Calderón, A., Ruiz, M., & O'Connor, R. V. (2018). A multivocal literature review on serious games for software process standards education. *Computer Standards & Interfaces*, *57*, 36-48.

A MULTIVOCAL LITERATURE REVIEW ON SERIOUS GAMES FOR SOFTWARE PROCESS STANDARDS EDUCATION

Alejandro Calderón[a, c], Mercedes Ruiz[a, d], Rory V. O'Connor[b, e]

[a] University of Cádiz, Cádiz, Spain

[b] Dublin City University, Dublin, Ireland.

[c] E-mail: alejandro.calderon@uca.es (corresponding author)

[d] E-mail: mercedes.ruiz@uca.es

[e] E-mail: rory.oconnor@dcu.ie

Abstract

Context: The interest in the use of serious games as learning resources for software process standards education and training has increased significantly in recent years.

Objective: The main purpose of this work is to record, analyze and characterize the state of the art related to serious games for software process standards education with the goal of identifying the current serious games in terms of the scope, their main features and the perceived benefits of integrating them in software process education, as well as, identifying new research opportunities.

Method: The study was conducted as a multivocal literature review that follows a predefined procedure in which studies from the scientific and grey literature are analyzed.

Results: A new selection process within the search strategy was defined to conduct this review. 190 papers were retrieved from the literature and 7 papers were selected as primary studies. Our multivocal literature review identified six different serious games for software process education, at the same time analyzed the main methods used to assess them as well as their main outcomes as learning resources.

Conclusion: The results of this review reveal that serious games have potential as supporting tools for software process standards education, but that more research and experimental outcomes are needed in order to observe the full potential of serious games as learning resources.

Keywords

Serious game; Systematic literature review; Software process standard; Multivocal literature review; Education;

1. Introduction

Software development is a critical activity that is supported by standards such as ISO/IEC 12207, ISO/IEC 29110, with the goal of helping companies create software systems, products or services with defined quality parameters. Such standards define a set of processes and activities needed to model and address a broader range of issues that take place during the software development. Regarding the relevance of software processes in software development success, equal attention should be given to the training that practitioners and future software engineers receive in this scope.

As a consequence, in recent years an increased interest has been observed in the development and use of new methods and techniques to teach in a highly practical way, promote active and interactive learning, increase the motivation and engagement of learners and design new training strategies to train software process practitioners as skilled and qualified professionals (Calderón & Ruiz, 2016) (Kosa, Yilmaz, O'Connor, & Clarke, 2016). Among these new approaches, we can find the use of serious games (Kosa, Yilmaz, O'Connor, & Clarke, 2016). Serious games (SGs), also called training or educational games, are designed for purposes other than pure entertainment (Abt, 2002). As learning tools, these kinds of games are powerful learning

resources that allow participants to experiment, learn from their own mistakes and acquire experience, in a safe way within risky environments.

Taking into account the recent interest of using SGs, we can find some systematic literature reviews (SLR) related to the scope of software processes such as the SLR conducted by (Kosa, Yilmaz, O'Connor, & Clarke, 2016) that presents an overview of the different uses of games in the scope of software engineering education or the mapping study performed by (Heredia, Colomo-Palacios, & Amescua-Seco, 2015) that structures the state of the practice on software process education to identify best practices and find new challenges in that field. However, we cannot find any SLR that provides an overview of the SGs to teach in software process standards (SPS). For that, we conducted this study with the goal of analyzing the current state of the art in applying SGs for software process standards education and get the main insights of this topic.

The study was conducted as a kind of SLR called Multivocal Literature Review (MLR) in which data from the scientific and grey literature are included (Garousi, Felderer, & Mäntylä, 2016). The inclusion of grey literature can give substantial benefits in certain areas of software engineering, can provide insight into the "state of the practice" in software engineering and brings forward certain challenges as evidence in them is often experience and opinion based (Garousi, Felderer, & Mäntylä, 2016). For that reason, in this work, we have defined a new procedure for performing MLRs based on the best practices and guidelines proposed by Kitchenham (MacDonell, Shepperd, Kitchenham, & Mendes, 2010) (Kitchenham & Charters, 2007) and the methods and procedures of SLRs that take into account the grey literature (Garousi, Felderer, & Hacaloğlu, 2017) (Calderón & Ruiz, 2015).

The contributions of this paper are: (i) the definition of a new procedure for performing a MLR, (ii) a case of application of our procedure to identify the state of the art related to the scope of SGs for software process standards education, and (iii) an overview of the features and perceived benefits of the current SGs for software process standards education.

The structure of the paper is as follows. Section 2 presents the background of this work and Section 3 analyzes the works related with our proposal. In Section 4, the method used for conducting our MLR is introduced. Section 5 shows the results of the review, and Section 6 offers discussions on these results. Section 7 shows the main threats found to the validity of this study. Finally, Section 8 summarizes the paper and presents our conclusions and future works based on the findings obtained.

2. Background

2.1. Software Process Standards

Software engineering is a relevant field that involves the study and application of engineering principles to the design, development and maintenance of software. Its main goal is to create software systems, services or products with quality, taking into account the constraints of cost and time (Humphrey, 1995). To achieve this objective, software development activity is usually supported by international standards that provide a set of software processes for covering all the software life cycle and define the activities needed to conceive, develop, deploy, and maintain a software system, product or service.

Regarding the software process standards, we can highlight the following as the main ones:

- ISO/IEC 12207 (ISO/IEC, 2008) is an international software engineering standard that establishes a common framework for software life cycle processes. It contains processes, activities, and tasks that are to be applied during the acquisition of a software product or service and during the supply, development, operation, maintenance and disposal of software products. Hence, the standard provides a set of processes that cover the software life cycle from conception to the end of product.
- ISO/IEC 29110 (ISO/IEC, 2016) is an international systems and software engineering standard that establishes lifecycle profiles for Very Small Entities (VSEs). The standard provides frameworks and guides for VSEs that do not have experience in applying or adapting the processes, activities, tasks and outcomes of ISO/IEC 12207 or ISO/IEC 15288 standards to their specific needs.

- ISO/IEC 15504 or SPICE (ISO/IEC, 2004) is an international standard for process assessment and improvement that provides a framework for the assessment of processes. This framework can be used by organizations involved in planning, managing, monitoring, controlling, and improving the acquisition, supply, development, operation, evolution and support of products and services. This international standard was initially derived from process lifecycle standard ISO/IEC 12207. On the other hand, it must be taken into account that the ISO/IEC 330xx family of standards is intended to supersede the ISO/IEC 155xx family.

