a result of the modules of the teaching proposal, the following grades were obtained in the first and second experiments, respectively: for gamification 1.57 and 1.48 of the total of 2, for the practical evaluation 3.3 and 3.4 of the total of 4, for theoretical evaluation 2.87 and 2.52 of the total of 4, and in the final grade was obtained 7.94 and 7.69 of the total of 10.

VI. DISCUSSIONS

As a result of the first experiment, a level of over 90% of evaluations was obtained with the good and excellent criteria for the theoretical class, 100% of criteria good and excellent for practical class and for practical evaluation, and 80% of the criteria were good or excellent for gamification. During the conduct of the first experiment, it was evident the level of acceptance of the proposal and the role-played by it. Their biggest contribution was to mitigate social dynamics, that is, often creates an environment where only a few students participate in the room. This phenomenon was amortized based on the game-based approach, because students compete and collaborate with each other.

Already the second experiment, for the most part, the evaluation was neutral or bad. In short, the first experiment presented better results; this can be understood by the profile of the students who were more receptive to approaches based on games. However, in both experiments the expected result was achieved, that of helping students to know more about the software measurement process. As indicative for this evidence can be used the final result, in which the two experiments had a grade above 7, that is, reached the concept Good by the standards of the author University.

It is worth mentioning that even the students, who did not have a constant participation or committed to the

gamification, ended up benefiting from the discussion of the software-metering theme during the classes, which was fomented by the application of gamification in the classroom.

Finally, on the null hypothesis described, we can say that, from the results obtained and described in Section 5, that the elements and mechanics of gamification generated significantly different results when adopting a teaching approach based on serious games, by the cooperation factor between the participants of the game, or by the competition that these assets allowed the students to promote when carrying out the planned activities. In Table X is presented the Strengths, Weaknesses, Opportunities and Threats (SWOT) to a better comprehension of the approach.

TABLE X. SWOT ANALYSIS

SWOT Analysis	
<i>Strengths</i>	<i>Weaknesses</i>
<ul style="list-style-type: none"> • More fluid learning and without the pressures promoted by traditional teaching; • Mitigate social dynamics, that is, often creates an environment where only a few students participate in the room. • The gamification itself is a way to measure students performance and give constant feedback. 	<ul style="list-style-type: none"> • Use case studies of software measurement during class; • The time and difficulty level established for dojo's tasks was unbalancing; • Go beyond the PBL (Points, Badges and Leaderboard). Many students felt the absence of more complex game elements.
<i>Opportunities</i>	<i>Threats</i>
<ul style="list-style-type: none"> • Present the techniques and paradigms in a more contextualized way. As a solution can be used the application of a paradigm within a case study; • Bring examples of measurement practice to experiment and discuss about it; • The serious game (X-MED) has Interface with confusing visual elements; absence of music or ambiance sound effects; conveyed more the feel of a traditional evaluation than a game itself; and has several bugs and wrong answers inside the game. Which is a great opportunity to change or develop a more feasible game. • A tool to see real-time progress of the classroom gamification. 	<ul style="list-style-type: none"> • Clarity was lacking for understanding the gamification rules; • Some students may not like the competition proposed by gamification; • Leaderboard can act to the last classified as a demotivating factor; • Only ludic measurement practices may not help students in real life situations activity that represents the reality of creating a measure program in companies.

VII. THREATS TO VALIDITY

This section presents and discusses different threats to the validity of this work, based on the four most common groups of threat to validity [16]: internal validity, external validity, construction validity and validity of conclusion.

The internal validity refers mainly to independent variables, that is, it is argued if the research was conducted in the correct way. In order to mitigate causes of internal validity, it was created and revised, together with the software engineering specialist, who is also the guiding principle of this work, a set of documents that presented each of the steps and tasks present in each class that would be seen in the classroom.

The external validity is concerned with identifying if other researchers can replicate this work, and if the results

would be consistent with those found by this study. In order to reduce the risk related to external validity, the following practices were adopted: class planning documents and constant meetings among researchers to verify the adherence between the execution of the study and what was being carried out.

The construction validity is related if the measures actually used represent what one wanted to measure, i.e., if the data collected meets the research questions. The research questions were created in conjunction with an education specialist and gamified education approach. Such a researcher has extensive experience in such areas and collects publications within these topics, such as [17][18][19]. Frequent consultations were carried out with this researcher in order to reduce issues regarding construction validity problems.

