Hindawi
International Journal of Computer Games Technology
Volume 2021, Article ID 8873997, 35 pages
https://doi.org/10.1155/2021/8873997

Review Article

Teaching Method for Software Measurement Process Based on Gamification or Serious Games: A Systematic Review of the Literature

Lennon Sales Furtado,[1,2] **Rafael Ferreira de Souza**,[1] **João Luís dos Reis Lima**,[1] **and Sandro Ronaldo Bezerra Oliveira**[1]

[1]*Graduate Program in Computer Science, Federal University of Pará, Belém 66075-110, Brazil*
[2]*Federal Institute of Pará, Paragominas 68625-000, Brazil*

Correspondence should be addressed to Lennon Sales Furtado; lennonsfurtado@gmail.com

Received 7 August 2020; Revised 3 February 2021; Accepted 6 May 2021; Published 11 June 2021

Academic Editor: Michael J. Katchabaw

Software process improvement programs are partly founded on software measurement. However, despite their importance, it has been pointed out in the literature that many students are leaving the academic world without the necessary skills to conduct this kind of process. This can be understood by people's attitudes to this process which is regarded as time-consuming and difficult to understand—factors that explain the lack of interest in it during a student's academic life. In light of this, the application of serious games or gamification can show useful alternative ways of meeting this need, because the strategies they involve are well accepted by students and have a motivational and engaging effect on them. The objective of this work is to discover different approaches to the teaching of software measurement and software process improvement through gamification projects and serious games. This involves carrying out a systematic review of the literature, which is aimed at characterizing the state-of-the-art on the use of methods related to gamification and serious games in the abovementioned subjects. We conducted a systematic review of the literature to identify primary studies that address the use, planning, or evaluation of gamification, serious games, their features, and game mechanics in software engineering. We located 137 primary studies, published between 2000 and 2019. Although the use of serious games and gamification in software engineering is not recent, there still remains a large area to be explored, especially in software process improvement and software measurement. The study expands and advances the research on how serious games and gamification proposals can be used for teaching software measurement in the context of software process improvement programs by conducting a systematic review of the literature.

1. Introduction

Software engineering is directly related to the generation of high-quality software product. This quality reduces the need for rework, and less rework results in a faster delivery time [1]. In other words, software engineering seeks to ensure the quality of processes that involve software development. Among the approaches adopted to achieve this goal are software process improvement (SPI) programs, which are based on measurement practices.

The software measurement process entails collecting, storing, analysing, and reporting the data on the products developed, as well as the implemented processes of a given organization, to further its organizational objectives [2]. This process is a key strategy in the software process improvement programs; however, the software industry has been hesitant in applying efficient measurement programs [3, 4]. This is due to the fact that many software managers and professionals, including academics in software engineering and computer science, are not fully aware of the application of this subject [5]. Although this work investigated only these two courses, it is understood that other courses derived from computing, such as Information Systems and Computer Degree, go through the same fundamental problem of not exploring software measurement in their curricula, consequently training professionals with little knowledge in this area.

People's attitudes are based on the assumption that the measurement process is difficult to master and time-consuming [6–9]. The first approach that is needed for an understanding of this problem lies in the question of how this subject should be taught [10], since it does not feature prominently in the undergraduate curriculum, and is often relegated to the background; therefore, these students receive little incentive to learn this practice. Another factor is the absence of guidelines for assisting students in the practice of measurement [11–13].

In general, human factor is determinant for the success of every measurement program, since if there is not a suitable degree of motivation and commitment to the measurement program, it is unlikely that it will achieve the desired result—the control of software metrics to assist decision-making. Among the alternative means of ensuring that people involved in the SPI program are fully engaged is the adoption of the gamification concept [14].

Gamification can be defined as the use of game elements and game design techniques outside the context of games [15]. According to Breuer and Bente [16], the games represent an intersection between the different learning strategies that allow serious games to serve as a subset of e-learning (electronic learning), educational entertainment, and game-based learning. According to Zyda [17], serious game defines a certain form of a game that uses computer games and simulation approaches and/or technologies for primarily non-entertainment purposes. The "serious" term refers to the game being aimed more at educational than entertainment purposes. These approaches seek to improve the engagement, motivation, and performance of a user in carrying out or learning some task or subject, by incorporating mechanics and game features, which makes them more attractive [18].

The objective of this work is to find teaching solutions for the subjects of software process improvement and software measurement by making use of gamification or a serious game, with a view to devising good practices, a suitable framework, validation methods, and in particular, the features, mechanics, and dynamics of games that can be more effectively employed for teaching purposes. As different authors in the literature have advocated numerous approaches for making use of games as a teaching tool, it has become necessary to find a mechanism that can allow a more suitable choice to be made from among different solutions. Mafra and Travassos [19] argue that the desired solutions should be found in an intensive and systematic adoption of an evidence-based approach.

A systematic review of the literature can be carried out as a reporting mechanism, which is the means by which a researcher can determine what expert knowledge is required in a given area to plan his research, while avoiding unnecessary duplication of effort and repetition of past errors [19]. A preestablished protocol is essential to mitigate errors related to the validity of the review carried out and ensure that this review does, in fact, have scientific value and potential for repetition. Unless this occurs, there is a risk that reviews become dependent on the researchers and hence reduce their reliability. Thus, a systematic review of the literature was carried out, in which 137 primary studies were investigated, these were analysed to find solutions for the teaching of software process improvement and software measurement based on the use of serious games and gamification.

The next sections of this article are structured as follows: Section 2 will provide an overview of software process improvement, software measurement, gamification, and serious games, Section 3 is aimed at setting out the methodological procedures of the systematic review of the literature (SRL) adopted in this work, Section 4 examines the results obtained from this SRL and attempts to answer the bibliometric and research questions defined in this work, Section 5 discusses the research carried out in the literature and the findings, Section 6 discusses different threats to the validity of this systematic review, Section 7 investigates some related work, and, finally, Section 8 summarizes the research undertaken, as well as making some recommendations for future work in the field.

2. Background

In this section, we define the underlying concepts that are needed for an understanding of this article, which are software process improvement, software measurement, gamification, and serious games.

2.1. Overview of Software Process Improvement and Software Measurement. A process works like a "glue" that keeps people, technologies, and procedures tightly bound together and is used by software engineers to design and develop computer programs [20]. Organizations involved in software development have their own processes, and the standard of these processes tends to influence the quality of the developed product. Thus, it is of great value for organizations to remain competitive by investing in software process improvement programs. These programs are aligned with process improvement goals, which are a set of desired and defined objectives that can guide the process improvement in a practical and measurable way [21]. The goals should give added value to a company's business and improve the quality of the goods produced.

Thus, it is necessary to have mechanisms capable of evidencing problems in the processes and to support the identification of improvement objectives [22]. The mechanism used as a thermometer to verify the health of a process is the measurement process, as it is the basis for the control, improvement, and understanding of the behaviour of a product or process from a quantitative evaluation [5]. This process serves as an aid in decision-making, because "you can only control what you can measure" [23], and "you can only predict what you can measure" [24].

Despite the importance of the measurement process and software improvement programs, the way they are taught has proved to be inefficient, as is pointed out by Jones [25]. This author lists 28 problems related to the area of software measurement, one of the more recurrent, being the lack of training of those involved. The findings of this study are corroborated in a survey conducted by the Brazilian-American Chamber of Commerce [26], where 44 IT executives were interviewed. It was reported that 86% of them were not satisfied with the way the measurement process was being conducted in their companies, since most of the problems arose from the inability of the professionals involved to solve them; this was because most of them lacked the necessary skills to conduct the process efficiently.

A striking feature, which should be noted for understanding this problem, is the difference between what is taught in educational institutions and what is required by industry [27]. The needs of industry can only be met by adopting innovative practices that go beyond traditional lectures. Among the various approaches available, the use of games is believed to be a powerful tool and has a high acceptance rate among students of different ages from different backgrounds [28]. In addition, according to Bjork and Holopainen [29], computer games can help create a more attractive and stimulating environment for the contemporary generation of students than "paper versions."

2.2. Gamification. Gamification is one of the different teaching techniques that seeks to improve user engagement and motivation in carrying out or learning tasks [18]. Gamification involves using the elements, mechanics, and dynamics of games outside the context of traditional games [15]. In addition, this author outlines three reasons why gamification can serve companies; these reasons can be easily adapted to different situations—so much so that they will be outlined in the teaching context. The reasons for the success of gamification are based on three cornerstones: engagement, experimentation, and results.

With regard to engagement, as Koster [28] maintains: "With games, learning is the drug." Gamification acts as a form of extrinsic motivation, as well as a reinforcing mechanism. It responds to one of the intrinsic needs of humans, that is, to seek the chemical rewards released by the brain as a motivating "engine" for the execution of tasks. Thus, the stimulus created by this feedback strengthens engagement with the class and the learning process and keeps the students motivated and hence eager to be engaged.

With regard to experimentation, games do not usually have permanent punishments for those who fail them. As a result, they create a safe and often competitive or cooperative environment, which tends to stimulate participation by trial and error. This safe environment can be characterized as one of the most valuable contributions of serious games. In addition, according to Werbach and Hunter [15], serious games can be seen as a special type of gamification, as they make use of nonfocused games for entertainment.

Finally, results, as depicted in the studies carried out by Hamari et al. [30] and Pedreira et al. [18], show that the adoption of gamification in organizations has had positive effects, depending on their application in a given context. In addition, large software organizations have employed gamification to encourage users to carry out ordinary tasks, since they know it achieves results.

2.3. Serious Games. According to Zyda [17], serious games raise the challenge of producing a set of rules which is aimed at training or teaching in a playful way. One of the distinguishing features of this type of game is that it is geared towards training and teaching; that is, its focus is not on entertainment, even though the fun engendered is usually a part of the user experience. In addition to this, serious games create a simulating atmosphere with real-life situations in such a way as to create a safe environment for users to experiment with different solutions and learn by trial and error or cause and effect. The main benefits of serious games are as follows [31]: (i) to derive pleasure from learning; (ii) to create an environment where the students construct their knowledge in a dynamic way; (iii) to formulate concepts that are difficult to understand, in a playful way; (iv) to enable students making decisions and then assessing them; (v) to foster socialization among the students; (vi) to allow the teacher to diagnose learning difficulties in what has been taught.

Dale's Cone of Learning [32] (see Figure 1) has been used as a reference point in planning instructional strategies in higher education [33]. This same study points out that simulating the real experience can improve the understanding of what is being taught more effectively than learning by just reading or listening, that is, passively. Dale states that people remember 90% of what they have learned through simulation. The study by Aydan et al. [34] corroborates this result by suggesting that there is a significant difference between students who have learned ISO 12207 by simulation and those who have learned it only by the traditional means, i.e., by reading sections of texts. This prompted the authors of this study to say: "we recommend the use of serious games that seem to be superior to a traditional paper-based approach."

3. Materials and Methods

This section outlines the following: objectives and research questions, the method and the search strategy used to mine the relevant papers/articles to this study, the procedure for selecting and classifying a primary study, and, finally, the method of conducting data extraction.

3.1. Goals and Research Questions. This systematic review of the literature (SRL) seeks to find different approaches for teaching the process of measurement and teaching processes related to SPI by using gamification systems and serious games. By determining these approaches, this SRL will highlight the dynamics, mechanics, and game components and show how they can assist in the development of an educational tool for the teaching of software measurement. A set of research questions was prepared to meet the planned objectives. Owing to the complexity of this SRL, the questions were divided into two groups: general questions and specific questions.

3.1.1. General Questions. General questions are pertinent to both areas of this research, namely software process improvement and software measurement. The general questions will be listed below:

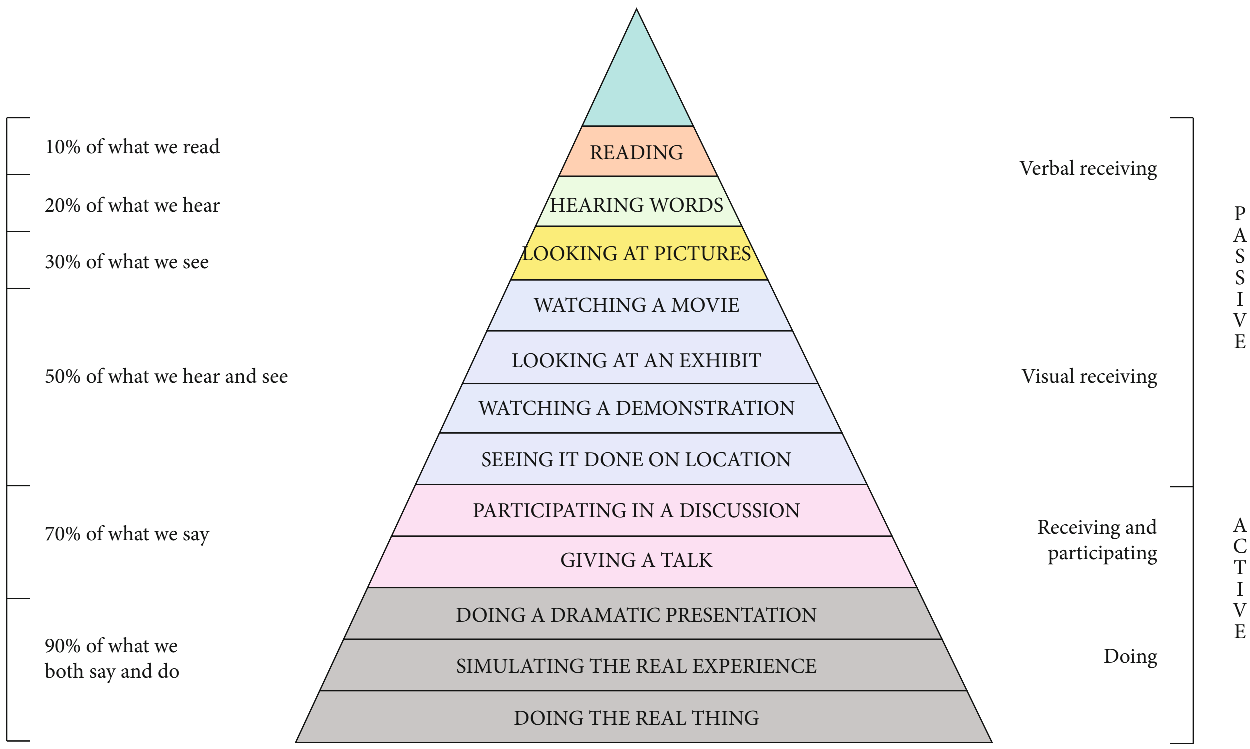

Figure 1: Dale's Cone of Learning [32].

(i) *GQ1*. In what contexts (i.e., academic or professional) did the gamification or serious game projects take place?

(ii) *GQ2*. What limitations have been reported in the use of gamification or serious games for teaching?

(iii) *GQ3*. What research methods were employed in the validation of the gamification or serious game projects?

(iv) *GQ4*. What game elements were included in the gamification or serious game projects?

(v) *GQ5*. What game mechanics were used in the gamification or serious game projects?

(vi) *GQ6*. What game dynamics were involved in gamification or serious game projects?

(vii) *GQ7*. What genres were included in the gamification or serious game projects?

(viii) *GQ8*. How does the effectiveness of learning through gamification or serious games compare with what is achieved by traditional learning?

3.1.2. Specific Questions. Specific questions, which as the name implies, are concerned with issues that are individually applied to each of the topics of this research. The specific question of software process improvement (SPIQ) is as follows:

(i) *SPIQ1*. In what processes (measurement and requirements collection, among others) were the gamification system or serious games applied in the area of SPI?

The specific questions about software measurement (MEAQ) are as follows:

(i) *MEAQ1*. Was the system employed based on a model or standard or paradigm? If so, which?

(ii) *MEAQ2*. What metrics were covered by the gamification or serious game projects?

(iii) *MEAQ3*. Which measurement activities (collect, store, analyse, and report) were covered by the systems?

(iv) *MEAQ4*. How can educators or the industry benefit from teaching or applying software measurement programs through gamification or serious game projects?

3.2. Method. This review lasted for 34 months, starting in February 2017 and continuing until December 2019, and was overseen by four researchers (one doctoral student, two undergraduates, and one supervisor) who carried out the activities of this systematic review of the literature, all of which were in the area of computer science. There were two searches in the selected databases: the first was in early 2018, and the second search was conducted in December of 2019. This SRL was based on the Kitchenham guidelines [35], and the method used is listed as follows:

(i) *Step 1*. To check and to validate the search strings to ascertain their accuracy in the return of the primary papers/articles and thus be able to create multiple instances of these strings adapted for each database,

(ii) *Step 2*. Search for possible primary papers/articles in the science citation index; there were available from the domain of the Federal University of Pará. This domain allowed free access to the papers/articles from the selected scientific databases,

(iii) *Step 3*. Read the titles and abstracts of the papers/articles returned by the search string, to create a list with the possible primary papers/articles,

(iv) *Step 4*. (a) Read the titles, abstracts, introductions, and conclusions of the papers/articles in the list of possible primary studies; (b) apply the inclusion and exclusion criteria to reject false positives; and (c) create a list of the primary papers/articles included and a list of those excluded,

(v) *Step 5*. Compare and combine the lists of different researchers, and if there is disagreement among the researchers over the inclusion or exclusion of a paper/article, this one should also be included,

(vi) *Step 6*. Read the papers/articles in the final list in full and apply the quality criteria to grade the remaining ones,

(vii) *Step 7*. Extract the data of all the papers/articles found in the list compiled previously.

In addition, all the documents and procedures were validated from meetings with the supervisor of the SPIDER Project (Software Process Improvement: Development, and Research) [36], Professor Sandro Oliveira. He has had practical experience of implementing the measurement process by consulting several Brazilian companies on this subject and is a credentialed evaluator, consultant/implementer, and official instructor of process improvement and software product models, such as CMMI, MPS.BR, Certics, Medepros, and QPS. For further details, Figure 2 provides an overview of all the phases followed in this work.

3.3. Search Strategy. There were two main research questions that were raised, one focused on software measurement and the other on software process improvement. Initially, the authors developed only one research question on software measurement. However, it was realized that the return of studies with an emphasis on software measurement was very scarce. Consequently, a second more broader research question was raised on software process improvement, bearing in mind that every software process improvement program uses the software measurement process as a framework. The authors realized that many studies that had no emphasis, but made use of software measurement, and were returned because of the research question on improving software processes. From this point, this research question was included in the systematic review of the literature and both main questions used the PICOC guidelines as a framework that helped to establish the search strings for each main research question.