2.2. SLR & MLR

A systematic literature review is a method used for identifying, analyzing and interpreting all available research relevant to a particular research question, topic area, or phenomenon of interest. Through a SLR, researchers are able to collect all the scientific literature related to a specific knowledge scope with the goal of observing its state of the art. We refer to the scientific literature as the academic and research works published in journals or conferences that are usually controlled by commercial publishers and passed a peer-review process. All the individual works that contribute to a SLR, are called primary studies. On the other hand, a SLR is a form of secondary study (Kitchenham & Charters, 2007).

The grey literature refers to print or electronic literature that is produced by government, academia, business and industry and is not controlled by commercial publishers (Auger, 1998). This can include materials such as unpublished studies or doctoral dissertations, conference proceedings, book chapters, government and agency reports, as well as blogs posts, white papers and presentation videos. In the last years, the number of works that complement their SLRs with studies from the grey literature (Garousi, Felderer, & Hacaloğlu, 2017) (Calderón & Ruiz, 2015) has risen considerably. This kind of SLRs in which data from the scientific and grey literature are included receives the name of Multivocal Literature Review (Tom, Aurum, & Vidgen, 2013).

Summarizing, SLR and MLR are two different methods to identify, analyze and interpret the state of the art of a particular research question, topic area, or phenomenon of interest. Both methods allow researchers to collect all the scientific literature related to a specific knowledge, but the main difference is that only the MLR method allows researchers to include and complete their insight of a topic with the information retrieved from the grey literature.

3. Related Works

This MLR focuses on the field of SGs as learning resources for teaching in software process standards education. Before performing this study, we conducted an initial study to identify the existing secondary studies related to our research scope. In order to obtain the maximum information about this topic, we performed automatic searches in different electronic databases (Wiley Online Library, Scopus, ACM Digital Library, IEEE Xplore, SpringerLink, ISI Web of Science and Science Direct) which are the most used and popular academic databases in the domain of software engineering (MacDonell, Shepperd, Kitchenham, & Mendes, 2010). These searches were conducted in April 2017 using the search string "(A1 **OR** A2 **OR** A3 **OR** A4 **OR** A5 **OR** A6 **OR** A7) **AND** B1 **AND** (C1 **OR** C2 **OR** C3 **OR** C4 **OR** C5 **OR** C6 **OR** C7 **OR** C8 **OR** C9)", where the search terms are shown in Table 1.

Table 1. Search terms to identify related secondary studies.

A1. Software process	B1. Game	C1. Systematic review
A2. ISO/IEC 12207		C2. Research review
A3. ISO 12207		C3. Systematic overview
A4. ISO/IEC 29110		C4. Systematic literature review
A5. ISO 29110		C5. Systematic mapping
A6. ISO/IEC 15504		C6. Mapping study
A7. ISO 15504		C7. Systematic mapping study
		C8. Multivocal review
		C9. Multivocal literature review

Calderón, A., Ruiz, M., & O'Connor, R. V. (2018). A multivocal literature review on serious games for software process standards education. *Computer Standards & Interfaces*, *57*, 36-48.

As a result of this initial study, the conducted searches helped us to find the secondary studies related to our research scope. Kosa et al. (Kosa, Yilmaz, O'Connor, & Clarke, 2016) conducted a SLR related to the use of games in software engineering education that were published in the last fifteen years and classified them in five categories regarding their use in software engineering education. This study presents an overview of the different uses of games in the scope of software engineering education in order to review and characterize its state of the art, but it does not focus on identifying the specific games or analyzing their features for teaching in this scope.

Following similar objectives, Heredia et al. (Heredia, Colomo-Palacios, & Amescua-Seco, 2015) conducted a mapping study for structuring and characterizing the state of the practice on software process education to identify best practices and find new challenges in that field. Although this study attempts to understand the software process subject and interpret its needs and implementation in academia, it does not focus on the specific topic of games as learning resources for software process education.

On the other hand, we found other secondary studies related to the scope of SGs for software engineering education. These studies are:

- The systematic mapping of Pedreira et al. (Pedreira, García, Brisaboa, & Piattini, 2015), which aims at characterizing the state of the art of gamification in software engineering in order to identify gaps and opportunities for further research.
- The systematic mapping of Marques et al. (Marques, Quispe, & Ochoa, 2014), which aims at identifying and characterizing the state of the art of the different approaches to teach in software engineering.
- The SLRs of Calderón et al. (Calderón & Ruiz, 2015) and Petri et al. (Petri & von Wangenheim, 2017) that focus on identifying the state of the art of the existing methods to assess SGs for computing or software engineering education.

As we described above, some secondary studies related to SGs to teach in software engineering have been already published. The majority of these secondary studies share among them that follow the guidelines and recommendations of Kitchenham et al. (Kitchenham & Charters, 2007) and Petersen et al. (Petersen, Feldt, Mujtaba, & Mattson, 2008) for conducting their systematic or mapping reviews in software engineering. Moreover, if we focus on the software process scope, no secondary studies which tackle the application of SGs for software process standards education has been found. For that reason, in our study, we analyze this scope with the goal of analyzing and characterizing the state of the art of the SGs for software process standards education and identifying new research opportunities.

4. Method

The main purpose of our work is to observe, document, analyze and characterize the state of the art related to SGs used for software process standards education. Our goal is to identify the current SGs within the scope, their main features and the perceived benefits of integrating them in software process education, as well as, identifying new research opportunities. For that reason, this study has been performed as a MLR based on the best practices and guidelines for conducting SLRs in software engineering proposed by Kitchenham et al. (MacDonell, Shepperd, Kitchenham, & Mendes, 2010) (Kitchenham & Charters, 2007) and the methods and procedures of SLRs that take into account the grey literature (Garousi, Felderer, & Hacaloğlu, 2017) (Calderón & Ruiz, 2015).

As (Kitchenham & Charters, 2007) established, a review should comprise three stages: planning, conducting and reporting. In the planning stage, the review protocol is established and how the researchers should work and interact to conduct the review is decided. This protocol defines the procedure for executing the review and includes research questions, search and evaluation strategies, inclusion/exclusion criteria, quality assessment, data collection form and methods of analysis. The second stage focuses on executing the protocol as it has been defined. Finally, the aim of the last stage is to elaborate the final report.