The validity of conclusion is whether the conclusions are correctly supported by the data raised. A threat to this validity is the non-comparison of the performance obtained in the approach with a class that does not use it. This hinders the interpretation of the data obtained related to the contribution of the approach to student performance. In addition, the absence of more in-depth questions to investigate the profile of the student as the familiarity with games and the age. These are some of the improvement points for the next versions of the experiment.

VIII. 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. And beyond to the generic contribution of gamification, which was to mitigate social dynamics, a competitive and a cooperative environment, and the constant presence of challengers in order to stimulate the students to seek, by himself, knowledge about the area.

More specific contributions were attributed to gamification. Like, gamification by itself is a form of measurement. As shown in the introduction, software measurement is divided into four parts: collect, store, analyze, and report. The gamification applied contemplates each of the parts. By collecting student data on each of the dimensions presented in the proposal on a daily basis. All the data collected were stored in a spreadsheet, where bar and radar graphs were generated to present respectively, the classroom leaderboard and each student progress in each independent dimension. At the end of each class, the respective graphs were reported and sent to the students to analyze their self-performance. That is, the student learned in practice the steps of a software measurement process with the gamification itself that served as a meta-measurement.

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 Information Technology;
- Few examples illustrated by figures and graphs;
- High complexity in practical tasks;
- Balancing the teaching approach by not allowing exploitation of its failures (exploitation);
- Present a more balanced rating scale, as there are two positive options ("excellent" and "good"), a neutral ("regular") option and a negative ("bad") option. It lacks the inclusion of one more negative option, as "poor" as opposed to "excellent";
- Select the improvements presented in the feedback class;
- Expand the student profile by applying Bartle's test [21] to identify students game profile.

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. Therefore, the null hypothesis present in this text remains open for future work.

Finally, the gamification has a positive effect on those involved when it is not only the use of game elements in a non-gaming context, but it is the intelligent use of this concept. As pointed out in [10], gamification proposals that do not bother focusing on the social, cognitive and emotional aspects of games pass, tend to be quickly ignored because they are shallow and do not hold the player's attention. Finally, as stated by Koster and Wright [20] "games in the end are teachers, and fun is another word for learning."

ACKNOWLEDGEMENTS

The authors would like to thank the students participating in the experiment and the collaborators who participated in the experiment in the Judge profile.

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 ccess: 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. Diaz-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.
- [16] R. Mauricio, L. Veado, R. Moreira, E. Figueiredo, and H. Costa, "A systematic mapping study on game-related methods for software engineering education". Information and Software Technology, 2017.
- [17] R. Chaves, C. von Wangenheim, J. Furtado, S. Oliveira, A. Santos, and E. Favero, "Experimental evaluation of a serious game for teaching software process modeling". IEEE Transactions on Education, 58(4):289–296, 2015.
- [18] I. Elgabry and S. Oliveira, "Gamification and evaluation of the use of agile tests in software quality disciplines: A case study". 13th International Conference on Evaluation of Novel Approaches to Software Engineering, pages 416–423, 2018.
- [19] L. Furtado and S. Oliveira, "A teaching method for software measurement process based on gamification". ICSEA 2017, 15(16):12, 2017.
- [20] R. Koster and W. Wright, "A Theory of Fun for Game Design". Paraglyph Press. ISBN 1932111972, 2004.
- [21] R. Bartle, "Hearts, clubs, diamonds, spades: Players who suit MUDs". Journal of MUD research 1,1. 1996.
- [22] M. Zyda, "From visual simulation to virtual reality to games". Computer, 38(9):25–32. 2005.
- [23] B. Oliveira, P. Afonso, and H. Costa, "Testeg A computational game for teaching of software testing". In Computer Science Society (SCCC), 2016 35th International Conference of the Chilean, pages 1–10. IEEE. 2016.
- [24] E. Herranz, J. G. Guzmán, A. de Amescua-Seco and X. Larrucea, "Gamification for software process improvement: a practical approach". in IET Software, vol. 13, no. 2, pp. 112-121, 2019.
- [25] R. Kasahara, K. Sakamoto, H. Washizaki, and Y. Fukazawa, "Applying Gamification to Motivate Students to Write High-Quality Code in Programming Assignments". In Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE '19), New York, USA, 92–98. 2019.
- [26] B. Marín, M. Vera, and G. Giachetti, "An Adventure Serious Game for Teaching Effort Estimation in Software Engineering". IWSM-Mensura. 2019.