The research questions that were raised in Section 3.1 were derived from the two main questions, which were arranged in accordance with the framework for Population, Intervention, Context, Outcomes, and Comparison (PICOC), recommended by Kitchenham [35]—with the exception of the comparison criterion which was not used, because the search string encompasses the papers/articles referenced in the other systematic review of the literature found by this study. In addition, the rest of the components of the structure were also used to define the following two main questions, namely,

(1) What is the state-of-the-art of research on the application or teaching of software process improvement (SPI) programs through the use of serious games or gamification?

   (a) *Population (P)*. Software Organizations and Teaching Institutions,

   (b) *Intervention (I)*. Approach used to apply or teach the software improvement process,

   (c) *Context (C)*. This article is aimed at making a comparison between papers/articles which are aimed at both the industrial and the academic sectors,

   (d) *Outcomes (O)*. To capture the dynamics, mechanics, and game components present in the systems discussed and the efficiency in teaching or practical application of the software improvement process when based on gamification or serious games,

   (e) *Comparison (C)*. This does not apply to this study.

(2) What is the state-of-the-art of research on the application or teaching of software measurement by making use of serious games or gamification?

   (a) *Population (P)*. Software Organizations and Teaching Institutions,

   (b) *Intervention (I)*. Approach used to apply or teach the measurement process,

   (c) *Context (C)*. This article is aimed at making a comparison between papers/articles which are aimed at both the industrial and the academic sectors,

   (d) *Outcomes (O)*. To capture the dynamics, mechanics, and game components present in the systems discussed and the efficiency in teaching or practical application of the software measurement process when based on gamification or serious games,

   (e) *Comparison (C)*. This does not apply to this study.

On the basis of the research questions, keywords were obtained in accordance with the framework for: Population, Intervention, Context, and Outcomes for the subsequent formulation of the search string. Here is the list of keywords defined for the first main search question:

(i) *Population (P)*. Project, Development, Organization, Enterprise, Company, Industry, Institute, Research Group, and Technology Center,

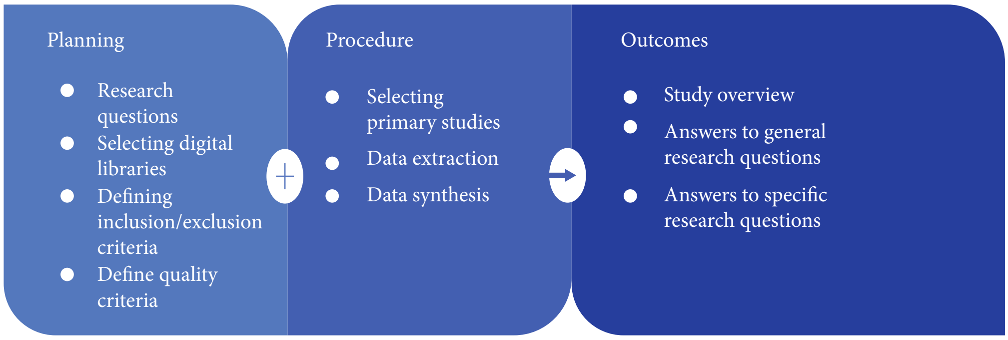

Figure 2: SRL overview [35].

(ii) *Intervention (I)*. Process, Improvement, and SPI,

(iii) *Context (C)*. Learning, Teaching, Education, Training, Practice, and Application,

(iv) *Outcomes (O)*. Gamification, Game, Serious Game, Funware, Game Elements, Game Mechanics, Game Component, Game factor, and Game appearance.

The following keywords were defined for the second main research question:

(i) *Population (P)*. Project, Development, Organization, Enterprise, Company, Industry, Institute, Research Group, and Technology Center,

(ii) *Intervention (I)*. Process, Measuring, Software, Measurement, Metrics, and Metrology,

(iii) *Context (C)*. Learning, Teaching, Education, Training, Practice, and Application,

(iv) *Outcomes (O)*. Gamification, Game, Serious Game, Funware, Game Elements, Game Mechanics, Game Component, Game factor, and Game appearance.

Later, the search string was assembled on the basis of the keywords using the AND and OR connectors, as follows: the AND connector was used to integrate the Population, Intervention, Context, and Outcomes, and the OR connector was used between keywords in the same category. After the search string was designed, it underwent a validation process and was incorporated in the search databases that have the following features: availability of papers/articles in full from queries by the UFPA web domain or when using Google or Google Scholar or Portal CAPES search engines and availability of papers/articles in English or Portuguese and academic libraries that have search engines. Thus, the following databases that comply with these criteria were established: IEEE Xplore, ACM DLL, Science Direct, Scopus, ISI of knowledge (Web of Science), and Ei compendex. Moreover, each database was checked to see whether applying the search returned its control papers/article strings. Previously, the researchers collected the following control papers/articles from the selected search databases [14, 18, 37–43]. Each paper/article was chosen in terms of its relevance to this study. By repeating the validation process of the string, it was possible to arrive at more precise strings for the subject of this research. The following are the final strings:

(1) title-abstr-key(Software AND (Project OR Development OR Organization OR Enterprise OR Academy OR Industry OR Learning OR Teaching OR Education OR Training OR Simulation) AND (Process OR Improvement) AND (Gam* OR Funware OR Ludification))

(2) title-abstr-key(Software AND (Project OR Development OR Organization OR Enterprise OR Academy OR Industry OR Learning OR Teaching OR Education OR Training OR Simulation) AND (Measu* OR Metr*) AND (Gam* OR Funware OR Ludification))

3.4. Study Selection. The scope of the research complies with the restrictions defined in Table 1, to ensure its viability.

Papers/articles were also included in the following areas: experimental studies, experience reports, systematic reviews of the literature, technical reports, bibliographic surveys, systematic study maps, and case studies. In addition, there were papers/articles written in Portuguese and English: the former because it is important to take account of national research, given the relevance of the MPS.BR Program to the study and the latter to broaden the scope of the research, since English is the language set as the standard in most journals and international conferences. Furthermore, the collected papers/articles were all written in the period 2000-2019. The first threshold was set in a way that ensured it was alongside the appearance of the term gamification and the second to be the currently closed year while this research was being conducted.

Additionally, the inclusion and exclusion criteria were employed to analyse the significance of a scientific paper/article while carrying out the systematic review of the literature, and this involved compiling a list of the primary papers/articles and another with the papers/articles that were excluded. The researchers involved in this SRL defined the criteria used in this research, and this is illustrated in Table 2, which shows the inclusion criteria and Table 3, which outlines the exclusion criteria defined for this SRL.

The evaluation of the quality of a paper/article allows works that are closely aligned to the objectives of the

Table 1: Scope of search and restrictions.

Scope	Restrictions
Availability of papers/articles in full through the UFPA domain or from the use of the Google search engine or Google Scholar or Portal CAPES	Papers/articles should mention at least one of the keywords.
Availability of papers/articles in English or Portuguese	Research should not impose a financial burden on researchers.
Search engines for the insertion of keywords	The research must be restricted to the works published between January 1, 2000, and December 31, 2019.

Table 2: Inclusion criteria.

Inclusion criteria
IC-01: papers/articles that provide reports of experience in the industry or academia or experimental or theoretical research, with examples of their application, a description of experiments, or actual cases of teaching approaches that were adopted and the application of software improvement programs or measurement systems through the use of gamification or serious games

Table 3: Exclusion criteria.

Exclusion criteria
EC-01: papers/articles that are not freely available for consultation or download (in full version) to students at the Federal University of Pará;
EC-02: papers/articles that were not found by using the search strings in research databases;
EC-03: papers/articles not included in the listed research databases;
EC-04: papers/articles that violate the principles of some research databases;
EC-05: only the first appearance of a paper/article will be included if the same one is repeated in more than one research database;
EC-06: only the latest or most complete version of a duplicate paper/article will be considered, except in cases where there is supplementary information;
EC-07: works classified as abstracts, keynote speeches, courses, tutorials, and posters;
EC-08: papers/articles that are not included in the teaching syllabus or application of software process improvement programs or measurement programs through the use of gamification or serious game;
EC-09: papers/articles that are not in English or Portuguese;
EC-10: papers/articles that have no relevance to the raised research questions.

projected SRL to make a greater contribution to the research questions. Thus, since the evaluation of the quality of a scientific paper/article is based on an assessment of its significance and content, this evaluative procedure cannot be used as one of the inclusion or exclusion criteria applied to the scientific output during the selection, since it reduces research bias and ensures the internal-external validation [35]. The following are the criteria for assessing the quality of the primary studies, adapted from [44]:

(1) Introduction/planning

 (a) Are the objectives or questions of the study clearly defined? And is the problem addressed in the research clearly described (including the justification for conducting the study)?

 (b) Is the type of study clearly defined?

(2) Development

 (a) Is there a clear description of the context in which the research was conducted?

 (b) Is the work suitably referenced (does it refer to related or similar works and is it based on models and theories in the literature)?

(3) Conclusion

 (a) Does the study support its results in a clear and unambiguous way?

 (b) Have the objectives been achieved and the research questions properly addressed?

(4) Criteria for the research question

 (a) Does the study adopt a primary or secondary approach or make use of a tool for teaching or applying software improvement programs or measurement systems through the use of gamification or serious games?

(5) Specific criteria for experimental studies

 (a) Is there a method or set of methods described in the study?

(6) Specific criteria for theoretical studies

 (a) Is there an unbiased system for choosing studies?

(7) Specific criteria for systematic reviews of the literature

 (a) Is there a strict protocol that has been described and followed?

(8) Specific criterion for industrial experience reporting

(a) Is there a description of the organization(s)/company where the study was conducted?

It should be noted that criteria (1) to (4) are generic, that is, they apply to all the primary studies evaluated, whereas criteria (5) to (8) are specific and correspond to the respective study types mentioned.

The studies in the list that were selected on the basis of the application of the inclusion and exclusion criteria were read in their entirety. When applying the quality criteria, the approach recommended by Costa [44] was adopted, in which the different levels of the Likert-5 scale were used to represent the study's compliance with the quality criteria. These levels are listed below.

(a) *Totally Agree (4)*. This should apply if the work fully meets the requirements of the criteria of the question,

(b) *Partially Agree (3)*. This applies if the work partially meets the criteria of the question,

(c) *Neutral (2)*. This applies if it is not clear whether or not the question has been answered,

(d) *Partially Disagree (1)*. This must apply if the criteria contained in the question are not met by the evaluated work,

(e) *Totally Disagree (0)*. This should apply if there is nothing in the work that meets the criteria of the question.

An evaluation scale is defined for each quality criterion previously established. Table 4 outlines the scale used for each quality criterion.

The two strings were applied to the search engines in the science citation indexes and returned a total of 19050 papers/articles. Scopus database had 30.4%, the largest number of papers/articles returned. Science Direct had 4.8%, IEEE 14.4%, ACM 11.5%, Ei compendex 22.5%, and Web of Science 16.1% of the number of papers/articles returned. Table 5 shows the number of studies returned and the remaining work after each of the criteria were processed; there was no occurrence of exclusion criteria 2, 3, and 4. In Table 5, the sum of the numbers in each row results in the total returned studies of each scientific indexer.

After that, a score was assigned for each paper/article evaluated that was based on the presence of each criterion in the Likert scale and the calculation was made by using the simple Rule of Three, so the papers/articles can be placed in one of the five quality levels defined by Beecham et al. [37] (as shown in Table 6).

An electronic spreadsheet was used to store the data of the papers/articles to answer the bibliometric questions and also calculate the grade (excellent, very good, good, fair, and poor) for the evaluated paper/article. The grade was calculated based on the attributes evaluated in the quality criteria and the Likert-5 scale, which represented the adherence of these attributes to the quality criteria. Table 6 shows the results of the quality evaluation.

The quality criteria were not exclusive, that is, there was no cut index for the evaluated papers/articles, because these papers/articles evaluated by the quality criteria had already passed through the exclusion and inclusion criteria. These quality criteria only served to categorize the writing of the papers/articles and not to exclude them. The quality criteria did not impact the number of accepted papers/articles, given that they did not have the role of excluding them, but rather qualifying them in five different levels, namely, poor, fair, good, very good, and excellent. Thus, all papers/articles were considered to be important for the data extraction from research questions.

The three researchers applied the quality criteria indicated in Table 4 in the 137 primary papers/articles and, whether was a conflict between the quality criteria applied by these different researchers, it was resolved through discussions between them supervised by the advisor of this study. Thus, a single document was generated containing all primary papers/articles qualified through quality criteria. Table 6 summarizes the results achieved by the researchers regarding the quality of the papers/articles. The percentages of this table were reached from the analysis of the complete reading of all papers/articles and the addition of the score according to the analysis of each criterion in Table 4. From the sum of these scores, each paper/article was framed in a quality range, according to Table 6, and a percentage was generated in relation to the total analysed.

As can be seen, few studies are in the poor range and 20 are in the fair range, while 23 studies (16.78%) are in the good range, 46 studies (33.57%) are in the very good range, and 43 studies (31, 38%) in the excellent range. Therefore, the analysed papers/articles present quality above average according to the criteria used. Thus, the quality evaluation criteria were used only to qualify the paper/article and not as an exclusion criterion.

3.5. Classification Study and Data Extraction. This stage involves arranging the data extracted for the display of the charts that provide a general overview and form the basis for future analysis. In addition to the analytical charts of the research questions, the following charts were also generated in response to bibliometric questions: (a) the number of papers/articles returned by the search database (see Figure 3), (b) the number of studies returned per year, (c) the number of studies returned per country, (d) the 5 authors with the highest number of publications, (e) the number of studies per type of study, (f) the number of experimental studies by type, (g) the number of studies returned by publication, (h) the number of studies returned by type of project, (i) the frequency of game elements, and (j) the frequency of game mechanics.

4. Results Achieved

In this section, the results of the systematic review of the literature will be examined. Section 4.1 provides an overview of the selected primary studies and Sections 4.2 to 4.14 describe the results of the research questions.

4.1. Overview. The selection of the studies resulted in a total of 137 primary studies published between 2000 and 2019 (see Table 7). Figure 4 plots a histogram displaying the

Table 4: Likert scale for the quality criteria. Source: adapted from [44].

Criterion	Scale
1a.	4: clearly defines the objectives and justifies the study 3: it clearly defines the study, but the justification is not clear or fails to justify the study. 2: fails to define the objectives, but justifies the study 1: the definition and justification of the study are not clear. 0: fails to define the objectives or justify the study
1b.	4: defines the type of study by referencing the methodology in the literature 3: defines the type of study, but without any reference to the methodology 2: does not define the type of study but can be easily inferred 1: does not define the type of study, but can be inferred with difficulty 0: it is not possible to infer the type of study.
2a.	4: clearly defines a section with the context of the research 3: the research context is included in a nonexclusive section. 2: the context of the research is dispersed throughout the text. 1: the context of the research is dispersed and is insubstantial. 0: the context of the research is not addressed.
2b.	4: the text includes a section of related works. 3: the text includes related works in a nonexclusive section. 2: the text includes related works scattered throughout the text. 1: the text does not include any related works, but is based on the literature. 0: the text does not include any related works nor is it based on the literature.
3a.	4: results are clearly shown in the conclusion section. 3: results are clearly referenced in the conclusion section. 2: results that appear in the conclusion are not clear. 1: results referenced in the conclusion are not clear. 0: no results found.
3b.	4: the results are completely pertinent to the objectives of the study. 3: the results are pertinent to the objectives of the study, although the author adds some subjective comments. 2: the results have some bearing on the objectives of the study. 1: the results are not pertinent to the objectives of the study. 0: no results were found.
4a.	4: a tool or approach is described in detail and validated with users. 3: a tool or approach is described in a detailed way, but not validated. 2: there is a tool or approach, but it is not described in detail. However, it is validated with users. 1: there is a tool or approach, but it is not described in detail or validated. 0: there is only a reference to a tool or no tool or approach is described.
4b.	4: the experimental method is defined and clearly referenced. 3: the experimental method is clearly defined. 2: the experimental method is referenced. 1: the experimental method is not referenced, but it can be inferred. 0: it is not possible to infer an experimental method.
5a.	4: the text describes the criteria for choosing the studies. 3: the text does not describe the criteria for choosing the studies, but includes material that contradicts the study being carried out. 2: the text only describes studies pertinent to the study being carried out. 1: the text does not describe a sufficient number of studies. 0: the text does not describe studies in the database.

Table 4: Continued.

Criterion	Scale
6a.	4: the review protocol is outlined, described, and followed. 3: the review protocol is outlined and described, but there is evidence that it was not followed adequately. 2: the review protocol has not been sufficiently described. 1: the review protocol was only cited at different stages in the text. 0: there is no review protocol.
7a.	4: there is information about the area of action, size, and origin of the organization. 3: there is only information about two of the features. 2: there is only information about one of the features. 1: there is no information about any of the features. 0: the study was not conducted in one or more organizations.
8a.	4: it clearly defines the objectives and justifies the study. 3: it clearly defines the study, but the justification is not clear or the study is not justified. 2: it fails to define the objectives, but justifies the study. 1: the definition and justification of the study are not clear. 0: it fails to define the objectives and does not justify the study.

Table 5: Studies returned and remaining studies through the stages of SRL (the table follows the S|M standard, in which S stands for software process improvement papers/articles and M for software measurement papers/articles).

		Primary studies							
Sources	Returned papers/articles	EC.1	EC.5	EC.6	EC.7	EC.8	EC.9	EC.10	IC.1
ACM	1747|461	0|0	24|25	0|2	10|3	1668|425	0|1	15|2	20|3
IEEE Xplore	2073|675	0|0	0|64	3|0	8|1	1903|595	5|0	86|9	36|6
Science Direct	555|362	0|0	268|198	0|0	5|2	261|162	0|0	12|0	9|0
Ei compendex	2583|1714	9|0	720|843	0|0	1|1	1798|865	0|0	9|0	16|5
Web of Knowledge	2098|986	10|1	131|131	2|0	1|0	1921|850	0|0	14|4	29|0
Scopus	3828|1968	17|3	459|302	0|0	42|29	3286|1624	0|0	15|6	9|4
Total	12884|6166	36|4	1602|1563	5|2	67|36	10837|4521	5|1	151|21	119|18

Table 6: Quality levels. Source: [37].