Calderón, A., Ruiz, M., & O'Connor, R. V. (2018). A multivocal literature review on serious games for software process standards education. *Computer Standards & Interfaces*, *57*, 36-48.

In order to conduct our MLR, we have defined a new procedure for executing the review that establishes all the steps involved in the search and evaluation strategies to get the primary studies from both the scientific and grey literature sources. Figure 1 shows the steps followed for conducting the MLR. In the following sub-sections, we describe the activities involved in each step of this procedure, at the same time that we define the specific information related to our MLR.

4.1. Research Questions

In this step of the review, the objectives of the study are defined and the research questions that guide the review are set. The research questions help to collect all the information needed to observe and analyze the state of the art of the topic under review.

Our review aims to identify useful information about SGs for software process standards education. For that reason, the goals of this MLR are: (i) Identify the SGs for software process standards education, (ii) collect and analyze the main features of the SGs for software process standards education, (iii) classify and characterize the SGs for software process standards education according its main features, (iv) identify the main strengths and weaknesses of the SGs for software process standards education, and (v) observe the perceived benefits of the application of SGs in the context of software process standards education.

In order to achieve these goals, we defined three research questions. The research questions addressed by this work are:

RQ1. What SGs for software process standards education are available?
RQ2. How have these SGs been evaluated?
RQ3. What are the perceived benefits of using SGs in the context of software process standards education?

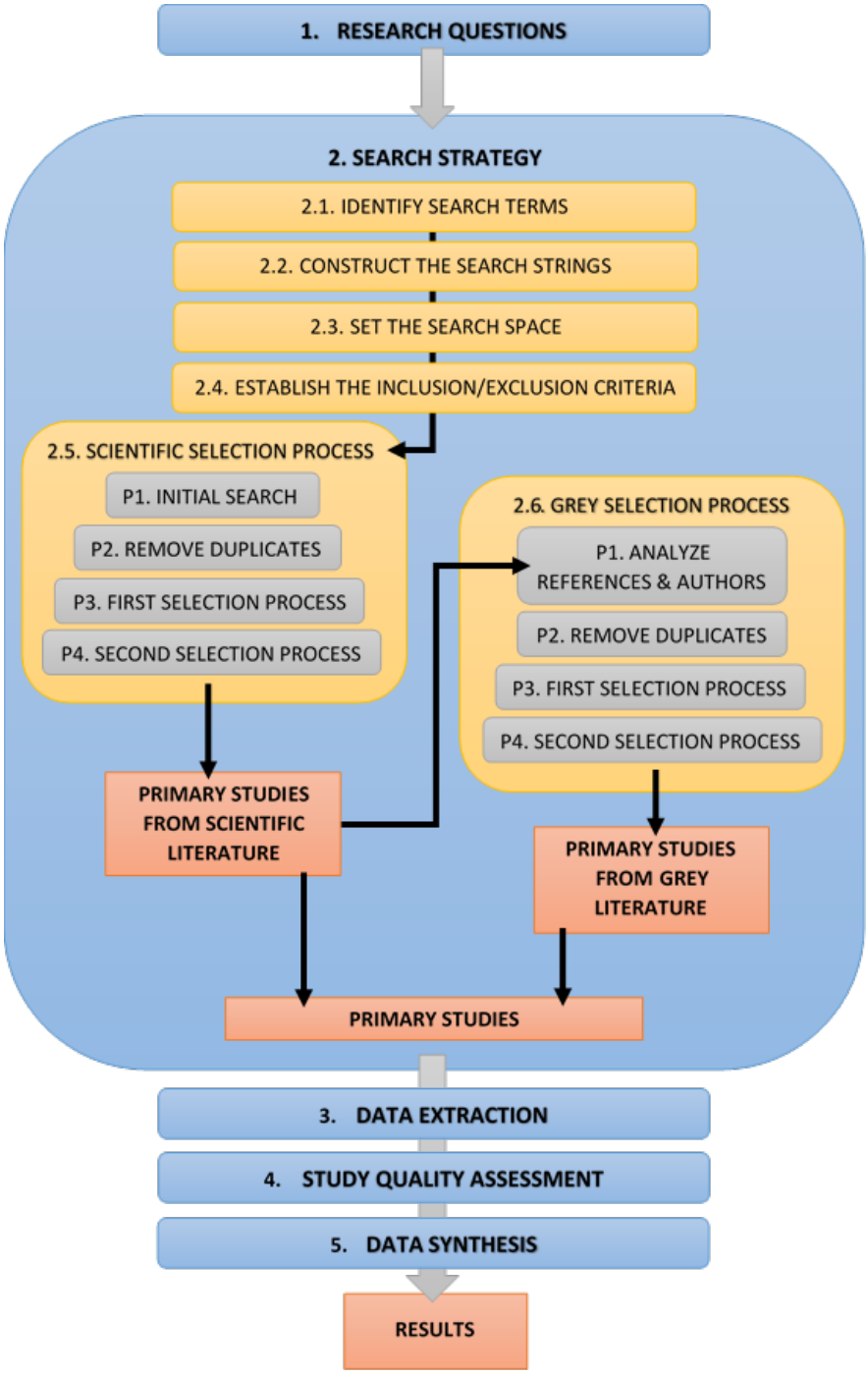

Figure 1. Procedure for conducting the MLR.

4.2. Search Strategy

The aim of this step is to define the search and evaluation strategy for identifying the primary studies. The search and evaluation strategy allows for performing of an exhaustive search of the published literature to find the primary studies from the scientific and the grey literature that answer the proposed research questions. For that reason, we need to identify the search terms, construct the search string, set the resources involved in the searches, and define the study selection process.

Identify Search Terms

The first activity of the search strategy is to identify the terms for the search. For that, we need to analyze the research questions in order to select the general terms related to the topic under review. This will help to assure that most of the relevant works are included in the study.

In our MLR, the main search terms will be "game", "software process standards" and "education".

Construct the Search Strings

Taking into account the main search terms, the search string will be constructed using the steps proposed by Brereton et al. (Brereton, Kitchenham, Budgen, Turner, & Khalil, 2007), which are the following:

1. Derive major terms from the research questions by identifying the main concepts.
2. Identify alternative spellings and synonyms for major terms.
3. Check the keywords in any relevant papers we already have.
4. Use the boolean OR to add alternatives spellings and synonyms.
5. Use the boolean AND to link the major terms.

In our MLR, we conducted some initial searches to test and fine-tune the search string. Finally, we defined the search string as the following Boolean expression "A1 **AND** (B1 **OR** B2 **OR** B3 **OR** B4 **OR** B5 **OR** B6 **OR** B7) **AND** (C1 **OR** C2 **OR** C3 **OR** C4 **OR** C5)" where search terms are shown in Table 2.

Table 2. Search terms

A1. Game	B1. Software Process Standard B2. ISO 29110 B3. ISO/IEC 29110 B4. ISO 12207 B5. ISO/IEC 12207 B6. ISO 15504 B7. ISO/IEC 15504	C1. Education C2. Educate C3. Teach C4. Train C5. Learn

Set the Search Space

Once the search string has been defined, the resources in where the searches will be performed will be decided. The responsible of conducting the searches must adapt the search string to the search engine of each digital resource involved in the study. Depending on the digital resource, the responsible has to select the direct, the advanced or the command search method in order to ensure that the searches take only into consideration the title, the keywords and the abstract of the publication.

In our MLR, the searches were performed in the following digital resources: IEEE Xplore, SCOPUS, ISI Web of Science, SpringerLink, ACM Digital Library, Wiley Online Library and Science Direct. These digital resources were used to find the scientific literature. The main reason for selecting these digital databases was the possibility of accessing its content thanks to the services offered by our institutions. Moreover, we can highlight that these digital databases are the main ones using in the published secondary studies related to our research scope (MacDonell, Shepperd, Kitchenham, & Mendes, 2010).

On the other hand, we retrieved the studies related to the grey literature from Google Scholar. Although there are other sources that provide the required information such as ResearchGate, ORCID, ResearcherID, personal websites of authors, etc., we decided to focus on Google Scholar because it easily allows finding the required information related to the authors of the scientific primary studies and the papers that cite the scientific primary studies. At the same time it allows defining specific strings for searching the information according to our needs. Thus, it makes it possible to conduct searches in a similar way than in the scientific digital databases. Nevertheless, the defined procedure does not limit the sources for retrieving the grey literature to Google Scholar. Therefore, researches can select those sources that they consider more suitable for conducting their MLR.

Establish the Inclusion/Exclusion Criteria

The inclusion/exclusion criteria establish the requirements that papers retrieved from the searches must accomplish in order to be included in the study. Our MLR identifies the studies that introduce serious, educational or simulation games for software process standards education. The studies must be written in English. In order to analyze the whole state of the art of our topic under review, we have not limited the start of the publication period. Table 3 summarizes the inclusion and exclusion criteria defined in our study.

Table 3. Inclusion/Exclusion criteria

Inclusion criteria	▪ Studies that present a game for software process standards education. ▪ Studies that explain the educational features of a game for software process standards education. ▪ Studies that add value to the scope of games for software process standards education.
Exclusion criteria	▪ Studies whose main objective is not related to games for software process standards education. ▪ Studies that present the outcomes of assessing a game but do not show any information about the educational features of the SG. ▪ Studies that discuss about games for software process standards education, but do not provide any information about a specific SG. ▪ Studies with the abstract and the full text unavailable. ▪ Studies not written in English. ▪ Duplicate studies (same paper retrieved from different resources).

Define the Scientific Selection Process

The scientific selection process allows selecting the primary studies from the scientific literature. This process is composed of four phases that help to find the scientific primary studies through a test-retest approach. The four phases of the scientific selection process are the following:

Phase 1. Initial Search: The researcher responsible of the searches will apply the search string to each digital resource with the goal to find all the studies related to the topic under review. The searches will be restricted to title, abstract and keywords.

Phase 2. Remove Duplicates: Papers found during Phase 1 of the scientific selection process will be checked in order to remove the duplicates. The duplicate papers appear when the same paper is retrieved from more than one digital resource. When the duplicates of a paper were identified, only the paper, whose digital resource provides more data about it such as the abstract or the full-text of the paper, the kind, volume or number of the source where the paper was published, the full reference of the publication, etc., will be taken into account for the next phases of the selection process.

Phase 3. First Selection Process: The resulting papers from Phase 2 will be evaluated for their suitability based on the analysis of their title and abstract. During this phase, the title and abstract of each paper will be reviewed against the inclusion and exclusion criteria. If a publication is clearly out of the inclusion criteria, it will be classified as non-selected paper (NS) and it will not be included in the following phases of the selection process. On the other hand, if a publication accomplishes with the inclusion criteria or the information provided by the title and abstract is not enough to make a decision, the publication will be classified as possible selected paper (PS) and will be included in the next phase of the selection process.

Phase 4. Second Selection Process: Publications classified as possible selected papers (PS) during Phase 3 will be thoroughly reviewed. For that, the full text of the publications will be analyzed. This phase will be done to ensure that the publication definitely contains relevant information for the study under review. During this phase, non-relevant publications will be classified as non-selected paper (NS) and relevant publications will be classified as selected paper (S).

At the end of the scientific selection process, the publications classified as selected paper (S) will form the scientific primary studies that are the primary studies retrieved from the scientific literature.

Regarding the researchers involved in the selection process, the proposed procedure is not limited to one researcher. To perform a MLR by more than one researcher in a collaborative way, the researchers involved in the selection process, need to divide each phase of the process in different tasks and allocate them to each participant. Moreover, they have to decide the tools that will use to collect the data and register every decision made (for instance, MS Excel, Mendeley, etc.) and discuss among them every doubts that can appear during all the steps of the MLR procedure. These activities of organizing the work apply not only to the selection process but to the whole process of performing the MLR and are the same activities that researchers need to do in order to conduct any kind of literature review.

In our MLR, three researchers were involved in the selection process: the first author of this paper conducted selection process and the rest of the authors reviewed the process, verified the outcomes and support the resolution of doubts. On the other hand, we used the spreadsheet tool MS Excel to manage all the phases of the scientific selection process and to store the information collected about the searches.