Grade range	Evaluation	Number of primary papers/articles	%
Excellent	>86%	43	31.38%
Very good	66%-85%	46	33.57%
Good	46%-65%	23	16.78%
Fair	26%-45%	20	14.59%
Poor	<26%	5	3.64%

frequency of primary studies per year, with the different colours representing the related papers/articles about serious games and gamification, and it shows a growing pattern until the year 2016 in the use of games and gamification for teaching.

Figure 4 shows a decrease in the use of games and gamification for teaching (2016-2019), but the authors cannot confirm with precision the reason for this event. It is possible to assume that the field has already reached a certain level of maturity and consequently had a reduction in the novelty factor due to already having a range of studies exploring the topic.

With regard to the distribution of papers/articles by type of publication, it was found that most of the primary studies (i.e., 73%) were published in conferences, 5% in workshops, and only 22% in journals, as shown by the chart in Figure 5. The authors consider conferences and workshops as two different events, because some papers point to workshops as a publication venue, for instance, the paper entitled "HALO (Highly Addictive, Socially Optimized) Software Engineering" that came with its DOI linked to the "Proceedings of the 1st International Workshop on Games and Software Engineering."

The systematic review of the literature is a method used to highlight trends. In spite of this, this work does not indicate a reason that justifies the conferences and workshops as the main means of publishing the studies of gamification and serious games analysed in this SRL. The authors believe that conferences and workshops have three main advantages over other venues, namely, (a) speed of publication: it usually took only a few months to have their work published in a conference or workshop, unlike journals that have a much longer time, and it can be from quarters to semesters to publish the same article; (b) full papers are generally papers of 8

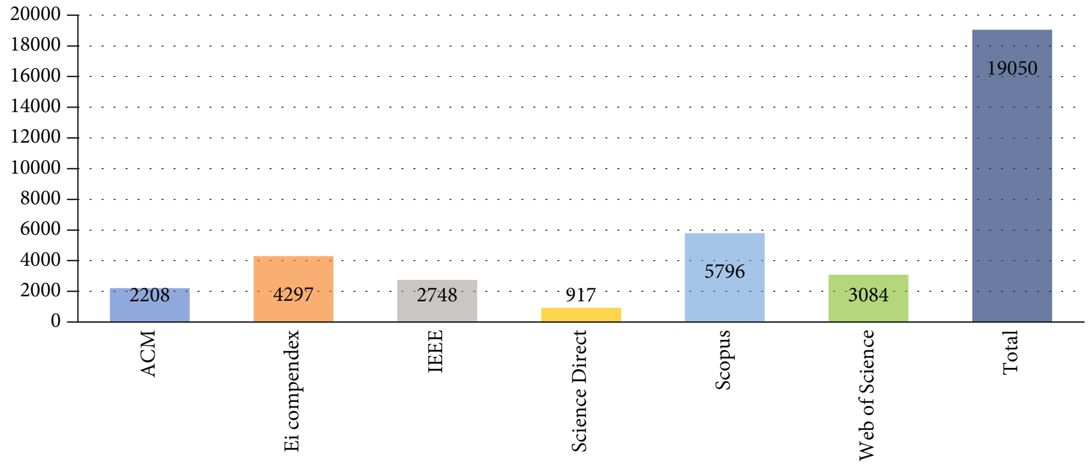

Figure 3: Number of studies returned by the search database.

or more pages that report the results of a research, unlike workshops that generally feature training, dynamics, or short papers of up to 4 pages to describe a work in progress; (c) network, which makes it possible to exchange knowledge with other researchers in the same area during the conference or workshop. Therefore, these characteristics can make it possible for conferences or workshops to be the main publication venue.

In addition, Table 8 shows the five conferences, journals, and workshops that have had more primary studies published.

The main contribution of Table 8 is to highlight the main publication venues for researchers who are working on the theme proposed by this SRL. In addition, the conferences, journals, and workshops allow inferring about the maturity of the researched field, that is, the higher the quality of the publication venue, the greater the maturity of the study topic.

In addition, the five universities, five authors, and the five countries that have had most published papers/articles between the primary studies were also analysed. The adopted method for counting authors, universities, and countries was the same that it was to account for all authors, universities, and countries present in the study using a spreadsheet that catalogued all metadata of selected studies, for instance, in the study that contained two or more countries all the countries involved were counted. The same occurs to authors and universities. This information can be seen in Table 9.

The bibliometric data, defined in Table 9, have the importance of showing a worldwide panorama of the research on the theme established by the SRL. Knowing the main authors, universities, and countries behind the advancement of the research field allows researchers to find possible mentoring or cooperation in research this field. In other words, it enables the creation of a network for the exchange of knowledge and the improvement of ideas related to the researched field.

4.2. GQ1: In Which Contexts (i.e., Academic or Professional) Were the Gamification or Serious Game Projects Applied? This question was addressed within two domains: academic and professional. In the academic domain, students are the target audience. In contrast, projects that took place in a professional context were applied in software organizations. When the research questions are answered, a code (P + number of study) will be used that represents each paper/article that is listed in Table 7. A total of 84 primary studies referred to the academic context. Projects are generally aimed at teaching some process or subject related to software process improvement programs, for instance, P3, P5, P9, P10, P11, P12, P13, P14, P16, P17, P18, P21, P23, P24, P25, P30, P32, P33, P34, P36, P37, and P39.

In the case of the professional context, the approaches fluctuated between (a) schemes that were aimed at encouraging patterns of behaviour within the organization, (b) adopting processes (P1, P2, P4, P6, P7, and P8), (c) making measurements in the development process and for their teams (P19 and P22), and (d) acculturation (P29, P31, and P35). Only paper P126 was applied within a combined academic and professional context, and this had a gamified tool to assist in the code review. Only P27 did not state what kind of environment it was in. Table 10 shows the percentage of papers/articles for each context and their code.

The SRL is a secondary research method, which is the synthesis of information and data on studies aligned with the theme of SRL collected in the selected databases. It is well known that the industrial or professional environment does not report all its practices and approaches used to have its competitive advantage. Consequently, it is plausible to have more studies which are aimed at the academic environment instead of studies with industrial reports.

4.3. GQ2: What Are the Limitations Reported in the Use of Gamification or Serious Games for Teaching? About 22% of the total number of papers/articles related to gamification reported some limitations. Among these, many referred to the need for some improvements to be made in the methodology employed in the work, but these were disregarded in this analysis. Others pointed out limitations that were only found in the gamified project, and these are listed below.

The primary study P1 included an interactive questionnaire as a personality assessment tool, which was specifically designed for software engineers. It was stated that there is a need to improve the aesthetics of the project, both in graphics as the sounds used because it was found that these areas were

Table 7: Primary papers/articles and their respective code (ID) and year.

ID	Primary studies
P1	Yilmaz, M., Yilmaz, M., O'Connor, R. V., & Clarke, P. (2016, June). A gamification approach to improve the software development process by exploring the personality of software practitioners. In International Conference on Software Process Improvement and Capability Determination (pp. 71-83). Springer, Cham.
P2	Ruiz, M., Trinidad, M., & Calderón, A. (2016, November). Gamification and functional prototyping to support motivation towards software process improvement. In International Conference on Product-Focused Software Process Improvement (pp. 697-704). Springer, Cham.
P3	Uyaguari, F. U., Intriago, M., & Jácome, E. S. (2015). Gamification proposal for a software engineering risk management course. In New Contributions in Information Systems and Technologies (pp. 795-802). Springer, Cham.
P4	Unkelos-Shpigel, N., & Hadar, I. (2015). Gamifying Software Development Environments uUsing Cognitive Principles. In CAiSE forum (pp. 9-16).
P5	Aydan, U., Yilmaz, M., & O'Connor, R. V. (2015, June). Towards a serious game to teach ISO/IEC 12207 software lifecycle process: an interactive learning approach. In International Conference on Software Process Improvement and Capability Determination (pp. 217-229). Springer, Cham.
P6	Morales-Trujillo, M. E., Oktaba, H., & González, J. C. (2014, April). Taking Seriously Software Projects Inception through Games. In International Conference on Evaluation of Novel Approaches to Software Engineering (pp. 109-124). Springer, Cham.
P7	Herranz, E., Colomo-Palacios, R., & de Amescua Seco, A. (2015, September). Gamiware: a gamification platform for software process improvement. In European Conference on Software Process Improvement (pp. 127-139). Springer, Cham.
P8	Jovanovic, M., Mesquida, A. L., & Mas, A. (2015, September). Process improvement with retrospective gaming in agile software development. In European Conference on Software Process Improvement (pp. 287-294). Springer, Cham.
P9	Kosa, M., & Yilmaz, M. (2015, September). Designing games for improving the software development process. In European Conference on Software Process Improvement (pp. 303-310). Springer, Cham.
P10	Manrique-Losada, B., Gasca-Hurtado, G. P., & Gómez Álvarez, M. C. (2015). Assessment proposal of teaching and learning strategies in software process improvement. Revista Facultad De Ingeniería Universidad De Antioquia, (77), 105-114.
P11	Ganesh, L. (2014, December). Board game as a tool to teach software engineering concept–technical debt. In 2014 IEEE Sixth International Conference on Technology for Education (pp. 44-47). IEEE.
P12	Oliveira, C., Cintra, M., & Neto, F. M. (2013, August). Learning risk management in software projects with a serious game based on intelligent agents and fuzzy systems. In 8th conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-13) (pp. 874-879). Atlantis Press.
P13	Galvão, T. A. B., Neto, F. M. M., Bonates, M. F., & Campos, M. T. (2011, July). A serious game for supporting training in risk management through project-based learning. In International Conference on Virtual and Networked Organizations, Emergent Technologies, and Tools (pp. 52-61). Springer, Berlin, Heidelberg.
P14	Galvão, T. A. B., Neto, F. M. M., Campos, M. T., & Júnior, E. D. L. C. (2012). An approach to assess knowledge and skills in risk management through project-based learning. International Journal of Distance Education Technologies (IJDET), 10(3), 17-34.
P15	Wautelet, Y., Kolp, M., & Neysen, N. (2012). E-SPM: an online software project management game. International Journal of Engineering Education, 28(6), 1316.
P16	Von Wangenheim, C. G., Thiry, M., & Kochanski, D. (2009). Empirical evaluation of an educational game on software measurement. Empirical Software Engineering, 14(4), 418-452 [43].
P17	Navarro, E. O., & Van Der Hoek, A. (2005). Software process modeling for an educational software engineering simulation game. Software Process: Improvement and Practice, 10(3), 311-325.
P18	Navarro, E. O., & van der Hoek, A. (2004, August). SIMSE: An Interactive Simulation Game for Software Engineering Education. In CATE (pp. 12-17).
P19	Cummins, D. (2004, June). Using competition to build a stronger team. In Agile Development Conference (pp. 137-141). IEEE.
P20	Häberlein, T., & Gantner, T. (2002, October). Process-Oriented Interactive Simulation of Software Acquisition Projects. In Eurasian Conference on Information and Communication Technology (pp. 806-815). Springer, Berlin, Heidelberg.
P21	Su, C. H. (2016). The effects of students' motivation, cognitive load and learning anxiety in gamification software engineering education: a structural equation modeling study. Multimedia Tools and Applications, 75(16), 10013-10036.

Table 7: Continued.

ID	Primary studies
P22	de Melo, A. A., Hinz, M., Scheibel, G., Berkenbrock, C. D. M., Gasparini, I., & Baldo, F. (2014, June). Version control system gamification: a proposal to encourage the engagement of developers to collaborate in software projects. In International Conference on Social Computing and Social Media (pp. 550-558). Springer, Cham.
P23	Papaloukas, S., Patriarcheas, K., & Xenos, M. (2011, October). Games' Usability and Learning-the Educational Videogame BeTheManager. In Proceedings of the 5th European Conference on Games Based Learning (pp. 449-456).
P24	Panetta, K., Dornbush, C., & Loomis, C. (2002). A collaborative learning methodology for enhanced comprehension using TEAMThink®. Journal of Engineering Education, 91(2), 223-229.
P25	Schriek, C., van der Werf, J. M. E., Tang, A., & Bex, F. (2016, November). Software architecture design reasoning: a card game to help novice designers. In European conference on software architecture (pp. 22-38). Springer, Cham.
P26	Sanchez-Gordón, M. L., Colomo-Palacios, R., & Herranz, E. (2016, September). Gamification and human factors in quality management systems: mapping from octalysis framework to ISO 10018. In European Conference on Software Process Improvement (pp. 234-241). Springer, Cham.
P27	Jovanović, M., Mesquida, A. L., Radaković, N., & Mas, A. (2016). Agile retrospective games for different team development phases. Journal of Universal Computer Science, 22(12), 1489-1508.
P28	Herranz, E., Palacios, R. C., de Amescua Seco, A., & Sánchez-Gordón, M. L. (2016). Towards a Gamification Framework for Software Process Improvement Initiatives: Construction and Validation. J. UCS, 22(12), 1509-1532.
P29	Medeiros, D. B., Neto, P. D. A. D. S., Passos, E. B., & De Souza Araújo, W. (2015). Working and playing with scrum. International Journal of Software Engineering and Knowledge Engineering, 25(06), 993-1015.
P30	Aguilar, R. A., Aké, I. E., & Ucán, J. P. Developing virtual learning environments for software enginnering education: a ludic proposal.
P31	Herranz, E., Palacios, R. C., de Amescua Seco, A., & Yilmaz, M. (2014). Gamification as a Disruptive Factor in Software Process Improvement Initiatives. J. UCS, 20(6), 885-906.
P32	Ramingwong, S. (2012). CutIT: a game for teaching process improvement in software engineering. In Proceeding of the Third International Conference on Information, Communication and Education Application (ICEA 2012).
P33	Von Wangenheim, C. G., Savi, R., & Borgatto, A. F. (2013). SCRUMIA—An educational game for teaching SCRUM in computing courses. Journal of Systems and Software, 86(10), 2675-2687.
P34	Vega, L., Cancela, N., Quintero, R., & Zepeda, L. (2013, August). Game-based ICT Project Formulation based on Agile Approach for Skills Development. In 2013 International Conference on Advanced ICT and Education (ICAICTE-13). Atlantis Press.
P35	Dorling, A., & McCaffery, F. (2012, May). The gamification of SPICE. In International Conference on Software Process Improvement and Capability Determination (pp. 295-301). Springer, Berlin, Heidelberg.
P36	Zapata Jaramillo, C. M. (2010). Communication and traceability game: a way to improve requirements elicitation process teaching. Revista Facultad de Ingeniería Universidad de Antioquia, (56), 213-221.
P37	Monsalve, E. S., Pereira, A. X., & Werneck, V. M. B. (2014). Teaching Software Engineering through a Collaborative Game. In Overcoming Challenges in Software Engineering Education: Delivering Non-Technical Knowledge and Skills (pp. 310-331). IGI global.
P38	Erfurth, I., & Kirchner, K. (2010, March). Requirements elicitation with adapted cuta cards: first experiences with business process analysis. In 2010 15th IEEE International Conference on Engineering of Complex Computer Systems (pp. 215-223). IEEE.
P39	Gamble, R. F., & Smith, M. L. (2008, October). Moving toward "reality" in team selection for software engineering. In 2008 38th Annual Frontiers in Education Conference (pp. F3H-21). IEEE.
P40	Yilmaz, M., & O'Connor, R. V. (2016). A Scrumban integrated gamification approach to guide software process improvement: a Turkish case study. Tehnički vjesnik, 23(1), 237-245.
P41	Silva, C. G., Farias, I. J., Teixeira M., Aquino C. (2016). gTest Learning: a game for learning basic software testing. Congresso Regional sobre tecnologias na Educação, 450-460.
P42	Sánchez-Gordón, M. L., O'Connor, R. V., Colomo-Palacios, R., & Sanchez-Gordon, S. (2016, June). A learning tool for the ISO/IEC 29110 standard: understanding the project management of basic profile. In International Conference on Software Process Improvement and Capability Determination (pp. 270-283). Springer, Cham.
P43	Calderón, A., & Ruiz, M. (2016, June). Coverage of ISO/IEC 12207 software lifecycle process by a simulation-based serious game. In International Conference on Software Process Improvement and Capability Determination (pp. 59-70). Springer, Cham.

Table 7: Continued.

ID	Primary studies
P44	Heikkilä, V. T., Paasivaara, M., & Lassenius, C. (2016, May). Teaching university students Kanban with a collaborative board game. In Proceedings of the 38th international conference on software engineering companion (pp. 471-480).
P45	Williams, L., Meneely, A., & Shipley, G. (2010). Protection poker: the new software security" game". IEEE Security & Privacy, 8(3), 14-20.
P46	Vega, K., Soares, A. P., Robichez, G., & Fuks, H. (2010, October). TREG Usability Tests: Evaluating a Training Game in Second Life. In 2010 Brazilian Symposium on Collaborative Systems-Simposio Brasileiro de Sistemas Colaborativos (pp. 63-70). IEEE.
P47	J. B. Hauge, H. Duin and K. Thoben, "Applying serious games for supporting idea generation in collaborative innovation processes," 2008 IEEE International Technology Management Conference (ICE), Lisbon, Portugal, 2008, pp. 1-8.
P48	Srinivasan, J., & Lundqvist, K. (2007, may). A constructivist approach to teaching software processes. In 29th International Conference on Software Engineering (ICSE'07) (pp. 664-672). IEEE.
P49	Carrington, D., Baker, A., & van der Hoek, A. (2005, October). It's all in the game: teaching software process concepts. In Proceedings Frontiers in Education 35th Annual Conference (pp. F4G-F4G). IEEE.
P50	Navarro, E. O., & Van Der Hoek, A. (2005, April). Design and evaluation of an educational software process simulation environment and associated model. In 18th Conference on Software Engineering Education & Training (CSEET'05) (pp. 25-32). IEEE.
P51	Baker, A., Navarro, E. O., & Van Der Hoek, A. (2003, May). Problems and Programmers: an educational software engineering card game. In 25th International Conference on Software Engineering, 2003. Proceedings. (pp. 614-619). IEEE.
P52	Alsaedi, O., Toups, Z., & cook, J. (2016, May). Can a team coordination game help student software project teams?. In Proceedings of the 9th International Workshop on Cooperative and Human Aspects of Software Engineering (pp. 33-39).
P53	Elezi, L., Sali, S., Demeyer, S., Murgia, A., & Pérez, J. (2016, May). A game of refactoring: studying the impact of gamification in software refactoring. In Proceedings of the Scientific Workshop Proceedings of XP2016 (pp. 1-6).
P54	Elm, D., Tondello, G. F., Kappen, D. L., Ganaba, M., Stocco, M., & Nacke, L. E. (2016, October). CLEVER: a trivia and strategy game for enterprise knowledge learning. In Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts (pp. 61-66).
P55	Mora, A., Planas, E., & Arnedo-Moreno, J. (2016, November). Designing game-like activities to engage adult learners in higher education. In Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality (pp. 755-762).
P56	Frtala, T., & Vranic, V. (2015, May). Animating organizational patterns. In 2015 IEEE/ACM 8th International Workshop on Cooperative and Human Aspects of Software Engineering (pp. 8-14). IEEE.
P57	Jurado, J. L., Fernandez, A., & Collazos, C. A. (2015, October). Applying gamification in the context of knowledge management. In Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business (pp. 1-4).
P58	Prause, C. R., & Jarke, M. (2015, August). Gamification for enforcing coding conventions. In Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering (pp. 649-660).
P59	Paasivaara, M., Heikkilä, V., Lassenius, C., & Toivola, T. (2014, May). Teaching students scrum using LEGO blocks. In Companion Proceedings of the 36th International Conference on Software Engineering (pp. 382-391).
P60	Snipes, W., Nair, A. R., & Murphy-Hill, E. (2014, May). Experiences gamifying developer adoption of practices and tools. In Companion Proceedings of the 36th International Conference on Software Engineering (pp. 105-114).
P61	Sukale, R., & Pfaff, M. S. (2014). QuoDocs: improving developer engagement in software documentation through gamification. In CHI'14 Extended Abstracts on Human Factors in Computing Systems (pp. 1531-1536).
P62	Costa, J. P., Wehbe, R. R., Robb, J., & Nacke, L. E. (2013, October). Time's up: studying leaderboards for engaging punctual behaviour. In Proceedings of the First International Conference on Gameful Design, Research, and Applications (pp. 26-33).
P63	Dubois, D. J., & Tamburrelli, G. (2013, August). Understanding gamification mechanisms for software development. In Proceedings of the 2013 9th Joint Meeting on Foundations of Software Engineering (pp. 659-662).
P64	Scharlau, B. A. (2013, July). Games for teaching software development. In Proceedings of the 18th ACM Conference on Innovation and Technology in Computer Science Education (pp. 303-308).
P65	Lotufo, R., Passos, L., & Czarnecki, K. (2012, June). Towards improving bug tracking systems with game mechanisms. In 2012 9th IEEE Working Conference on Mining Software Repositories (MSR) (pp. 2-11). IEEE.

Table 7: Continued.

ID	Primary studies
P66	Sheth, S., Bell, J., & Kaiser, G. (2011, May). Halo (highly addictive, socially optimized) software engineering. In Proceedings of the 1st International Workshop on Games and Software Engineering (pp. 29-32).
P67	Erfurth, I., & Rossak, W. (2010, May). CUTA4UML: bridging the gap between informal and formal requirements for dynamic system aspects. In Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering-Volume 2 (pp. 171-174).
P68	Shaw, K., & Dermoudy, J. (2005, January). Engendering an empathy for software engineering. In Proceedings of the 7th Australasian Conference on Computing Education-Volume 42 (pp. 135-144).
P69	Adachi, H., Myojin, S., & Shimada, N. (2015). ScoringTalk: a tablet system scoring and visualizing conversation for balancing of participation. In SIGGRAPH Asia 2015 Mobile Graphics and Interactive Applications (pp. 1-5).
P70	Aydan, U., Yilmaz, M., Clarke, P. M., & O'Connor, R. V. (2017). Teaching ISO/IEC 12207 software lifecycle processes: a serious game approach. Computer Standards & Interfaces, 54, 129-138.
P71	Flores, N. H., Paiva, A. C., & Letra, P. (2016). Software engineering management education through game design patterns. Procedia-Social and Behavioral Sciences, 228, 436-442.
P72	Wu, W. H., Yan, W. C., Kao, H. Y., Wang, W. Y., & Wu, Y. C. J. (2016). Integration of RPG use and ELC foundation to examine students' learning for practice. Computers in Human Behavior, 55, 1179-1184.
P73	Herranz, E., Colomo-Palacios, R., & Amescua-Seco, A. (2013). Towards a new approach to supporting top managers in SPI organizational change management. Procedia Technology, 9, 129-138.
P74	Baker, A., Navarro, E. O., & Van Der Hoek, A. (2005). An experimental card game for teaching software engineering processes. Journal of Systems and Software, 75(1-2), 3-16.
P75	Bartel, A., & Hagel, G. (2016, April). Gamifying the learning of design patterns in software engineering education. In 2016 IEEE Global Engineering Education Conference (EDUCON) (pp. 74-79). IEEE.
P76	Sharma, V. S., & Kaulgud, V. (2016, August). Agile workbench: tying people, process, and tools in distributed agile delivery. In 2016 IEEE 11th International Conference on Global Software Engineering (ICGSE) (pp. 69-73). IEEE.
P77	Valencia, D., Vizcaíno, A., Garcia-Mundo, L., Piattini, M., & Soto, J. P. (2016, August). GSDgame: a serious game for the acquisition of the competencies needed in GSD. In 2016 IEEE 11th International Conference on Global Software Engineering Workshops (ICGSEW) (pp. 19-24). IEEE.
P78	Przybylek, A., & Olszewski, M. K. (2016, September). Adopting collaborative games into Open Kanban. In 2016 Federated Conference on Computer Science and Information Systems (FedCSIS) (pp. 1539-1543). IEEE.
P79	Maxim, B. R., Kaur, R., Apzynski, C., Edwards, D., & Evans, E. (2016, October). An agile software engineering process improvement game. In 2016 IEEE Frontiers in education Conference (FIE) (pp. 1-4). IEEE.
P80	Piras, L., Giorgini, P., & Mylopoulos, J. (2016, September). Acceptance requirements and their gamification solutions. In 2016 IEEE 24th International Requirements Engineering Conference (RE) (pp. 365-370). IEEE.
P81	Sharma, V. S., Kaulgud, V., & Duraisamy, P. (2016, May). A gamification approach for distributed agile delivery. In Proceedings of the 5th International Workshop on Games and Software Engineering (pp. 42-45).
P82	Oliveira, B., Afonso, P., & Costa, H. (2016, October). TestEG—a computational game for teaching of software testing. In 2016 35th International Conference of the Chilean Computer Science Society (SCCC) (pp. 1-10). IEEE.
P83	Chaves, R. O., von Wangenheim, C. G., Furtado, J. C. C., Oliveira, S. R. B., Santos, A., & Favero, E. L. (2015). Experimental evaluation of a serious game for teaching software process modeling. ieee Transactions on Education, 58(4), 289-296.
P84	Pieper, J. (2015, March). Discovering the essence of Software Engineering an integrated game-based approach based on the SEMAT Essence specification. In 2015 IEEE Global Engineering Education Conference (EDUCON) (pp. 939-947). IEEE.
P85	Monsalve, E. S., do Prado Leite, J. C. S., & Werneck, V. M. B. (2015, May). Transparently teaching in the context of game-based learning: the case of simulES-W. In 2015 IEEE/ACM 37th IEEE International Conference on Software Engineering (Vol. 2, pp. 343-352). IEEE.
P86	Sonchan, P., & Ramingwong, S. (2015, June). ARMI 2.0: an online risk management simulation. In 2015 12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON) (pp. 1-5). IEEE.
P87	Letra, P., Paiva, A. C. R., & Flores, N. (2015, October). Game design techniques for software engineering management education. In 2015 IEEE 18th International Conference on Computational Science and Engineering (pp. 192-199). IEEE.

Table 7: Continued.

ID	Primary studies
P88	Pötter, H., Schots, M., Duboc, L., & Werneck, V. (2014, April). InspectorX: a game for software inspection training and learning. In 2014 IEEE 27th Conference on Software Engineering Education and Training (CSEE&T) (pp. 55-64). IEEE.
P89	Akpolat, B. S., & Slany, W. (2014, April). Enhancing software engineering student team engagement in a high-intensity extreme programming course using gamification. In 2014 ieee 27th conference on software engineering education and training (csee&t) (pp. 149-153). IEEE.
P90	Saito, D., Takebayashi, A., & Yamaura, T. (2014, October). Minecraft-based preparatory training for software development project. In 2014 IEEE International Professional Communication Conference (IPCC) (pp. 1-9). IEEE.
P91	Snipes, W., Augustine, V., Nair, A. R., & Murphy-Hill, E. (2013, May). Towards recognizing and rewarding efficient developer work patterns. In 2013 35th International Conference on Software Engineering (ICSE) (pp. 1277-1280). IEEE.
P92	Potineni, S., Bansal, S. K., & Amresh, A. (2013, August). ScrumTutor: a web-based interactive tutorial for Scrum Software development. In 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 1884-1890). IEEE.
P93	Peixoto, D. C., Resende, R. F., & Padua, C. I. P. (2013, October). An educational simulation model derived from academic and industrial experiences. In 2013 IEEE Frontiers in Education Conference (FIE) (pp. 691-697). IEEE.
P94	Singer, L., & Schneider, K. (2012, June). It was a bit of a race: gamification of version control. In 2012 Second International Workshop on Games and Software Engineering: Realizing User Engagement with Game Engineering Techniques (GAS) (pp. 5-8). IEEE.
P95	Zuppiroli, S., Ciancarini, P., & Gabbrielli, M. (2012, April). A role-playing game for a software engineering lab: developing a product line. In 2012 IEEE 25th Conference on Software Engineering Education and Training (pp. 13-22). IEEE.
P96	Lima, T., Campos, B., Santos, R., & Werner, C. (2012, October). UbiRE: a game for teaching requirements in the context of ubiquitous systems. In 2012 XXXVIII Conferencia Latinoamericana En Informatica (CLEI) (pp. 1-10). IEEE.
P97	Rusu, A., Russell, R., & Cocco, R. (2011, July). Simulating the software engineering interview process using a decision-based serious computer game. In 2011 16th International Conference on Computer Games (CGAMES) (pp. 235-239). IEEE.
P98	Lynch, T. D., Herold, M., Bolinger, J., Deshpande, S., Bihari, T., Ramanathan, J., & Ramnath, R. (2011, October). An agile boot camp: using a LEGO®-based active game to ground agile development principles. In 2011 Frontiers in Education Conference (FIE) (pp. F1H-1). IEEE.
P99	Passos, E. B., Medeiros, D. B., Neto, P. A., & Clua, E. W. (2011, November). Turning real-world software development into a game. In 2011 Brazilian Symposium on Games and Digital Entertainment (pp. 260-269). IEEE.
P100	Fernandes, J. M., & Sousa, S. M. (2010, March). Playscrum-a card game to learn the scrum agile method. In 2010 Second International Conference on Games and Virtual Worlds for Serious Applications (pp. 52-59). IEEE.
P101	Nerbråten, Ø., & Røstad, L. (2009, March). HACMEgame: a tool for teaching software security. In 2009 International Conference on Availability, Reliability and Security (pp. 811-816). IEEE.
P102	Beatty, J., & Alexander, M. (2008, September). Games-based requirements engineering training: an initial experience report. In 2008 16th IEEE International Requirements Engineering Conference (pp. 211-216). IEEE.
P103	Bachtiar, F. A., Pradana, F., Priyambadha, B., & Bastari, D. I. (2018, July). CoMa: Development of Gamification-based E-learning. In 2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE) (pp. 1-6). IEEE.
P104	Mi, Q., Keung, J., Mei, X., Xiao, Y., & Chan, W. K. (2018, July). A gamification technique for motivating students to learn code readability in software engineering. In 2018 International Symposium on Educational Technology (ISET) (pp. 250-254). IEEE.
P105	Stettina, C. J., Offerman, T., De Mooij, B., & Sidhu, I. (2018, June). Gaming for agility: using serious games to enable agile project & portfolio management capabilities in practice. In 2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC) (pp. 1-9). IEEE.
P106	Kemell, K. K., Risku, J., Evensen, A., Abraharnsson, P., Dahl, A. M., Grytten, L. H., ... & Nguyen-Duc, A. (2018, June). Gamifying the Escape from the Engineering Method Prison. In 2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC) (pp. 1-9). IEEE.
P107	Marques, R., Costa, G., da Silva, M. M., & Gonçalves, P. (2017, September). Gamifying software development scrum projects. In 2017 9th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games) (pp. 141-144). IEEE.
P108	Frącz, W., & Dajda, J. (2018, September). Developers' Game: A Preliminary Study Concerning a Tool for Automated Developers Assessment. In 2018 IEEE International Conference on Software Maintenance and Evolution (ICSME) (pp. 695-699). IEEE.

Table 7: Continued.

ID	Primary studies
P109	Toscani, C., Gery, D., Steinmacher, I., & Marczak, S. (2018, October). A gamification proposal to support the onboarding of newcomers in the flosscoach portal. In Proceedings of the 17th Brazilian Symposium on Human Factors in Computing Systems (pp. 1-10).
P110	Molléri, J. S., Gonzalez-Huerta, J., & Henningsson, K. (2018, June). A legacy game for project management in software engineering courses. In Proceedings of the 3rd European Conference of Software Engineering Education (pp. 72-76).
P111	Pimentel, J., Santos, E., Pereira, T., Ferreira, D., & Castro, J. (2018, April). A gamified requirements inspection process for goal models. In Proceedings of the 33rd Annual ACM Symposium on Applied Computing (pp. 1300-1307).
P112	Chow, I., & Huang, L. (2017, January). A software gamification model for cross-cultural software development teams. In Proceedings of the 2017 international conference on management engineering, software engineering and service sciences (pp. 1-8).
P113	Khandelwal, S., Sripada, S. K., & Reddy, Y. R. (2017, February). Impact of gamification on code review process: an experimental study. In Proceedings of the 10th innovations in software engineering conference (pp. 122-126).
P114	Kuo, J. H., Wu, T. H., Ye, H. B., & Jiau, H. C. (2018, October). A competitive platform for continuous programming skill enhancement. In Proceedings of the 28th Annual International Conference on Computer Science and Software Engineering (pp. 30-39).
P115	Fernandez-Reyes, K., Clarke, D., & Hornbach, J. (2018, October). The impact of opt-in gamification on students' grades in a software design course. In Proceedings of the 21st ACM/IEEE International Conference on Model Driven Engineering Languages and Systems: Companion Proceedings (pp. 90-97).
P116	Calderón, A., Ruiz, M., & O'Connor, R. V. (2018). A serious game to support the ISO 21500 standard education in the context of software project management. Computer standards & interfaces, 60, 80-92.
P117	García, F., Pedreira, O., Piattini, M., Cerdeira-Pena, A., & Penabad, M. (2017). A framework for gamification in software engineering. Journal of Systems and Software, 132, 21-40.
P118	Mesquida, A. L., & Mas, A. (2018). Experiences on the use of a game for improving learning and assessing knowledge. Computer Applications in Engineering Education, 26(6), 2058-2070.
P119	Suescún Monsalve, E., Toro, M., Mazo, R., Velasquez, D., Vallejo, P., Cardona, J. F., ... & Leite, J. C. S. D. P. (2018). SimulES-W: a collaborative game to improve software engineering teaching. Computación y Sistemas, 22(3), 953-983.
P120	Furtado, J. C. C., & Oliveira, S. R. B. (2018, March). A Methodology to Teaching Statistical Process Control in Computer Courses. In ENASE (pp. 424-431).
P121	Aydan, U., Yilmaz, M., Clarke, P. M., & O'Connor, R. V. (2017). Teaching ISO/IEC 12207 software lifecycle processes: a serious game approach. Computer Standards & Interfaces, 54, 129-138.
P122	Herranz, E., & Colomo-Palacios, R. (2018, September). Is gamification a way to a softer software process improvement? A preliminary study of success factors. In European Conference on Software Process Improvement (pp. 207-218). Springer, Cham.
P123	Gañán, D., Caballé, S., Clarisó, R., Conesa, J., & Bañeres, D. (2017). ICT-FLAG: a web-based e-assessment platform featuring learning analytics and gamification. International Journal of Web Information Systems.
P124	Calderón, A., Ruiz, M., & O'Connor, R. V. (2017, September). ProDecAdmin: a game scenario design tool for software project management training. In European Conference on Software Process Improvement (pp. 241-248). Springer, Cham.
P125	Üsfekes, Ç., Yilmaz, M., Tuzun, E., Clarke, P. M., & O'Connor, R. V. (2017, September). Examining reward mechanisms for effective usage of application lifecycle management tools. In European Conference on Software Process Improvement (pp. 259-268). Springer, Cham.
P126	Khandelwal, S., Sripada, S. K., & Reddy, Y. R. (2017, February). Impact of gamification on code review process: an experimental study. In Proceedings of the 10th innovations in software engineering conference (pp. 122-126).
P127	Liechti, O., Pasquier, J., & Reis, R. (2017, May). Supporting agile teams with a test analytics platform: a case study. In 2017 IEEE/ACM 12th International Workshop on Automation of Software Testing (AST) (pp. 9-15). IEEE.
P128	Mesquida, A. L., Jovanović, J., Jovanović, M., & Mas, A. (2018). Agile software process improvement: a collaborative game toolbox. IET Software, 13(2), 106-111.
P129	Neto, P. S., Medeiros, D. B., Ibiapina, I., & da Costa Castro, O. C. (2018). Case study of the introduction of game design techniques in software development. IET Software, 13(2), 129-143.
P130	Naik, N., & Jenkins, P. (2019, August). Relax, it'sa game: utilising gamification in learning agile scrum software development. In 2019 IEEE Conference on Games (CoG) (pp. 1-4). IEEE.

Table 7: Continued.