Define the Grey Selection Process

The grey selection process allows selecting the primary studies from the grey literature. This process is composed of four phases that help to find the grey primary studies through a test-retest approach. The four phases of the grey selection process are the following:

- **Phase 1. Analyze References & Authors**: The scientific primary studies retrieved from the scientific selection process will be analyzed in order to find additional studies. In this phase, first forward and backward snowballing (Wohlin, 2014) will be conducted on the scientific primary studies. Snowballing, in this context, refers to using the reference list of a paper (backward snowballing) or the citations to the paper to identify additional papers (forward snowballing) (Wohlin, 2014). Second, the researcher responsible of the searches will apply the search string to the list of works of each of the authors of the scientific primary studies. The aim of this search is to find all the works of each author related to the topic under review. At the end of this phase, a set of additional studies will be retrieved from the analysis of the grey literature.
- **Phase 2. Remove Duplicates**: The studies retrieved during Phase 1 of the grey selection process will be checked in order to remove the duplicates, the studies that are clearly irrelevant to the topic under review and the papers that belong to the scientific primary studies.
- **Phase 3. First Selection Process**: Once the duplicates have been removed, in this phase, the researcher responsible will carry out the same activities described in Phase 3 of the scientific selection process.
- **Phase 4. Second Selection Process**: The works classified as possible selected papers (PS) during Phase 3 will be thoroughly analyzed by reading the full text. In this phase, the researcher responsible will carry out the same activities described in Phase 4 of the scientific selection process.

At the end of the grey selection process, works classified as selected paper (S) will form the grey primary studies that are the primary studies retrieved from the grey literature.

In our MLR, we used Google Scholar for conducting the forward snowballing and the searches in the list of works of each of the authors of the scientific primary studies. Moreover, as during the scientific selection process, we used MS Excel to manage the grey selection process.

Primary Studies

The union of the scientific and grey primary studies defines the primary studies. These primary studies will be analyzed in order to get the needed information needed to answer the research questions the study addresses.

4.3. Data Extraction

During the execution of the MLR, a substantial amount of data will be collected. There are two main sources of data: data collected during the search process and data collected during the extraction process. The data collected during the search process allow collecting general data to identify the papers retrieved from both, scientific and grey, selection processes, support decision-making in order to select the primary studies of the

review and document the selection process. On the other hand, the data collected during the extraction process are the specific data needed to achieve the objectives of the study and answer the research questions under review.

During the extraction process, the primary studies will be read completely to collect all the needed information and ensure that the data will be accurate. All the collected data will be stored in a spreadsheet. This allows placing all the information of the study in the same location, at the same time that makes the analysis and comparison of the collected data during the synthesis process easier.

4.4. Study Quality Assessment

In this step, each primary study should be assessed for internal validity using a standardized approach for rating the quality of the individual studies. The results of this assessment can be used to reduce the number of primary studies regarding their relevance and the information they provide to the review (Calderón & Ruiz, 2015). This step is useful when the number of retrieved primary studies is very high.

In our MLR, each publication was assessed for quality at the same time as the publication data extraction process was performed. A questionnaire, which had to be filled in for each primary paper, was defined as the quality instrument. The assessment questionnaire consisted of seven quality assessment questions (QA) to identify the quality of the information that the paper provided us, and rate its relevance within the MLR. According with the above information, the seven questions used were the following:

- **QA1.** Does the paper introduce a SG for software process standards education?
- **QA2.** Does the paper allow finding the information related to the main features of the SG?
- **QA3.** Does the paper allow knowing the type of SG?
- **QA4.** Does the paper allow knowing the application domain of the SG?
- **QA5.** Does the paper allow finding the information related to the learning objectives of the SG?
- **QA6.** Does the paper allow answering RQ2?
- **QA7.** Does the paper allow answering RQ3?

Each question was answered YES (Y) or NO (N). The questions were scored as follows:

- QA1: Y, the paper presents a SG for software process standards education. N, the publication does not show a SG for software process standards education.
- QA2: Y, the publication provides the information about the name, description and main features of a SG for software process standards education or a SG that adds value to the scope of games for software process standards education. N, the publication does not show the required information.
- QA3: Y, the paper provides the information needed to know the type of SG. N, the publication does not present any information about the type of SG and cannot be readily inferred.
- QA4: Y, the paper shows the application domain of the SG. N, the application domain of the SG is not defined and cannot be readily inferred.
- QA5: Y, the paper provides the information needed to know the learning objectives of the SG and the software processes or standards that the SG aims to train. N, the paper does not provide any information about the learning objectives or the software processes the SG provides training for, and cannot be readily inferred.
- QA6: Y, the paper provides the information about the SG evaluation (i.e. the paper answers RQ2). N, the information about the SG evaluation is not defined and cannot be readily inferred.
- QA7: Y, the paper provides the information about the perceived benefits of the SGs for software process standards education (i.e. the paper answers RQ3). N, the information about the perceived benefits is not defined and cannot be readily inferred.

The scoring procedure will be Y = 1 and N = 0. Hence the total number of Ys will define the quality assessment score of each publication.

4.5. Data Synthesis

Once the primary studies have been thoroughly analyzed and all data have been collected, the last step of the review, before reporting the outcomes, is the synthesis process. In the synthesis process, data collected during the extraction process will be analyzed to compare the information, answer the research questions addressed by the study, and obtain conclusions about the topic under review.

In our MLR, the data of the primary studies was classified according to the research questions addressed, as Table 4 shows.

Table 4. Extracted data classification.

Data	Research question addressed
The name of the SG	RQ1
A brief description of the SG	
The type of game	
The application domain of the SG	
The learning objectives of the SG	
If the SG has been assessed	RQ2
The technique used to assess the SG	
The assessment's objectives	
The size of the population that took part in the evaluation	
The outcomes of the SG's assessment	RQ3
The perceived benefits of the SG for the software process education	

5. Results

The MLR process was conducted during the spring of 2017. The process lasted three months, in which we developed the MLR protocol, identified and selected the primary studies from the scientific and grey literature, performed the data extraction and synthesis processes and reported the results of our study. Three researchers performed the MLR process: the first author of this paper conducted the MLR process and the rest of the authors reviewed the process and verified the work as conducted by the leading author.

In the following subsections, we present the results of the MLR. The results of the search process are analyzed taking into account the different primary studies retrieved from both, grey and scientific literature.

5.1. Scientific Selection Process Results

During the scientific selection process, the papers related to SGs for software process standards education were retrieved in May 2017. As we have previously mentioned, the scientific selection process performed during our study consists of four phases. In the first phase, we performed the initial search of studies applying the search string to each digital electronic resource. As a result, we found 20 papers. In the second phase, we analyzed these papers to remove duplicates. Consequently, we excluded 11 duplicated papers. In the third phase, we checked the title and abstract of each paper according to the inclusion and exclusion criteria. This review reduced the list of papers to 5 possible selected papers to be considered as scientific primary studies. During the last phase of the scientific selection process, the full text of the papers was analyzed for ensuring that suitability of the papers to our review. As a result, we excluded 2 additional papers. At the end of the scientific selection process, the list of papers retrieved from the scientific literature was formed by 3 papers. These 3 papers defined the scientific primary studies of our MLR.

Table 5 shows the evolution of the list of papers during the scientific selection process regarding each digital resource. This figure represents: first, the papers that were found in each digital resource during the first phase of the scientific selection process; second, the number of different papers that were collected from each digital resource after removing the duplicates; third, the number of papers that passed the first selection process; finally, it represents the number of papers that were included as scientific primary studies in this MLR.

Table 5. Evolution of the studies retrieved in each digital resource.

Digital resource	Studies retrieved	Distinct studies retrieved	Studies that passed the first selection process	Primary studies
Science Direct	1	0	0	0
Web of Science	5	0	0	0
SpringerLink	5	5	2	1
ACM Digital Library	1	0	0	0
SCOPUS	6	3	3	2
IEEE Xplore	1	1	0	0
Wiley Online Library	1	0	0	0

5.2. Grey Selection Process Results

Taking into account the three scientific primary studies retrieved during the scientific selection process, we performed the grey selection process. In the first phase of the process, we collected the studies from the grey literature using three different techniques: forward snowballing, backward snowballing and searching in the list of publications of each author of the scientific primary studies. At the end of the first phase of the grey selection process, we retrieved 120 studies using the forward snowballing technique, 7 studies using the backward snowballing technique, and 43 studies searching in the author's list of publications. Hence, 170 studies were collected from the grey literature. Figure 2 shows the percentage of studies collected using the backward snowballing technique (Reference_BS), the percentage of studies collected using the forward snowballing technique (Reference_FS), and the percentage of studies collected using the author's list of publications (Author) regarding the total number of studies collected from the grey literature.

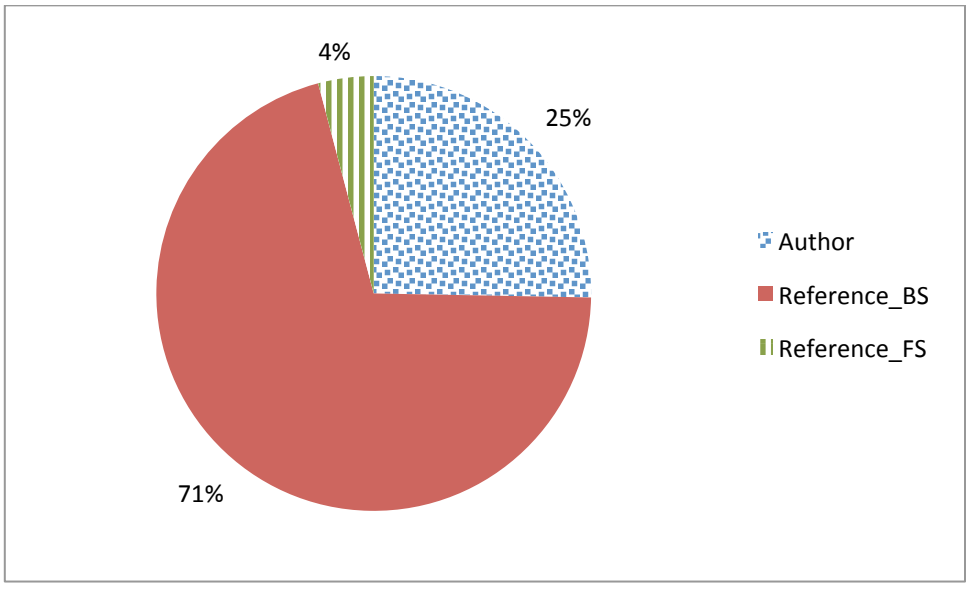

Figure 2. Percentage of studies retrieved from the grey literature.

The second phase of the grey selection process started with 170 studies. After removing the duplicates, we reduced the list of grey literature studies to 22 studies. Then we carried out the third phase of the grey selection process in which we analyzed the title and abstract of each study regarding the inclusion and exclusion criteria. After the first selection process, we excluded 11 studies from the list of grey literature studies. Then we carried out the last phase of the grey selection process to analyze the full text of the studies. As a result, we excluded 7 additional studies from the grey literature list. At the end of the grey selection process, the list of studies retrieved from the grey literature was formed by 4 studies. These four studies defined the grey primary studies of our MLR.

Figure 3, shows the evolution of the list of papers during the grey selection process regarding each technique used to find the studies from the grey literature. This figure represents first: the papers that were retrieved applying each technique during the first phase of the grey selection process; second, the number of different papers that were collected from each technique after removing the duplicates; third, the number of papers from the grey literature that passed the first selection process; finally, the number of papers that were included as grey primary studies in this MLR.