ID	Primary studies
P131	Herranz, E., Guzmán, J. G., de Amescua-Seco, A., & Larrucea, X. (2018). Gamification for software process improvement: a practical approach. IET Software, 13(2), 112-121.
P132	Ivan, G., Carla, P., & Antonio, C. M. J. (2019). Introducing gamification to increase staff involvement and motivation when conducting SPI initiatives in small-sized software enterprises. IET Software, 13(5), 456-465.
P133	Üsfekes, Ç., Tüzün, E., Yılmaz, M., Macit, Y., & Clarke, P. (2019). Auction-based serious game for bug tracking. IET Software, 13(5), 386-392.
P134	Kasahara, R., Sakamoto, K., Washizaki, H., & Fukazawa, Y. (2019, July). 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 (pp. 92-98).
P135	Furtado, J., Oliveira, S. R. B., Chaves, R. O., Telles, A., & Colares, A. (2021). An Experimental Evaluation of a Teaching Approach for Statistical Process Control in Computer Courses. International Journal of Information and Communication Technology Education (IJICTE), 17(1), 154-171.
P136	Furtado, L., & Oliveira, S. R. B. (2018). A Teaching Approach of Software Measurement Process with Gamification-A Experimental Study. In ICSOFT (pp. 267-274).
P137	Marín, B., Vera, M., & Giachetti, G. (2019). An Adventure Serious Game for Teaching Effort Estimation in Software Engineering. In IWSM-Mensura (pp. 71-86).

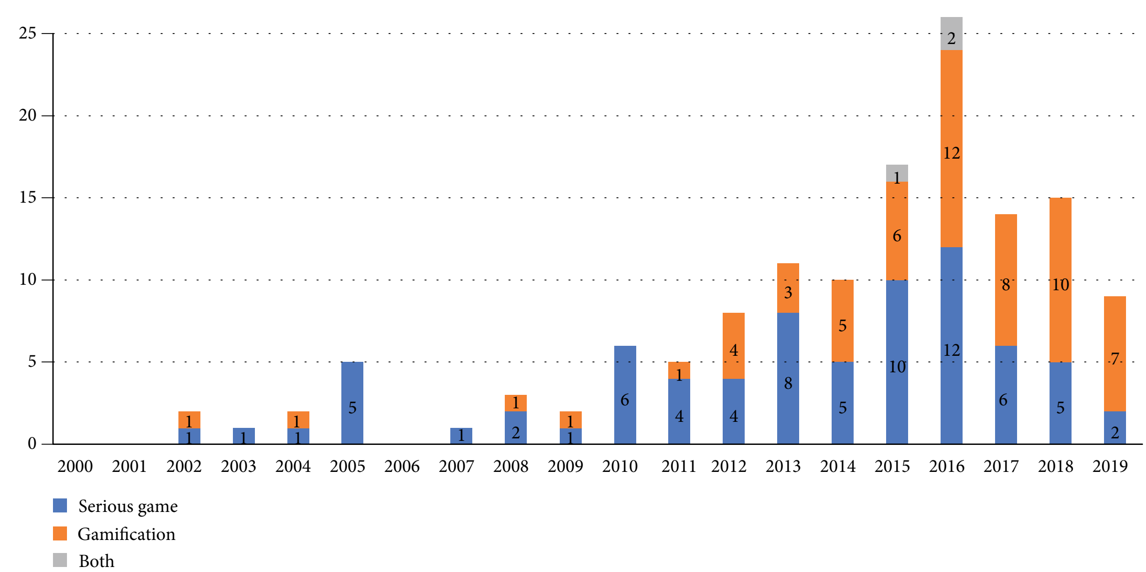

Figure 4: Publications per year.

unattractive, so at the risk of the participants lost interest in the application. In addition, in this same work, it was stated that the participants were bored, as they considered the project repetitive and with many interruptions. This contrasted with the P7 study, which includes the Gamiware tool, a gamification platform which is aimed at increasing motivation in software projects. It was reported that its main limitation was the lack of integration with a consolidated framework for project management, such as Jira or Redmine.

The study P24 is a learning process that features a quiz-making and quiz-solving game, where students work together in teams to create multiple-choice quiz questions which challenge the knowledge of the other students. However, some of the students complained that there was a lack of balance in the project which discouraged them, because it was very difficult to obtain points. The purpose of study P39 was to allow the students to compete, by using gamification, and play a role, join a team, and choose a project. However, it was found that not all students liked the competitive atmosphere surrounding gamification, and they argued that this great pressure led to a sense of demotivation. As in the case of P53, which seeks to encourage code refactoring through the use of gamification, it was found that some students unnecessarily refactored the code just to get more points in the game. As a result, some students admitted that they were cheating and did not deserve their score, and hence, this mistaken assignment of points ended up discouraging those students who had earned the points honestly. In

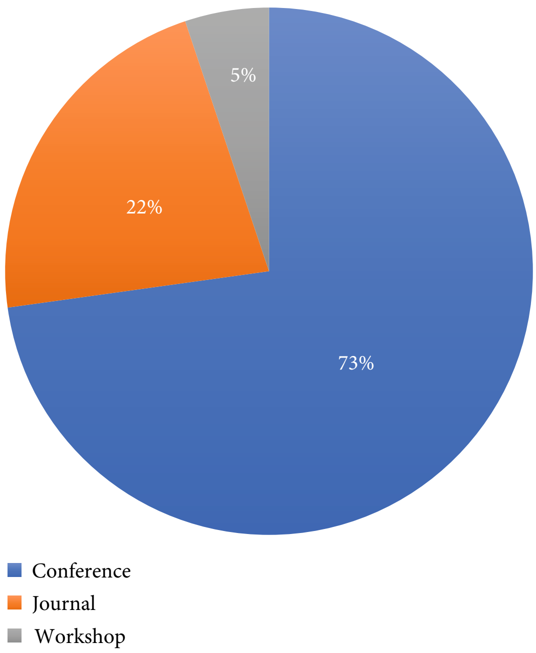

Figure 5: Types of events.

addition, the P53 showed that the use of the leaderboard, which is always available on the screen during the sessions of the game, ended up discouraging the students, because it caused some of them to abandon the gamification when they found themselves at the bottom of the leaderboard.

In P62, it was a gamification to check the punctuality of those who attend group meetings, which showed that if there is a lack of penalties in the project, it made it easier for the participants to become disengaged. In study P82, it was argued that the aesthetic features are not shown in a user-friendly way, since the participants are burdened with a lot of textual information, arising from the use of the quiz mechanics that proved to be unattractive to some students. The study P89 suggested that the lack of feedback in the small cycles increased the anxiety of some students and thus hampered their focus, because the feedback from the leaderboard was only provided once, at the end of a week.

Another factor that was regarded as a limitation was the "replay value," as was the case with P106. The purpose of this study is to create an educational board game that shows the basic concepts of the essence of software engineering in an enjoyable way. But, even though, the participants found the game experience fun for a few rounds, it gradually became less interesting, as the replay value diminished. In the case of others, like P108, where gamified tool was used to track the code review, many badges were awarded without any motivational value, and as the leaderboard was easy to handle, it meant that only small contributions were made, without much relevance. Thus, this system did not take into account the complexity of the task, but only the number of tasks performed by the same user.

Approximately 26% of all the papers/articles dealing with serious games described their limitations. The following limitations are related to the game itself and not the methodology of the study.

In studies P16, P37, P85, and P119, it was reported that there was no statistical difference between the control group and the experimental group, which suggests there is an equivalence between teaching methods. In addition, most of the participants in the P16 study did not enjoy playing the game and some of them said they would prefer a written exercise to the game itself. This same game used quiz mechanics and had unattractive graphics and no sound effects. Apparently, the aesthetic features and mechanics of the game were not ideally suited to meeting the requirements of the target audience.

In paper P32, it was found that the low learning curve discouraged students from engaging with it. However, study P43 pointed out that the tool that was used (ProDec) should only be regarded as a support tool because it just helps students to apply their knowledge. In light of this, the students had to obtain knowledge by other methods, such as traditional classes. Similarly, study P100 maintained that its PlayScrum tool should not be used alone, but in conjunction with traditional teaching, like P110 that introduced legacy, which is a board game for teaching project management, who had a serious limitation, and the high difficulty of the challenge, which made it hard to engage students who had little or no practical experience?

In the case of papers P49, P50, and P51, the drawback was the degree of realism of the game. Some students complained that the game did not reflect reality, and they had limited content. They were aware that the game was supposed to deal with the teaching of life cycles but in fact only included the Waterfall model and left other life cycle models aside. Moreover, several students thought that the "requirements and design" phases were boring. Another serious drawback is the rejection or scepticism of people who are not familiar with the use of games in the workplace, as mentioned in paper P67, which is concerned with a card game to meet requirements in an organizational environment. At the same time, some limitations were also noted in the teaching environment. As was mentioned in work P83 which taught software process modelling through a serious game, in this study, it was noticed that the game-imposed constraints on the creativity of the students in the answers put forward in its teaching method and also induced the students only to memorize the answers of the game and remember its predefined models.

4.4. GQ3: What Research Methods Were Employed in the Validation of Gamification or Serious Game Projects? Two methods were employed for the validation of the projects—-validation with users and validation involving experts. Not all the papers/articles validated their projects, and about 41% of the primary studies were purely theoretical and set out schemes without implementing them. However, about 55% of the remaining papers/articles were validated through the application of questionnaires, and 12.5% were validated through experiments involving two groups—one for control and another in the experiment. This meant that the results could be compared and it could be determined if the data had any statistical significance. In addition, the following methods were employed: user interviews and mixed method research (i.e., research that relies on more than one of the

Table 8: Publication venues.

Number of studies	Conferences
4	Frontiers in Education (FIE)
4	International Conference on Software Process Improvement and Capability dEtermination (SPICE)
4	European Conference on Software Process Improvement (EUROSPI)
7	International Conference on Software Engineering (ICSE)
7	Conference on Software Engineering Education & Training (CSEE&T)
Number of studies	**Journals**
2	Computer Systems and Software Engineering (CSSE)
2	Revista Facultad de Ingeniería Universidad de Antioquia (RFIUA)
3	Journal of Universal Computer Science (JUCS)
4	Computer Standards & Interfaces (CSI)
7	IET Software (IET)
Number of studies	**Workshops**
2	International Workshop on Cooperative and Human Aspects of Software Engineering (IWCHASE)
3	International Workshop on Games and Software Engineering (GAS)

Table 9: Authors, universities, and countries that have had most published primary studies.

Universities	Studies per university
Østfold University College	4
Çankaya University	5
University of California	6
Dublin City University	6
Universidad Carlos III de Madrid	8
First authors	**Studies per author**
Christiane Gresse von Wangenheim	2
Emily oh Navarro	3
Alejandro Calderón	3
Elizabeth Suescún Monsalve	3
Eduardo Herranz	6
Countries	**Studies per country**
Republic of Ireland	7
Germany	9
Spain	21
USA	23
Brazil	25

previous methods). Each of them had a 2.5% share of all the primary studies. Furthermore, a usability test had a 3.75% share of the total. Finally, the user observation testing technique (used in ethnography) was an evaluation method that comprised 1.25% of the total number of analysed papers/articles.

Two approaches were adopted to ensure a high-quality assessment. The first entailed conducting an interview with experts, and this was used in 16.25% of the papers/articles, and on the basis of this, it was possible to conduct a survey or make a presentation of the projects to allow the experts give their opinions. The second strategy was the application of the Delphi method for the validation of 2.5% of the projects. In this technique, a questionnaire is sent to a team of experts who are not aware of the views of other experts and in consecutive rounds, and the experts' opinions are analysed until a consensus is reached. Figure 6 shows each of the validation methods found in the papers/articles of this review, and in Table 11, each study is examined by means of this validation method.

As can be seen in Figure 6, most studies analysed used questionnaires or interviews as a validation method. This points out that the vast majority of studies have not been concerned with verifying the accuracy of their proposals in the face of the traditional teaching method. In other words, only 12.5% of the studies that reported their validation method used a control group and another experiment group to compare their results and verify the effectiveness of their proposals. This can be related to the difficulty and time needed to perform an experiment. Consequently, few of the 137 primary studies will be used to support the development of a teaching tool on software measurement, which is future work for these authors.

4.5. GQ4: What Game Elements Were Used in the Gamification or Serious Game Projects? The elements of a game are the components used in its interface and are intrinsically related to dynamics and mechanics. A mechanic makes use of the implementation of one or more game elements, and a dynamic is the composition of one or more game mechanics. Since gaming components form the atomic structure of a gamified project, the central task of gamification is to combine them in order to create motivating and engaging mechanics and dynamics. In other words, when designing gamified proposals, the concern must be with the intelligent concatenation of the elements, mechanics, and dynamics of the games. Otherwise, these proposals are usually doomed to fail. This study was based on the book

Table 10: Application contexts.

Category	Primary papers/articles	Percentage
Professional	P1 P2 P4 P6 P7 P8 P15 P19 P20 P22 P26 P28 P29 P31 P35 P38 P40 P45 P47 P54 P56 P57 P58 P60 P61 P65 P66 P67 P69 P72 P75 P77 P79 P80 P90 P98 P102 P105 P107 P112 P114 P117 P122 P125 P127 P128 P129 P131 P132 P133	36.76%
Academic	P3 P9 P10 P11 P12 P13 P14 P16 P17 P18 P21 P23 P24 P25 P30 P32 P33 P34 P36 P37 P39 P41 P42 P43 P44 P46 P48 P49 P50 P51 P52 P53 P55 P59 P62 P63 P64 P68 P70 P71 P73 P74 P76 P78 P81 P82 P83 P84 P85 P86 P87 P88 P89 P91 P92 P93 P94 P95 P96 P97 P99 P100 P101 P103 P104 P106 P108 P109 P110 P111 P113 P115 P116 P118 P119 P120 P121 P123 P124 P130 P134 P135 P136 P137	61.76%
Both	P126	0.74%
Not reported	P27	0.74%

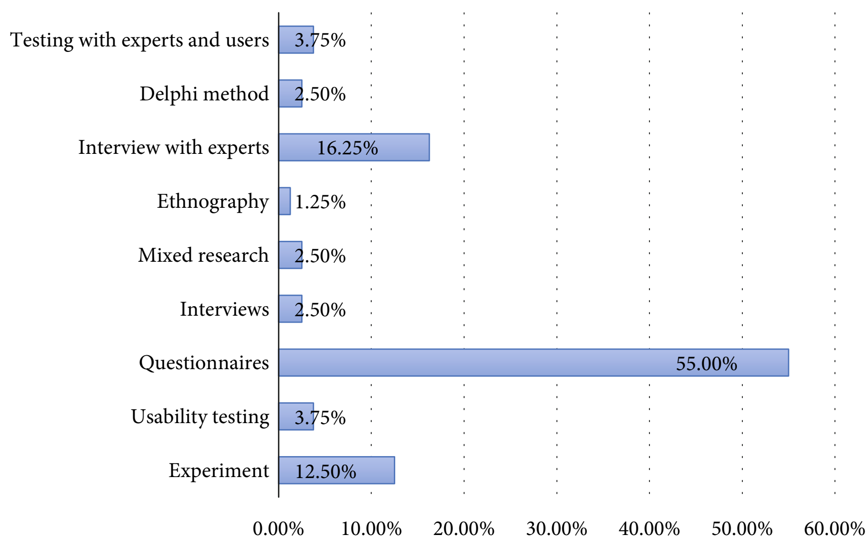

Figure 6: Validation methods.

Table 11: Studies by validation methods.

Validation with users	Experiment	P16 P86 P87 P88 P90 p92 P100 P102 P134 P135
	Usability testing	P46 P96 P137
	Questionnaires	P1 P2 P11 P6 P7 P15 P19 P21 p23 P24 P25 P33 P36 P37 p38 P39 P44 P47 P48 p49 P50 P51 P52 P53 P55 P57 P59 P60 P64 P68 P69 P74 P75 P79 P83 P85 P89 P94 P95 P97 P98 P101 P129 P131
	Interviews	P61 P62
	Mixed research	P70 P72
	Ethnography	P45
Validation with experts	Interview with experts	P5 P28 P29 P40 P42 P58 P77 P78 P80 P93 P99 P105 P109
	Delphi method	P31 P43
Validation with users and experts	Testing with experts and users	P123 P108 P118

by [15] to select a set of game elements that would be investigated in the primary studies, and these game elements can be listed and explained as follows:

(a) *Avatar*. A visual representation of the player in the game world,

(b) *Virtual Goods*. In-game items that players can collect and use in a virtual rather than real fashion, but which still generate endogenous value to the players. Players can pay for items either with game currency or with real money,

(c) *Boss*. A generally difficult challenge at the end of a level that has to be overcome before an advance can be made in the game,

(d) *Collections*. Formed of items accumulated within the game. Badges and medals often form a part of the collections,

(e) *Combat*. A dispute that occurs which allows the player to defeat opponents in a confrontation,

(f) *Achievements*. A reward that the player receives for carrying out a set of specific tasks,

(g) *Unlockable Content*. The ability to unblock and access certain content in the game if a number of prerequisites are fulfilled. The player needs to do something specific to be able to unlock the content,

(h) *Badges*. Visual emblems of in-game achievements,

(i) *Social Graph*. A diagram that makes it possible to see friends who are also in the game and interact with them. A social graph makes the game an extension of one's social networking experience,

(j) *Mission*. Similar to "achievements." It is an assignment in which the player must carry out some activities that are specifically defined within the framework of the game,

(k) *Levels*. Numerical representation of the progress made by the player. The player's level rises as the player becomes better at the game,

(l) *Points*. In-game actions that score points. They are often linked to levels,

(m) *Gifts*. The possibility of providing items or virtual currency to other players,

(n) *Leaderboard*. A means of listing players who have the highest scores in a game,

(o) *Teams*. Possibility of playing with other people who have the same goal.

Of all the game elements that were searched, only the transaction and boss had no correspondence, while the collections and gifts had only one incidence among the primary studies. As expected, the most widely used game elements in the primary studies were the PBL triad: points with 59.85%, badges with 22.63%, and leaderboards with 25.55%. The other elements of the games, together with their frequencies of use and respective primary studies, can be seen in Table 12. However, many primary studies do not appear in this Table because they are often purely theoretical and do not include game elements or report them in their work.

4.6. GQ5: What Game Mechanics Were Used in the Gamification or Serious Game Projects? Mechanics are the second layer discussed by [15], and these are similar to the rules of the game and how the player should interact with it. That is, they steer the players' activities in the required direction by setting out what a player can or cannot do during the game. The mechanics also have an intrinsic relation to the genre of the game; for example, the RPG or Board usually employs the mechanics of turns. In addition, the implementation of a dynamic can be operated by the use of one or more mechanics; for example, the dynamics of progression can be implemented by the use of mechanical feedback and rewards. The mechanics defined by [15] are listed and explained below.

(a) *Resource Acquisition*. The player can collect items that might help him achieve his goals,

(b) *Feedback*. The evaluation allows players to see how they are progressing in the game,

(c) *Chance*. The results of the player's activities are random and can thus create a sense of surprise and uncertainty,

(d) *Cooperation and Competition*. This creates a sense of triumph or disappointment in defeat,

(e) *Challenges*. The goals that the game defines for the player,

(f) *Rewards*. The benefits that the player can gain from achievements in the game,

(g) *Transactions*. The means of buying, selling, or exchanging something with other players in the game,

(h) *Turns*. Each player in the game has his/her own time and opportunity to play. Traditional games such as card games and board games often rely on turns to maintain a balance in the game, while many modern computer games take place in real time,

(i) *Win State*. The "state" that defines who is winning the game.

Table 13 lists the mechanics and the primary studies that implemented them. The process used to identify the mechanics involved a complete reading of the paper/article in search of the mechanics used by the author. Even if the author did not supply this information, a search was made to find evidence of the use of the mechanics listed by Werbach and Hunter [15]. There were no occurrences of transaction mechanic in the primary papers/articles.

As can be seen in Table 13, the most used mechanics were feedback (25.55%) and cooperation and competition (29.20%). The feedback was used to engage the user with

Table 12: Game elements per primary studies.

Game elements	Primary studies	Percentage
Avatar	P3 P4 P17 P18 P29 P40 P52 P66 P79 P87 P94 P99 P101 P106 P111 P117	11.68%
Collections	P29	0.73%
Combat	P3 P54	1.46%
Achievements	P3 P7 P29 P60 P66 P91 P99	5.11%
Unlockable content	P3 P109	1.46%
Badges	P3 P4 P7 P22 P29 P40 P52 P57 P58 P60 P63 P79 P80 P91 P99 P103 P104 P107 P108 P109 P111 P112 P115 P117 P123 P126 P127 P129 P131 P132 P136	22.63%
Social graph	P3 P58 P65 P76 P79 P81 P117 P129 P132	6.57%
Levels	P3 P7 P29 P40 P58 P75 P99 P101 P107 P117 P123 P129 P132	9.49%
Points	P3 P4 P7 P12 P13 P14 P15 P16 P17 P18 P19 P22 P24 P29 P30 P32 P33 P37 P39 P40 P42 P44 P45 P47 P50 P51 P52 P53 P54 P55 P57 P58 P60 P61 P62 P63 P65 P66 P69 P70 P72 P75 P76 P77 P79 P80 P81 P82 P84 P85 P86 P87 P88 P89 P92 P94 P95 P96 P97 P99 P100 P101 P103 P104 P107 P108 P111 P112 P115 P117 P118 P119 P123 P125 P126 P127 P129 P131 P132 P133 P134 P136	59.85%
Gifts	P19	0.73%
Leaderboard	P3 P4 P7 P13 P14 P19 P22 P29 P40 P42 P52 P53 P57 P62 P79 P80 P82 P86 P88 P89 P94 P103 P104 P109 P112 P114 P115 P117 P123 P126 P127 P129 P132 P134 P136	25.55%
Virtual goods	P3 P48 P112	2.19%

Table 13: Mechanics per primary studies.

Game mechanics	Primary studies	Percentage
Resource acquisition	P3 P12 P13 P14 P15 P17 P18 P33 P87 P117	7.30%
Feedback	P3 P7 P12 P13 P14 P15 P16 P17 P18 P22 P29 P33 P40 P42 P53 P54 P57 P58 P59 P60 P61 P62 P66 P67 P68 P69 P75 P87 P99 P103 P117 P118 P129 P132 P134	25.55%
Chance	P3 P12 P13 P14 P15 P17 P18 P44 P48 P51 P87 P110 P111	9.49%
Cooperation and competition	P3 P4 P7 P19 P22 P24 P29 P32 P33 P34 P37 P39 P40 P49 P51 P52 P53 P54 P55 P58 P63 P74 P75 P76 P81 P85 P100 P103 P106 P107 P112 P115 P118 P119 P129 P130 P131 P132 P133 P134	29.20%
Challenges	P3 P29 P53 P75 P77 P83 P84 P103 P109 P117 P129 P132	8.76%
Rewards	P19 P29 P40 P58 P75 P80 P89 P107 P117 P133	7.30%
Turns	P36 P37 P44 P48 P50 P51 P54 P85 P100 P106 P110 P111 P119	9.49%
Win state	P3 P40 P68 P70 P72 P75 P84 P89 P92 P118	7.30%

the game approach by presenting the progress with different game elements, like points, leaderboards, levels, and social graphs, as a constant factor to encourage the user to keep the game on. The cooperation and competition used similar game elements to create an environment of player versus player or team versus team. These two mechanics were the most common ones used in games.

4.7. GQ6: What Game Dynamics Were Used in the Gamification or Serious Game Projects? Dynamics are the highest level of abstraction of the set of game components outlined by [15]. They are usually related to the sensations that gamification seeks to arouse in users. Below are listed and conceptualized the dynamics defined by [15]:

(a) *Emotions*. Games can induce different types of emotions, especially a sense of fun (an emotional reinforcement that keeps people playing),

(b) *Narrative*. The structure that makes the game coherent. The narrative does not have to be explicit, like a story in a game. It can also be implicit, in so far as all the experience has a purpose in itself,

(c) *Progression*. The idea of giving players the feeling of advancing within the game,

(d) *Relationships*. Refers to the interaction between the players, whether it be between friends, companions, or adversaries,

(e) *Restrictions*. Refers to limiting the freedom of the players within the game.

The use of dynamics is not mandatory, but is important because of its effects on user engagement within gamification. As an example, in the primary study P3, which makes use of gamification for teaching risk management, it was recommended that restriction, narrative, and progression dynamics should be employed to encourage users to carry out the project in the best possible way. When the dynamics were implemented as a means of restricting the user's time for decision-making, this helped create a sense of urgency. Moreover, the narrative provided the activities with a context and sense of order, while the progression was implemented with represented graphics that showed the advances made by the participants to ensure a sense of achievement or triumph.

The dynamics that were used less frequently were the "Emotion" category, which was only found in article P3 and papers P75 and P112. This was designed to achieve the following goal, as stated in the primary study P3, "The game must cause any imaginable emotion. During the gamification course, it is expected to arouse emotions in the participants to encourage them to complete identification and analysis tasks, and enable students to earn points, complete all the levels and win the game." While the most widely used dynamics in primary studies were "Relationship" (as can be seen in Table 14), these studies were generally implemented by means of cooperation and competition and using game elements such as a social graph, leaderboard, gifts, levels, and avatars.

Remember that a dynamic is only achieved from the application of mechanics and game elements, according to Table 14, which presented the greatest occurrence, among dynamics, the use of relationships, which has competition and cooperation mechanics as one of its bases. In addition, the relationship dynamic is one of the most applied in commercial games and also in educational games, due to the majority of games exploring aspects of multiplayer. So, unsurprisingly, this was the dynamic most explored by the studies analysed in this SRL.

4.8. GQ7: What Genres Were Used in the Gamification or Serious Game Projects? The game genre offers an established classification of entertainment games that provides a useful way of identifying characteristics that the games have in common. One of these categories which is well accepted by industry is defined by Herz [45].

Herz distinguishes between the following game genres: action, adventure, fighting, puzzles, RPG (role-playing games), simulation, sport, strategy, cards, and board. In addition to the standard classification of Herz, we add two more categories, which are without genres and collaborative games, because some primary studies were classified this way. Table 15 shows the relationship between genres and primary studies.

The most notorious category was the without a genre, and this can be understood because gamification is not necessarily a game, and hence, it is not always possible to classify it in terms of a genre. The second most frequent genre was simulation, which allows the researcher to create a safe environment for their students to learn by trial and error in a game that simulates the software process in real-life situations.

4.9. GQ8: How Does the Effectiveness of Learning through Gamification or Serious Games Compare with That of Traditional Learning? By investigating the primary studies found in this SRL, it could be determined that some studies (P36, P39, P70, P71, and P72) showed significant gains with regard to student learning. The studies (P36 and P39) show a gain with regard to the student's awareness of the concepts defined for the studied subject. In article P36, a board game was provided to teach efficient communication during the elicitation requirements process, and the results were an improvement in the users' perception of the importance of communication in this process, as is made clear in the article. Before the use of the game, only 27% of the students thought this kind of process was important but after the application of the game, this perception increased to 68%. In concordance, in the P39 study, 67% of the students felt that they learned the concepts of software engineering more easily by using the game-based strategy and 80% agreed that they had much more practical knowledge through this teaching method than by the traditional approach. In the traditional approach, teaching tends to be teacher-centered and the students only carry out tasks that are prescribed for them [46].

On the other hand, articles P70, P71, P72, and P135 showed a significant statistical gain with regard to the traditional methods of teaching. As was shown in P70 when comparing Cohen's stipulated value of 0.8 with the obtained value of 1.35, the Cohen's d effect compares the average of two groups (like a "control" and an "experimental" group), by subtracting the mean average of the control group from that of the experimental one and dividing this result by the average standard deviation. The result of this equation can be interpreted in 3 ranges: 0 to 0.2 is regarded as a small effect, 0.2 to 0.5 is a medium effect, and greater than 0.5 is a large effect. Thus, in this case, it can be stated that there was a large effect on student learning. However, in P71 when the answers of the knowledge questionnaires were compared between group A and group B, there was an improvement which suggests that the game actually promoted the acquisition of knowledge about project management. Furthermore, this was also the case in article P135, where statistical process control is taught by means of collaborative games. Apart from this, the students designed a measurement plan with the aid of Goal Question Metrics (GQM) and discussed what might be the most feasible chart to represent it. The results of this work compare the grades of the students in the control and experimental group and found the planned approach leads to more effective learning since the average score obtained in the experimental group was 30% higher than that

Table 14: Dynamics per primary studies.

Game dynamics	Primary studies	Percentage
Emotions	P3 P75 P112	2.19%
Narrative	P3 P12 P13 P14 P15 P16 P17 P18 P20 P29 P33 P34 P48 P50 P56 P64 P68 P75 P87 P92 P94 P137	16.06%
Progression	P3 P7 P12 P13 P14 P17 P18 P22 P84 P87 P92 P101 P107 P118 P129 P132	11.68%
Relationships	P3 P4 P7 P12 P13 P15 P19 P25 P29 P32 P38 P40 P42 P44 P45 P47 P49 P51 P52 P54 P55 P57 P58 P59 P60 P61 P62 P63 P64 P65 P66 P67 P69 P74 P75 P76 P81 P98 P99 P100 P103 P106 P107 P117	32.12%
Restrictions	P3 P12 P13 P14 P15 P32 P48 P50 P51 P68 P75 P130 P137	9.49%

Table 15: Genres per primary studies.

Game genres	Primary studies	Percentage
Simulation	P1 P12 P13 P14 P15 P16 P17 P18 P20 P32 P33 P34 P43 P46 P48 P50 P52 P53 P55 P56 P58 P60 P61 P62 P65 P68 P70 P71 P72 P77 P79 P82 P83 P84 P88 P92 P93 P96 P105 P116 P121 P124	30.66%
RPG	P6 P21 P23 P29 P66 P95 P99	5.11%
Collaborative games	P8 P59 P64 P78 P90 P102 P120 P128 P135	6.57%
Cards	P25 P30 P38 P42 P45 P49 P51 P55 P67 P74 P100 P119	8.76%
Board	P36 P37 P44 P63 P65 P100 P106 P110 P111 P119	7.30%
Adventure	P41 P137	1.46%
Strategy	P40 P54 P57 P65	2.92%
Without a genre	P2 P3 P4 P5 P7 P9 P10 P11 P19 P22 P24 P26 P27 P28 P31 P35 P39 P47 P55 P69 P73 P75 P76 P80 P81 P91 P94 P101 P102 P103 P104 P107 P108 P109 P112 P113 P114 P115 P117 P122 P 123 P125 P126 P127 P129 P130 P131 P132 P133 P134 P136	37.23%

obtained by the control group. Finally, in the P72 study, it could be concluded that there was a positive gain when a comparison was made between pre and postquestionnaires in a class of 42 students who achieved scores of 39.05 and 61.91, respectively, after they had been taught by a method based on RPG that was used in the classroom for the teaching of measurement and analysis through estimates of cost, time, and risk.

However, most studies such as P16, P37, P44, P53, P83, P85, P104, P113, P115, P119, P121, P126, and P131 showed no statistically significant gain or signs that the game-based schemes were superior to the traditional teaching environment, although they were often referred to as having the same effect and being equally effective in teaching. For example, in the case of article P16 where software measurement is taught through the GQM paradigm, no statistical differences were found between the control and experimental groups. As discussed in P44, the results obtained in these studies, may show that games or gamification are not more efficient in teaching than in traditional classes. In contrast, these studies suggest that these schemes are no less efficient than the traditional medium. That is, even if the value of teaching software process improvements and software measurement is questionable, this does not preclude the qualities involved in games and gamification from being motivational or mean they are unable to create a safe and simulating application environment for the deployment of practical knowledge. As article P16 makes clear, these schemes can be beneficial when used as a teaching support tool.

4.10. SPIQ1: In What Processes (Measurement and Requirements Collection, among Others) Were the Gamification System or Serious Games Applied in the Area of SPI? Software process improvement carries out practical activities to optimize the processes in the organization and ensure that they meet the business objectives, more effectively [20], i.e., to deliver software faster to the market, improve quality, and reduce waste. The goal is to make the organization more competitive by producing higher-quality software in less time and at a more affordable price.

There are international and national standards and models that are designed to optimize the organizational processes and, hence, make an improvement in the quality of the software. These include the CMMI model, the MPS.BR model, and ISO 15504. The CMMI model that was developed by the CMMI Institute (organization belonging to ISACA) includes best practices for software and systems processes, since it is an internationally adopted standard. Thus, this

Table 16: Studies with gamified software processes.

Process area	Primary studies	Percentage
Requirements management	P7 P17 P18 P21 P29 P32 P33 P34 P36 P38 P39 P46 P49 P51 P55 P89 P90 P92 P93 P96 P97 P98 P100 P110	17.52%
Project planning	P12 P13 P14 P17 P18 P21 P29 P32 P33 P34 P39 P40 P42 P43 P44 P48 P49 P50 P51 P59 P64 P68 P79 P89 P90 P92 P93 P95 P98 P100 P105 P106 P132	24.09%
Project monitoring and control	P15 P23 P29 P33 P34 P76 P79 P81 P87 P89 P92 P93 P98 P100 P132	10.95%
Supplier agreement management	P70 P93	1.46%
Measurement and analysis	P10 P12 P16 P17 P19 P22 P29 P34 P43 P44 P53 P59 P63 P72 P89 P93 P94 P98 P99 P110 P124 P131 P133 P134 P136 P137	18.98%
Process and product quality assurance	P10 P15 P26 P41 P53 P58 P63 P65 P66 P70 P71 P74 P82 P88 P89 P91 P93 P127 P133 P134	14.60%
Configuration management	P93 P94	1.46%
Requirements development	P6 P7 P17 P18 P21 P29 P32 P33 P36 P38 P39 P46 P49 P51 P55 P67 P89 P90 P92 P93 P96 P97 P100 P102	17.52%
Technical solution	P4 P15 P25 P60 P75 P93 P112 P115	5.84%
Product integration	P93	0.73%
Verification	P4 P27 P29 P33 P70 P89 P92 P93 P100 P108 P111 P126	8.76%
Validation	P6 P27 P29 P33 P70 P80 P89 P92 P93 P100 P126	8.03%
Organizational process focus	P19 P27 P29 P33 P38 P56 P60 P61 P92 P93 P100	7.30%
Organizational process definition	P27 P29 P33 P38 P56 P60 P83 P92 P93 P100	7.30%
Organizational training	P15 P29 P33 P52 P62 P92 P93 P100 P109 P114	7.30%
Integrated project management	P76 P81 P93	2.19%
Risk management	P3 P12 P13 P14 P15 P37 P85 P86 P93 P119	7.30%

model was chosen as a reference point for the processes addressed in this work because it allows a company to improve its processes based on the application of maturity and capacity, enabling an analysis the performance of these processes.

During the data extraction, some of the processes listed by CMMI were identified in the primary studies. Table 16 lists the processes and studies that addressed these processes. The papers/articles included in the primary study are papers/articles that deal directly with the SPI area, as well as papers/articles that address some SPI process (even if indirectly). Most of the studies were about project planning (24%), while the measurement and analysis process was only found in 18% of all the primary studies. It was noted that a considerable proportion of the studies used more than one process area.

The most relevant process area to this research was the measurement and analysis with 18.98% (see Table 16) of the total papers/articles. This process was the second most frequent process, second only to project planning with 24.09%, which can be considered the most important because it allows developing plans to describe what is needed to accomplish the work within the standards and constraints of the organization. One of the reasons the measurement and analysis takes second place is that this process is used as a basis for the other processes, as project planning. Although the measurement process has a significant occurrence, the vast majority of primary papers/articles do not address the software measurement process as an exclusive focus, and only 4% (P16, P29, P134, P135, P136, and P137) of all primary papers/articles have an exclusive focus on it, which is a very low percentage, showing that this field has a lot to be explored. In addition, as can be seen in Table 16, the primary papers/articles usually have more than one related software process.

4.11. MEAQ1: Was the Scheme Examined Based on a Model or Standard or Paradigm? If So, Which? Concepts such as GQM (Goal-Question-Metric) and Practical Software Measurement (PSM) paradigms, norms such as ISO 15939 and ISO 25000, and models such as MPS.BR and CMMI-DEV are commonly found in the subject of software measurement. For this reason, this research question is aimed at identifying the paradigms, norms, and models used in primary studies.

Primary study P16 that showed a serious game for software measurement teaching, based on the Goal-Question-Metric (GQM) paradigm, laid exclusive emphasis on the software measurement process and its different stages (collect, store, analyse, and report). Another article that clearly made use of GQM was P135, which was employed to design a measurement plan with the purpose of teaching statistical process control by means of collaborative games. The only approach that used COSMIC function points was paper P137.