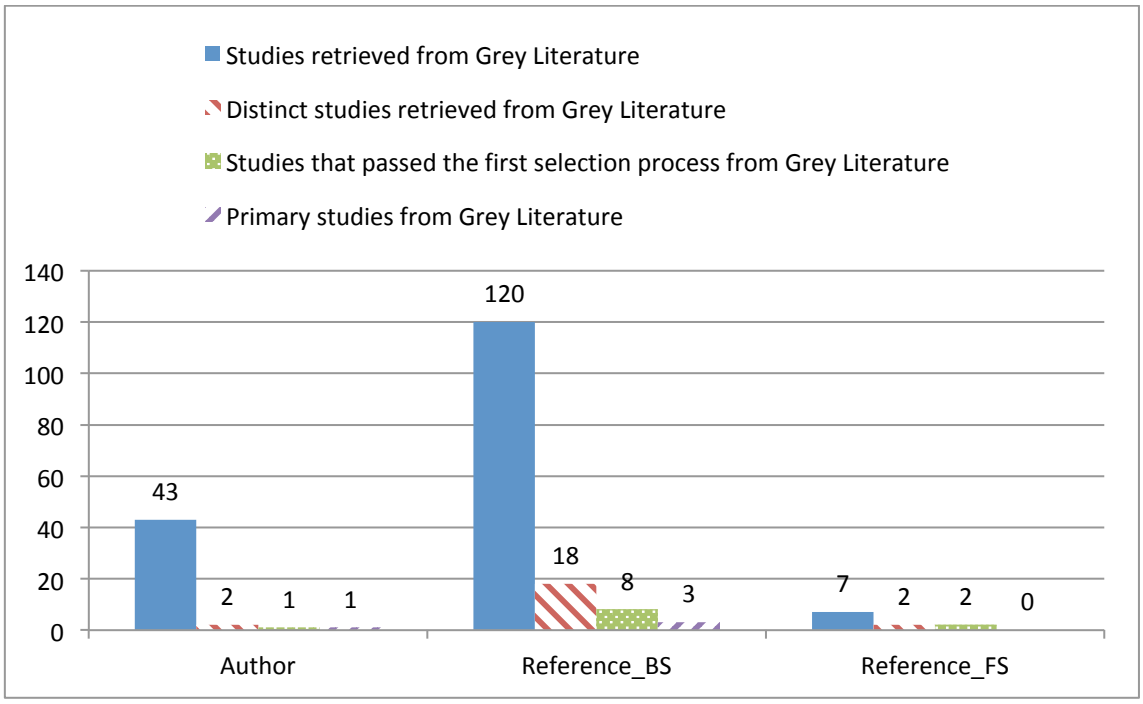

Figure 3. Evolution of the studies retrieved from the grey literature.

5.3. Primary Studies

The results of both, the scientific and grey selection process allowed us to define the primary studies of our MLR. Our primary studies were formed by 7 studies taking into account the scientific primary studies and the grey primary studies. Table 6 shows the information related to the title, authors and publication date of the primary studies of our MLR, as well as an identifier (PS_ID) that allows referencing each primary study in a unique way within the MLR. The first three primary studies (PS_01, PS_02 and PS_03) were retrieved from the scientific literature and the rest (PS_04, PS_05, PS_06 and PS_07) were retrieved from the grey literature. These studies were the basis for our extraction and synthesis processes.

Table 6. Primary Studies

PS_ID	Title	Author/s	Publication Date
PS_01	Coverage of ISO/IEC 12207 Software Lifecycle Process by a Simulation-Based Serious Game	Calderón, A.; Ruiz, M.	2016
PS_02	Teaching ISO/IEC 12207 software lifecycle processes: A serious game approach	Aydan, U.; Yilmaz, M.; Clarke, P.M.; O'Connor, R.V.	2017
PS_03	A learning tool for the ISO/IEC 29110 standard: Understanding the project management of basic profile	Sánchez-Gordón, M.-L.; O'Connor, R.V.; Colomo-Palacios, R.; Sanchez-Gordon, S.	2016
PS_04	An experimental card game for teaching software engineering processes	Baker, A.; Navarro, E.O.; van Der Hoek, A.	2005
PS_05	Experimental evaluation of a serious game for teaching software process modeling.	Chaves, R.O.; von Wangenheim, C.G.; Furtado, J.C.C.; Oliveira, S.R.B.; Santos, A.; Favero, E.L.	2015
PS_06	SimSE: an interactive simulation game for software engineering education	Navarro, E.O.; van der Hoek, A.	2004

Calderón, A., Ruiz, M., & O'Connor, R. V. (2018). A multivocal literature review on serious games for software process standards education. *Computer Standards & Interfaces*, *57*, 36-48.

PS_07	Coverage of ISO/IEC 29110 Project Management Process of Basic Profile by a Serious Game	Calderón, A.; Ruiz, M.; O'Connor, R.V.	2017

Figure 4 shows the percentage of the primary studies retrieved from the scientific and grey literature, that are included in our MLR.

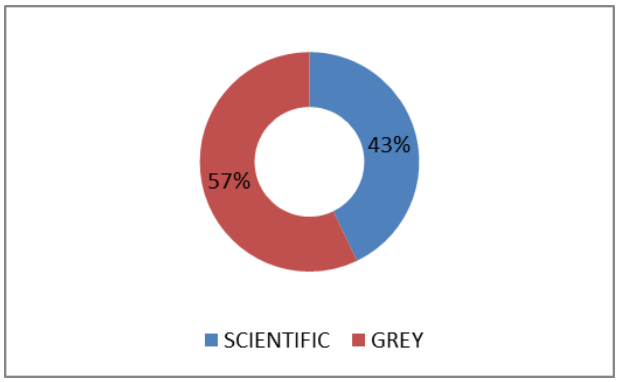

Figure 4. Percentage of primary studies provided by each literature.

Figure 5 presents the number of primary studies regarding their year of publication. As we can observe, all of the primary studies retrieved were published between 2004 and 2017, with a high number (5 out of 7) of studies published since 2015. This allows us to see that the topic under review, SGs for software process standards education, is an emerging research scope in which the interest is growing.

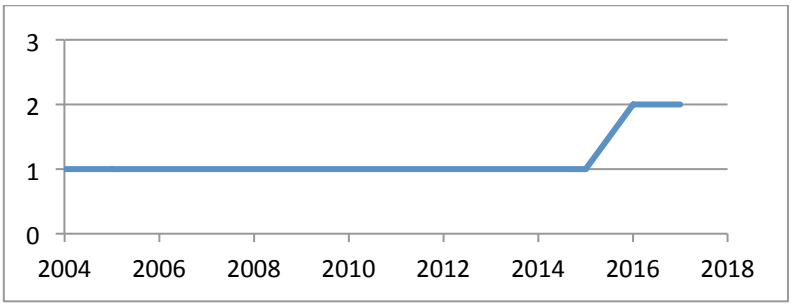

Figure 5. Number of studies published regarding the year of publication.

Finally, our 7 primary studies have been published in conferences or journals. Table 7 shows the place of publication of each primary study and presents the information related to its identifier. More precisely, the title of the place where each primary study has been published or is going to be published, the kind of place of publication that it means if the primary study belongs to a conference or journal, and the quality of the place of publication. Regarding the quality of the journals, we have classified the studies regarding the quartile in which the journal is placed according to the Journal Citation Report (JCR) (Clarivate, 2017). On the other hand, conference studies have been classified regarding the CORE ranking (Computing Research and Education Association of Australasia, 2016).

Table 7. Primary studies' place of publication.