Other studies have addressed software measurement, but have not limited themselves to investigating this process. In other words, these studies only mentioned this type of process, but did not describe in detail the serious game or the planned gamification, for example, in the article P28 and papers P43 and P93. Study P28 provides an overview of a system to gamify a software process improvement program without implementing it. A team of gamification and SPI experts has only validated it, as measurement is one of the basic requirements for any SPI. This article includes such a process, although it is not restricted to it, but just sees it as another process within the same program. As it is an article that provides only an overview of a gamification structure of an SPI program, it does not specify in detail what the measurement process and how the gamification of this process should be. It only guides the use of the GQM paradigm, as one of the foundations of this process. In addition, similarly, studies P43 and P93 make use of a serious game for teaching software process models and this involves the use of ISO 12207 and CMMI. As mentioned earlier, these works do not investigate the measurement process, but still refer to it because it is part of the software process models addressed.

Thus, it is clear that making use of serious games and gamification as a teaching or training tool for software measurement has still not been adequately explored, since out of 19050 papers/articles, it was only possible to extract six which concentrated exclusively on this process.

4.12. MEAQ2: What Metrics Were Covered by the Gamification or Serious Games Schemes? This question highlights which metrics are being explored and which need more attention. Moreover, how they were used and how they relate to gamification. Gamification, in general, is intended to measure certain desired behaviours; consequently, the metrics were used within gamification as mechanics, which verified the behaviour of the user within the reach of the milestones stipulated by the game. Therefore, the metrics acted as a basis for gamification, acting as part of the rules of the game and also used as an evaluative component.

Since metrics are one of the key components of gamification, they must be correctly selected to encourage the desired behaviour. Thus, it is worth drawing attention to some criticisms that have been made about the use of metrics in gamification. As previously noted in study P63, the lines of code and coverage of test cases were adopted as metrics and the mechanical features of its gamification. As a result, the author states that some students tried to exploit the system by adding lines of code or test cases that had no significance to the project. As a countermeasure, the author notified the students that the codes produced would be randomly selected for manual or automatic evaluation. This same behaviour was apparent in other works such as the P53 study, which made use of gamification to encourage the practice of code refactoring, but found some students made use of "insipid refactoring" to increase their score and move up in the ranking. This distorted the balance in the game and allowed students to exploit it.

Table 17 lists the metrics identified in the different primary studies. It should be reiterated that few metrics include an explanation of their calculation.

As can be seen in Table 17, the most widely used metric in the primary papers/articles were lines of code (with 3.65%). This is understandable because this is one of the simplest metrics that can be applied in a software project. However, its great weakness is in an industrial application, because depending on the language used, this metric varies a lot. For example, a language like Java is much more verbose than Python, and since multiple languages are commonly used for software projects in the same project, this metric ends up being inaccurate. Nevertheless, the lines of code metric is an excellent starting point for introducing concepts of software measurement because of its simplicity and the fact that it is easy to understand.

In second place was sprint velocity with 2.92%. This is understandable since the Scrum method is one of the most widely used in academia and industry [47]. On account of its simplicity in managing teams and effectiveness as a method, some of its practices are recommended. Among these practices, it has the capacity to measure the development process and divide tasks into user stories that have their difficulties linked to points. This method recommends measuring the number of user story points the team can produce in a time interval (sprint). This relationship between sprint and the number of user story points is what is called sprint velocity. This is one of the ways to measure the productivity of a team in a Scrum environment. However, papers P12 and P93 and article P17, which represent 2.19% of the primary papers/articles, measured productivity differently. P12 was the only system that showed the calculation made by the productivity metric. In the paper, the productivity per employee was calculated, that is, it was estimated how many tasks a team member was able to perform in a time box. However, since it is a personal metric, care should be taken in how it is used, as it can make a team member feel he/she is at risk, if there is a low value in that metric.

Another categories of most widely used metrics were the estimates of time (2.92%), size (2.19%), and effort (1.46%). These are generally estimated at the beginning of a software project, by means of historical data and reflect the team's previous results to provide stakeholders with estimated values for project completion and price. Furthermore, these metrics serve as milestones that can enable the team to adjust their productivity to achieving these goals. It should be noted that these estimation metrics generally use a model as a benchmark, as was the case in paper P137, which used the COSMIC method [48] and taught this reference model through a serious game.

4.13. MEAQ3: Which Areas of Measurement (Collect, Store, Analyse, and Report) Have Been Covered by the Schemes? Only two of all primary studies, P16 and P135, covered

Table 17: Metrics found in the primary studies.

Metrics	Primary studies	Percentage
Average time for completion of tasks	P29	0.73%
Burndown chart	P129 P130	.1.46%
Commits per developer	P94	0.73%
Cumulative Flow Diagram	P44	0.73%
Construction time	P10	0.73%
Cost Performance Index (CPI)	P16	0.73%
Cyclomatic complexity (CC)	P22 P134	1.46%
Defects in requirements	P93	0.73%
Defects injected per phase	P10	0.73%
The density of comment lines for documentation	P112	0.73%
The density of duplication code	P63 P112	1.46%
Real effort	P16	0.73%
Error rate per developer	P17	0.73%
Estimate of effort/COSMIC	P131 P137	1.46%
Estimate of cost	P72 P116	1.46%
Estimate of risk	P72	0.73%
Estimate of task time spent	P129	0.73%
Estimate of time	P72 P116 P129 P131	2.92%
Estimation of size	P43 P116 P124	2.19%
Leadtime	P44	0.73%
Lines of code (LOC)	P17 P63 P89 P112 P134	3.65%
Mean time to repair (MTTR)	P133	0.73%
Number of changes made to the project	P108	0.73%
Number of comments received during the code reviews	P108	0.73%
Number of comments written during the code reviews	P108	0.73%
Number of completed tasks	P29 P99	1.46%
Number of errors and warnings in “builds”	P19	0.73%
Number of flawless changes	P108	0.73%
Number of items approved or rejected at the retrospective meetings	P29	0.73%
Number of known errors	P17 P112	1.46%
Number of patch sets in changes made by the author	P108	0.73%
Number of peer code reviews carried out for a single change	P108	0.73%
Number of reopened and persistent tasks	P107	0.73%
Number of repeated code reviews carried out	P108	0.73%
Number of tasks per sprint and (un)resolved tasks	P107	0,73%
Number of unknown errors	P17	0.73%
Designing one’s own metrics in the classroom or ludic metrics for teaching	P135 P136	1.46%
Participation in the daily Scrum meeting	P29	0.73%
Percentage of completion	P17	0.73%
Coverage percentage of the test case of the branch	P63 P99	1.46%
Phase time	P10	0.73%
Productivity	P12 P17 P93	2.19%
The ratio of project activities	P131	0.73%
The ratio of project activities correctly identified	P131	0.73%
The ratio of project activities of actual and estimated time effort	P131	0.73%
Schedule Performance Index (SPI)/SPI variation	P16	0.73%
Schedule variation	P16	0.73%
Sprint time	P29	0.73%

Table 17: Continued.

Metrics	Primary studies	Percentage
Sprint velocity	P34 P59 P98 P107	2.92%
Team work self-evaluation score	P112	0.73%
Total errors per phase	P10	0.73%
Unit test coverage	P112	0.73%

all the stages of the measurement process with clarity and in-depth.

P16 presented a serious game that simulated the adoption of metrics in a small software company. During the game, an expert asked a number of questions in the form of a quiz designed to analyse the situation and assist the player to choose the best metrics. The metrics that were collected and stored were the Schedule Performance Index (SPI), Cost Performance Index (CPI), level of effort activity vs. planned effort, schedule variation, and SPI variation. Following this, when conducting the analysis, the charts that best suited these metrics were chosen in the game, for example, the Gantt chart was chosen to represent schedule variation. After this analysis of the chart, the results were reported to the development team and stakeholders.

On the other hand, P136 proceeded with these stages through a teaching methodology called Dojo Handori, and this process was carried out in the classroom. It involved creating a measurement plan by means of the Goal-Question-Indicator-Metric (GQIM) method that was used to measure features related to ISO 25010 that deal with the quality of a software product. The software product that was chosen was the serious game put forward in article P16 which raises problems related to software quality. Right after the measurement plan was created, the metrics established that they should be collected and stored in a spreadsheet during the lesson. Finally, the students created a report based on the analysis of the collected metrics and spoke to the room instructor about possible improvements in the chosen game. In other words, all the stages of the measurement process were covered.

4.14. MEAQ4: What Were the Elements, Mechanics, Dynamics, and Genres of the Games Covered by the Schemes? The purpose of this research question is to find evidence of the customary practices in the teaching or training of the software measurement process. Table 18 lists the genres, dynamics, mechanics, and game elements identified in the primary software measurement studies.

As can be seen in Table 18, the majority of the works (57.78%) dealt with gamification, while the largest proportion of the serious games, which were the rest of these works, were in the simulation genre. This can be understood, as an attempt by the authors to create a safe environment that allows the learning of the selected topic through trial and error. In other words, the user is not severely penalized for his failures, but on the contrary, is encouraged to fail until he reaches the desired result, this being one of the means that serious games surpass other approaches.

As expected, the most widely used game elements in the primary papers/articles were the PBL triad: points with 66.67%, badges 33.33%, and leaderboard 39.39%. Most examples of gamification make use of this triad, as there are many benefits to be derived from adopting these elements. One of the main purposes of the point system is to provide immediate feedback to users, or in other words, it allows the player's performance to be monitored. This suggests that awarding points is an excellent way to represent the player's progress and provides an accurate metric to show the balance of the approach. Similarly, badges offer the following benefits: they visually represent something achieved by the user. In other words, they are a way of visualizing the user's progress and setting objectives or "milestones" that can be achieved. In addition, leaderboard publicly shows the progress of the players and can serve as a motivating factor in encouraging players to rise to greater heights. However, as stated in the book by [15] although the PBL triad is an excellent starting point, it is strongly recommended that other elements in addition to PBL are used to achieve more significant results and a greater diversity of gamification.

In addition, it is worth mentioning that the elements of the games often recurred several times in the same scheme. An example of this is study P29, which adopts the RUPGY approach, which is the application of dynamics, mechanics, and game elements. This is a kind of gamification that is aimed at motivating a development team, by giving visibility to software processes through the metrics used by gamification, and allowing the software development with Scrum to be more attractively displayed in the form of a game. It included achievements and badges based on historical data from a software house where this scheme was applied. In addition, the game elements used in this scheme were as follows: avatar, collections, achievements, badges, levels, points, leaderboard, and teams. The present elements lead the mechanics of competition and/or cooperation, achievements, rewards, and feedback. And such mechanics lead to the progression, relationships, and narrative dynamics. Finally, by way of illustration, there was a challenge called Clockwork Developer, which was an achievement based on the number of tasks that the developer completed during a sprint. This achievement had three levels of completion, i.e., related badges, which were as follows: bronze (50%), silver (75%), and gold (100%). Thus, this achievement gave visibility to the most productive developers and could be used as a balancing parameter in the training of teams.

5. Discussion

Studies that addressed SPI or software measurement through serious games have made a number of contributions in

Table 18: Game elements identified in the primary software measurement studies.

Game genre	Primary studies	Percentage
Adventure	P137	3.03%
Board	P44 P63	6.06%
Gamification	P10 P19 P22 P34 P43 P53 P59 P89 P94 P107 P108 P112 P129 P130 P131 P133 P134 P135 P136	57.58%
RPG	P29 P99	6.06%
Simulation	P12 P16 P17 P72 P93 P98 P116 P124 P136	27.27%
Game elements	**Primary studies**	**Percentage**
Achievements	P29 P99 P107	9.09%
Avatar	P17 P29 P94	9.09%
Badges	P22 P29 P63 P99 P107 P108 P112 P129 P131 P135 P136	33.33%
Collections	P29	3.03%
Gift	P19	3.03%
Leaderboard	P12 P19 P22 P29 P53 P63 P89 P94 P112 P129 P134 P135 P136	39.39%
Levels	P29 P53 P99 P107 P131	15.15%
Not reported	P10 P34 P43 P93 P98 P116 P124 P130 P137	27.27%
Points	P12 P16 P17 P19 P22 P29 P44 P53 P63 P72 P89 P94 P99 P107 P108 P112 P129 P131 P133 P134 P135 P136	66.67%
Task	P94	3.03%
Team	P29 P59	6.06%
Virtual goods	P107 P112	6.06%
Visual graph	P129	3.03%
Game mechanics	**Primary studies**	**Percentage**
Challenges	P22 P29 P53 P129	12.12%
Chance	P12 P17 P44	9.09%
Cooperation and competition	P19 P29 P34 P53 P63 P107 P112 P129 P130 P131 P133 P134 P135 P136	42.42%
Feedback	P12 P53 P17 P22 P29 P53 P59 P72 P99 P129 P134 P135	36.36%
Not reported	P10 P16 P43 P89 P93 P94 P98 P108 P116 P124 P137	33.33%
Resource acquisition	P12 P17	6.06%
Rewards	P19 P22 P29 P89 P133	15.15%
Turns	P44	3.03%
Win state	P89	3.03%
Game dynamics	**Primary studies**	**Percentage**
Emotions	P112 P136	6.06%
Narrative	P12 P16 P17 P29 P34 P59 P72 P94 P98 P137	30.30%
Not reported	P10 P43 P89 P93 P108 P116 P124 P131 P133	27.27%
Progression	P12 P16 P17 P22 P29 P53 P63 P107 P129	27.27%
Relationships	P12 P19 P29 P44 P59 P63 P98 P99 P107 P134 P135 P136	36.36%
Restrictions	P12 P130	6.06%

various contexts, including education, and industry. Their main achievement is to create a simulating environment in which students can apply their theoretical knowledge in a practical way and through this means learning by doing. As can be seen in the Dale cone in Figure 1, learning is theoretically more effective when carried out in an atmosphere that simulates actual experience. As can be seen in article P16, educators can benefit from using this kind of tool as a pedagogical aid for the subject of software measurement because it is very difficult to cover all the software engineering processes. Moreover, when being involved in serious games, students can learn measurement practices in their spare time.

As in the case of P44, involvement with serious gameplay through classroom dynamics has led students to understand the CFD (Cumulative Flow Diagram), which enables them to become aware of the importance of measuring a software process. As a result, the students can understand how the CFD evidences the WIP (work in progress) in the course of time and hence the bottlenecks of the process. This approach has also taught students how to draw a CFD chart and how to interpret it when searching for information, such as the average leadtime.

Both article P16 and paper P44 followed a pattern that was evident in GQ8, which was to ask questions about the effectiveness of learning through gamification or serious games when compared with traditional learning. Although the application of serious games theoretically has the capacity to pass on knowledge in a more efficient way, this effect cannot be determined by statistical data, and neither of the papers/articles provided evidence of statistical gains when contrasted with the traditional method of teaching. Several factors may have had an influence on this result, one of them being the limited attractiveness of the planned games. This is because the aesthetic appeal of the games that were shown is far less than the games offered by the game industry, and the players are already accustomed to a high standard of quality. Another point to consider is that, according to McGonigal [49], games should be a voluntary experience, while in the case of the analysed works; it is evident that they have been introduced in a mandatory way, which can cause the users to lose interest. Moreover, no simulation model can accurately replicate the real world, and this is a factor that should be taken into account. To sum up, the use of games provides solid support for teaching, but they should not be used without being supplemented by other methods, whether traditional methods or otherwise.

With regard to gamification, in P19, the author states that "people love competition, it is the fuel that drives them to follow organizational processes and their daily activities with greater impetus and will power." If gamification can act as a driving force for the employees and encourage them to carry out their ordinary tasks, it means they will be more willing to comply with an organizational process and their obligations or duties will be more palatable, that is, more pleasant and subject to less resistance. In fact, gamification acted as a behaviour measurement tool and increased the visibility of different individuals in development teams. This attribute can be seen in paper P22, where developers working on critical areas of projects had higher points and rewards than those who only refactored simple programs. A more detailed application of this scenario was shown in paper P99, which transformed the goals inherent in software, into real achievements. Thus, encouraging the developers to be more willing to achieve them assisted the management to observe the progress of the team by being provided with instant feedback.

The intelligent application of gamification in SPI and software measurement processes had a positive effect on those involved in them. Despite this, it should be borne in mind that gamification is not just the use of gaming elements in a nongaming context. Rather, it is the intelligent use of this concept, but as [15] points out, there is a problem that arises because many gamification schemes do not bother to focus on the social, cognitive, and emotional factors that the games address. These schemes tend to be quickly ignored because they seem to be superficial and fail to catch the attention of the player. Table 19 lists the strengths, weaknesses, opportunities, and threats (SWOT) for a better understanding of this work.

Most of the studies analysed were not concerned with verifying the accuracy of their proposals. That is, not all studies used a control group and another experiment group to compare their results and verify the effectiveness of their methods. Even less was the number of studies that showed positive results when compared to the traditional teaching method. It is a common understanding that games are an optimized way to learn a new skill, especially if the target audience is the current generation. However, when comparing educational games with commercial games, it is possible to notice a great disparity in graphics, design, sound effects, ambient music, and game production. The production of a commercial game is usually done by a multidisciplinary team of experts containing dozens and even hundreds of members and has a production time of months to years of development. On the other hand, a small team develops most educational games with little diversity of experts, which generally do not have experts in game design, and short development time. Therefore, even with the undeniable potential of games to teach, far less effort is devoted to developing an educational game compared to commercial games and, consequently, the result is also below those achieved by them. It is well known that game-based education, especially the teaching of software measurement and software process improvement, has a lot to mature as a research field.

6. Threat to Validity

This section examines different threats to the validity of this systematic review of the literature, based on the four most common "threat to validity" categories [50]: internal validity, external validity, construct validity, and conclusion validity. There is also a discussion about how these threats were mitigated in this article.

6.1. Internal Validity. The internal validity refers in particular to the independent variables, that is, it questions whether the research was conducted in the correct way. An extensive protocol was designed to mitigate the causes of internal validity, and this set out the steps that had to be followed and the tasks that needed to be carried out together with descriptions of each of them. This protocol was updated and validated for each new version by the expert of software engineering, who is also the supervisor of this study. Moreover, the researchers continuously consulted the protocol to clarify certain stages of the process and to discuss any possible deviations from it.

6.2. External Validity. The external validity is concerned with determining whether this study can be replicated by other researchers and if their results are consistent with those found here. The following practices were adopted to reduce the risks posed by external validity: the creation of a protocol and constant meetings with all the researchers to check the degree of conformity to the protocol and to validate the articles that might be selected at each stage of the process. Before an article could be included, it had to be accepted by at least two of the four researchers. Thus, on the basis of these

Table 19: SWOT analysis.

SWOT analysis	
Strengths	Weaknesses
Review of five relevant scientific databases 19050 analysed papers/articles An extensive protocol was designed and followed An expert of software engineering, who is also the supervisor of this work, validated each step followed in this systematic review.	Only English and Portuguese papers/articles could be analysed The analysers had a good command of English Only 4 members of the research team conducted this review
Opportunities	Threats
Inclusion of more scientific databases	Internal validity
Expansion of the range of languages	External validity
Inclusion of paid papers/articles	Construct validity
An increase in the age range to another ten years	Conclusion validity

practices, it was possible to reduce the risk of primary articles being excluded.

6.3. Construct Validity. Construct validity concerns the measures taken to ensure they really represent what they are intended to measure, i.e., whether the data collected assists in answering the research questions. The research questions were devised in conjunction with an educational expert, who is also responsible for game-based educational strategies. This researcher has extensive experience in these areas and collects publications, such as those listed by [38, 49, 51]. Frequent consultations were conducted with this researcher with the aim of overcoming construct validity problems.

6.4. Conclusion Validity. The purpose of the conclusion validity is to determine whether the conclusions are correctly supported by the collated data. There is a threat to this validity in the data extraction stage, since many articles found did not answer the research questions directly, and so it was necessary to infer the information. The authors met to address these issues and discussed the reliability of the inference, and if two authors agreed about this, this information was included.

7. Related Works

During the systematic analysis of the collected papers/articles, some systematic reviews of the literature were found that addressed similar topics or that covered topics of interest in this study. In the present systematic review of the literature, the objective was to find and extract information from gamified schemes or serious games that could be used for teaching software process improvement or software measurement. In addition, most of the works found were about the use of games for teaching software processes, and the most significant of the related works are examined below.

The study by [53] provides a review that investigates the method of evaluating serious games that are aimed at training students in the subject of project management. This review has obtained 102 primary studies and summarizes information about the following: methods, processes, assessment techniques, the field of application (education, health, and welfare), different types of serious games, the number of users, and the main features of the games evaluated. The results shown in the article are useful since they assist the process of evaluating serious game projects, especially those aimed at project management, but not restricted to this process.

In the work of [54], a systematic mapping procedure was conducted that involved selecting 173 primary papers/articles with the objective of classifying works that referred to practical experiences in the teaching of software engineering. The systematic mapping sought to answer the following key questions: What are the main approaches used to address practical experiences in software engineering education? Is there an emerging trend in addressing such a need? What software process models are used to support hands-on experience in software engineering courses? Have universities changed the means of conducting these experiments over the years? What are the main forums for seeking information on practical approaches to teaching software engineering?

As a result, the most frequent practical experiences were determined and classified, such as gaming learning, a case study, simulations, inverted classrooms, project maintenance, service-learning, and open source development. Additionally, the authors found that methodologies for guiding the development of projects only appeared in 40% of the studies and mostly involved flexible methods. In conclusion, the author provided evidence to show that there is a clear concern about how to adopt practical approaches in the teaching of software engineering, and also, there are also numerous alternative ways of filling this gap. Among these, it is worth mentioning the current trend for a method of teaching based on games.

In [55] a systematic review of the literature was conducted based on papers/articles written in the period 2000-2015, where a total number of 53 primary papers/articles were analysed relating to games for the teaching of software engineering. These were classified as follows: games for the students to play, games for students to carry out as a project, curricular projects, innovative web design tools, a framework, suggestions, and other factors. In short, this study has shown that software engineering and games are being approached in different ways and that investment in software engineering education will have an influence on future software engineers, by enabling them to achieve the broader goal of software process improvement.

The study by [56] detected 42 primary studies between 1992 and 2013 that made use of Software Process Simulators for teaching software engineering. As a result, the authors confirmed that there had really been a positive impact on the teaching of software processes, and in addition, simulators were provided with their individual capacities and features and their respective evaluations. The studies that addressed the question of software process improvement and software measurement were also analysed by our systematic review of the literature.

The work of [18] outlined a systematic mapping of the field of gamification when applied to software engineering and to characterize its state-of-the-art. As a result, 29 primary studies, published between 2011 and 2014, were identified. These were classified and analysed in terms of the software process area addressed, the gamification elements used, the type of search method followed, and the type of forum in which the paper/article was published. As a result, the research discovered that the most frequent process areas were system implementation, collaboration, project planning, project control, and evaluation. With regard to the last of these, 5 works were found that make use of gamification for project management, and these were included in the present systematic review of the literature. It was also stated by the authors that most of the gamification projects were based on PBL (points, levels, and leaderboards), which for some critics like Margaret Robertson [15] show they are being used for superficial purposes.

Thus, we can point out that the differentials of this work in relation to the others discussed in this section are as follows: initially, this work focuses on a research field (software process improvement and software measurement) not found in other SRL, which allows an analysis of the application of these fields in serious games and gamification, another point concerns the analysis and discussion of a large number of general and specific questions related to the measurement and process improvement, enabling an identification of the current scenario of their application in teaching from games and gamification, and finally, the details of all the steps followed for the execution of the SRL, enabling its reapplication in other contexts, since the related works do not detail them.

8. Conclusion and Future Work

This systematic review of the literature was aimed at examining genres, dynamics, mechanics, and game elements present in gamification projects and serious games for teaching software process improvement, with an emphasis on software measurement. A protocol was used for this that allowed 137 primary articles to be analysed from a total of 19050 articles in the IEEE, Scopus, ACM, Ei compendex, Web of Knowledge, and Science Direct databases. In addition, this study had the following objective: to present the state-of-the-art at using gamification and serious games in the teaching of software measurement and software process improvement programs. To achieve this objective, we established the following steps:

(i) To make prior definitions, such as research restrictions, criteria for inclusion and exclusion of primary studies, and quality criteria for these studies, among others, as guidelines for the systematic review of the literature

(ii) To conduct a systematic review of the literature following the specifications previously established in the protocol

(iii) To analyse the results of the review from the characterization of the selected studies

There has been a growing interest in addressing software processes through the use of serious games, and in 2016, it comprised the largest output of scientific studies in this area. However, there was also a notable underexploitation of the software measurement process among the researched studies. As reported in this SRL, only 4% (P16, P29, P134, P135, P136, and P137) of the studies concentrated exclusively on this process and its different stages. Thus, it is clear that this area needs more exploratory studies to teach this process.

In addition, the most widely used genre was simulation, which entails creating a simulated environment for the practical application of the students' theoretical knowledge. Relationships were the most widely used gamification dynamics, since this category was designed to create a relationship between those involved in competition or cooperation and thus to engage the users in social factors. And as a reflection of the relationship dynamic, the most used mechanics were feedback, cooperation, and competition. Finally, the most widely used game elements were points, leaderboards, badges, and avatars.

With regard to future work, this review will serve as input to guide the development of an educational tool for teaching the measurement process in the context of software projects. This kind of tool will be based on the elements of games, good practices, and strengths that were found in the primary studies. In addition, for the purposes of validation, the tool will be evaluated with the user to decide whether the game-based project (gamification and serious games) can be regarded as appropriate in terms of relevance of content, correctness, degree of difficulty, method of teaching, and duration within the context for which it is intended.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to thank the Coordination for the Improvement of Higher Education Personnel (CAPES) for awarding a doctoral scholarship to Graduate Program in Computer Science (PPGCC) at Federal University of Pará (UFPA) in Brazil. This scholarship was used to foster the research that led to the development of the article entitled "Teaching Method for Software Measurement Process based on Gamification or Serious Games: A Systematic Review of the Literature," which was supervised by Professor and Doctor Sandro Ronaldo Bezerra Oliveira. This research belongs to the SPIDER-UFPA (https://spider.ufpa.br) research group. The authors would also like to thank the Dean of Research and Graduate Studies at the Federal University of Pará (PROPESP/UFPA) for providing financial support through the Qualified Publication Support Program (PAPQ) Public Notice 06/2021.

References

[1] R. S. Pressman, *Software Engineering: A Practitioner's Approach. 7a edição*, McGraw Hill, Ed, 2011.

[2] Softex, "MPS.BR - Melhoria de Processo do Software Brasileiro guia geral mps de software," 2016, 2018, https://www.softex.br/wp-content/uploads/2016/04/MPS.BR_Guia_Geral_Software_2016-com-ISBN.pdf.

[3] C. A. Dekkers and P. A. McQuaid, "The dangers of using software metrics to (mis) manage," *IT Professional*, vol. 4, no. 2, pp. 24–30, 2002.

[4] M. Kasunic, "The state of software measurement practice: results of 2006 survey," in *Procedures for performing systematic reviews*, B. Kitchenham, Ed., vol. 33, pp. 1–26, Keele University, Keele, UK, 2006.

[5] C. Jones, "Software metrics: good, bad, and missing," *Computer*, vol. 27, no. 9, pp. 98–100, 1994.

[6] M. Díaz-Ley, F. García, and M. Piattini, "Implementing a software measurement program in small and medium enterprises: a suitable framework," *IET Software*, vol. 2, no. 5, pp. 417–436, 2008.

[7] A. Gopal, M. S. Krishnan, T. Mukhopadhyay, and D. R. Goldenson, "Measurement programs in software development: determinants of success," *IEEE Transactions on Software Engineering*, vol. 28, no. 9, pp. 863–875, 2002.

[8] 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*, vol. 26, no. 1, pp. 30–43, 2006.

[9] C. G. von Wangenheim, T. Punter, A. Anacleto, and R. Anacleto, "Software measurement for small and medium enterprises - a Brazilian-German view on extending the GQM method," in *In Proceeding 7th International Conference on Empirical Assessment in Software Engineering (EASE)*, 2003.

[10] S. Löper and M. Zehle, *Evaluation of Software Metrics in the Design Phase and their Implication on Case Tools*, Blekinge Institute of Technology, 2003.

[11] M. Diaz-Ley, F. García, and M. Piattini, "Mis-pyme software measurement maturity model-supporting the definition of software measurement programs," in *International conference on product focused software process improvement*, pp. 19–33, Berlin, Heidelberg, 2008.

[12] J. J. Trienekens, R. J. Kusters, M. J. van Genuchten, and H. Aerts, "Targets, drivers and metrics in software process improvement: results of a survey in a multinational organization," *Software Quality Journal*, vol. 15, no. 2, pp. 135–153, 2007.

[13] M. Villavicencio and A. Abran, "Towards the development of a framework for education in software measurement," in *Software Measurement and the 2013 Eighth International Conference on Software Process and Product Measurement (IWSM- MENSURA)*, pp. 113–119, Ankara, Turkey, 2013.

[14] E. Herranz, R. C. Palacios, A. de Amescua Seco, and M. Yilmaz, "Gamification as a disruptive factor in software process improvement initiatives," *Journal of Universal Computer Science*, vol. 20, no. 6, pp. 885–906, 2014.

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

[16] J. S. Breuer and G. Bente, "Why so serious? On the relation of serious games and learning," *Eludamos. Journal for Computer Game Culture*, vol. 4, no. 1, pp. 7–24, 2010.

[17] M. Zyda, "From visual simulation to virtual reality to games," *Computer*, vol. 38, no. 9, pp. 25–32, 2005.

[18] O. Pedreira, F. García, N. Brisaboa, and M. Piattini, "Gamification in software engineering - a systematic mapping," *Information and Software Technology*, vol. 57, pp. 157–168, 2015.

[19] S. N. Mafra and G. H. Travassos, *Estudos Primários e Secundários Apoiando a Busca por Evidência em Engenharia de Software*, vol. 687, no. 6, 2006Relatório Técnico, RT-ES, 2006.

[20] G. O'Regan, *Introduction to Software Process Improvement*, Springer Science & Business Media, 2010.

[21] CMMI Institute, *CMMI for Development, version 2.0*, CMMI Institute, 2018.

[22] A. Rocha, G. Souza, and M. Barcellos, *Medição de Software e Controle Estatístico de Processos. Ministério da Ciência*, Tecnologia e Inovação; Secretaria de Política de Informática, 2012.

[23] T. DeMarco, *Controlling Software Projects: Management, Measurement, and Estimates*, Prentice-Hall PTR, 1986.

[24] N. E. Fenton and S. L. Pfleeger, *Software Metrics: A Rigorous and Practical Approach, Revised*, CRC press, 1998.

[25] C. Jones, *Applied Software Measurement: Global Analysis of Productivity and Quality*, McGraw-Hill Education Group, 2008.

[26] AMCHAM, "American Chamber of Commerce only 14% of companies are completely satisfied with the results of it measurement systems," 2012, 2018, http://www.amcham.com.br.

[27] M. Villavicencio and A. Abran, "Software measurement in software engineering education: a comparative analysis," in *International Conferences on Software Measurement IWSM/-MetriKon/Mensura*, pp. 633–644, 2010.

[28] R. Koster, *Theory of Fun for Game Design*, O'Reilly media, Inc., 2013.

[29] S. Bjork and J. Holopainen, *Patterns in Game Design*, Charles River Media, Inc, 2004.

[30] J. Hamari, J. Koivisto, and H. Sarsa, "Does gamification work? A literature review of empirical studies on gamification," in *System Sciences (HICSS), 2014 47th Hawaii International Conference*, pp. 3025–3034, Waikoloa, HI, USA, 2014.

[31] 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, pp. 1–10, Valparaiso, Chile, 2016.

[32] E. Dale, *Audiovisual Methods in Teaching*, The Dryden Press, 1969.

[33] J. Hamilton and S. Tee, "Smart utilization of tertiary instructional modes," *Computers & Education*, vol. 54, no. 4, pp. 1036–1053, 2010.

[34] U. Aydan, M. Yilmaz, P. M. Clarke, and R. V. O'Connor, "Teaching ISO/IEC 12207 software lifecycle processes: a serious game approach," *Computer Standards & Interfaces*, vol. 54, pp. 129–138, 2017.

[35] B. Kitchenham, *Guidelines for Performing Systematic Literature Reviews in Software Engineering*, Keele University, Keele, UK, 2007.

[36] Spider, "Software process improvement: development and research," 2018, 2018, https://www.spider.ufpa.br/.

[37] S. Beechan, N. Baddoo, T. Hall, H. Robinson, and H. Sharp, "Motivation in software engineering: a systematic literature review," in *Information and Software Technology*, vol. 50, pp. 860–878, Elsevier, 2007.

[38] R. O. Chaves, C. G. von Wangenheim, J. C. C. Furtado, S. R. B. Oliveira, A. Santos, and E. L. Favero, "Experimental evaluation of a serious game for teaching software process modeling," *IEEE Transactions on Education*, vol. 58, no. 4, pp. 289–296, 2015.

[39] A. Dorling and F. McCaffery, "The gamification of spice," in *International Conference on Software Process Improvement and Capability Determination*, pp. 295–301, Berlin, Heidelberg, 2012.

[40] J. M. Fernandes and S. M. Sousa, "Playscrum-a card game to learn the scrum agile method," in *Games and Virtual Worlds for Serious Applications (VS-GAMES), 2010 Second International Conference on. [S.l.]*, pp. 52–59, Braga, Portugal, 2010.

[41] C. R. Prause, *Improving the internal quality of software through reputation-based gamification, [Ph.D thesis]*, Citeseer, 2013.

[42] A. Bartel and G. Hagel, "Gamifying the learning of design patterns in software engineering education," in *Global Engineering Education Conference (EDUCON), 2016 IEEE*, pp. 74–79, Abu Dhabi, United Arab Emirates, 2016.

[43] C. G. Von Wangenheim, M. Thiry, and D. Kochanski, "Empirical evaluation of an educational game on software measurement," *Empirical Software Engineering*, vol. 28, no. 9, pp. 418–452, 2009.

[44] C. S. Costa, *An Evidence-Based Approach to Project Management in Distributed Software Development [M.S thesis]*, Graduate Program in Computer Science - Federal University of Pernambuco, Recife, 2010.

[45] J. C. Herz, *Joystick Nation: How Videogames Ate our Quarters, Won our Hearts, and Rewired our Minds*, Brown & Co. Inc., Little, 1997.

[46] M. Mizukami, *Ensino: As Abordagens do Processo*, Epu, São Paulo, 1986.

[47] D. Stankovic, V. Nikolic, M. Djordjevic, and D.-B. Cao, "A survey study of critical success factors in agile software projects in former Yugoslavia IT companies," *Journal of Systems and Software*, vol. 86, no. 6, pp. 1663–1678, 2013.

[48] B. Marín, O. Pastor, and A. Abran, "Towards an accurate functional size measurement procedure for conceptual models in an MDA environment," *Data & Knowledge Engineering*, vol. 69, no. 5, pp. 472–490, 2010.

[49] J. McGonigal, *Reality Is Broken: Why Games Make us Better and how they Can Change the World*, Penguin, 2011.

[50] M. R. de Souza, L. Veado, R. T. Moreira, E. Figueiredo, and H. Costa, "A systematic mapping study on game-related methods for software engineering education," *Information and Software Technology*, vol. 95, pp. 201–218, 2018.

[51] I. Elgabry and S. R. B. Oliveira, "Gamification and evaluation of the use of agile tests in software quality subjects: the application of case studies," in *13th International Conference on Evaluation of Novel Approaches to Software Engineering*, pp. 416–423, Belém, Pará, Brazil, 2018.

[52] L. S. Furtado and S. R. B. Oliveira, "A teaching method for software measurement process based on gamification," *ICSEA*, vol. 15, no. 16, p. 12, 2017.

[53] A. Calderón and M. Ruiz, "A systematic literature review on serious games evaluation: an application to software project management," *Computers Education*, vol. 87, pp. 396–422, 2015.

[54] M. R. Marques, A. Quispe, and S. F. Ochoa, "A systematic mapping study on practical approaches to teaching software engineering," in *2014 IEEE Frontiers in education conference (FIE) proceedings*, pp. 1–8, Madrid, Spain, 2014.

[55] M. Kosa, M. Yilmaz, R. V. O'Connor, and P. M. Clarke, "Software engineering education and games: a systematic literature review," *Journal of Universal Computer Science*, vol. 22, no. 12, pp. 1558–1574, 2016, 1st gamification workshop at the 22nd EuroAsiaSPI conference, Ankara Univ, Ankara, TURKEY, SEP 30, 2015.

[56] S. Jiang, H. Zhang, C. Gao, D. Shao, and G. Rong, "Process simulation for software engineering education," in *Proceedings of the 2015 International Conference on Software and System Process, ICSSP 2015*, pp. 147–156, New York, NY, USA, 2015.