PS_ID	Place of Publication Title	Kind	Quality
PS_01	16th International Conference on Software Process Improvement and Capability Determination (SPICE 2016)	Conference	CORE A
PS_02	Computer Standards and Interfaces	Journal	COMPUTER SCIENCES, SOFTWARE ENGINEERING (Q2)
PS_03	16th International Conference on Software Process Improvement and Capability Determination (SPICE 2016)	Conference	CORE A
PS_04	Journal of Systems and Software	Journal	COMPUTER SCIENCE, SOFTWARE

			ENGINEERING (Q1)
PS_05	IEEE Transactions on Education	Journal	EDUCATION, SCIENTIF DISCIPLINES (Q2); ENGINEERING, ELECTRICAL & ELECTRONIC (Q2)
PS_06	International Conference on Computers and Advanced Technology in Education (CATE 2004)	Conference	CORE C
PS_07	24th European System, Software & Service Process Improvement & Innovation Conference (EuroSPI 2017)	Conference	CORE B

Figure 6 shows the number of primary studies regarding the quality of their place of publication. As we can observe, all the studies belong to conferences or journals indexed in the Journal Citation Report or the CORE ranking.

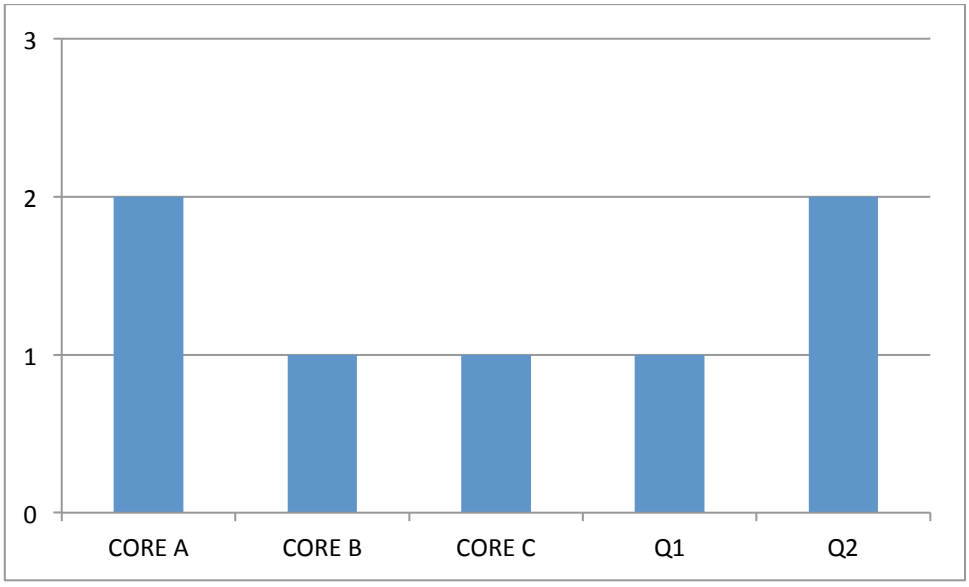

Figure 6. Number of studies regarding the quality of their place of publication.

5.4. Study Quality Assessment

We evaluated the primary studies for quality using a questionnaire (see Section 4.4). The seven quality questions were rated for each primary study in accordance with the criteria established in the quality questionnaire. Figure 7 shows the coverage of every QA in the primary studies. It shows that QA2, QA3, QA4 and QA5 were covered 100% by Yes answer. In addition, 6 out of 7 questions were covered in a rate higher than 80% by Yes answer. QA1 was the question with the lowest rate (57%). This is because, although every primary study introduces a SG related with software process education, only 4 out of 7 primary studies introduce a SG that really focused on software process standards education. Nevertheless, all the quality assessment questions were covered in a rate higher than 50% by Yes answer.

The score of the quality assessment takes a value in a range from 0 (it does not provide enough information) to 7 (it provides all the information required). In general, all the primary studies achieved a quality score equal or greater than 6, with the exception of one that got a value of 4. Then, 5 out of 6 primary studies were considered to provide the information needed to answer our research questions.

Due to the fact that the number of retrieved primary studies is not very high, in this MLR we are reporting this step to exemplify its execution within the defined procedure. As it can be seen, executing this step did not help to reduce the number of primary studies that participate in the data synthesis.

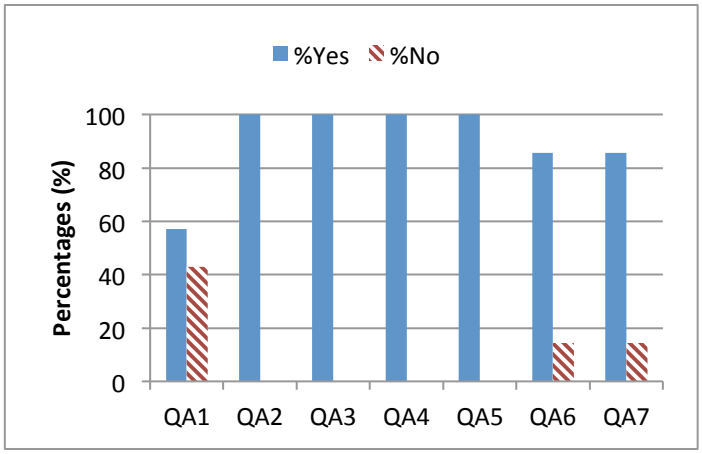

Figure 7. Quality assessment results per question and type of assessment response.

6. Discussion

In this section, we discuss our findings and the answers to the research questions that addressed this MLR addresses.

6.1. RQ1. What SGs for software process standards education are available?

The goal of this question was to identify the existing SGs within the scope of software process standards education and classify them according to their features. To answer this question, for each SG, we collected: the information related to its name, its description, its type, the application domain and its learning objectives.

We found a total of 6 different SGs in this review, which are the following:

- **ProDec** is a simulation-based SG to teach and motivate software developers in learning and practicing the principles of software project management, as well as supporting the comprehension and knowledge acquisition of project management lifecycle processes and activities, in a risk-free environment (Calderón & Ruiz, 2016) (Calderón, Ruiz, & O'Connor, 2017).
- **Floors** is a SG that proposes an interactive learning experience to introduce ISO/IEC 12207:1995 by creating different floors of a virtual environment where various processes of the standard are discussed and implemented (Aydan, Yilmaz, Clarke, & O'Connor, 2017).
- **Go for it!** is a non-technological educational game for contributing to teaching the ISO/IEC 29110 standard elements where players are encouraged to understand the project management process of the Basic profile (Sánchez-Gordón, O'Connor, Colomo-Palacios, & Sanchez-Gordon, 2016).
- **Problems and Programmers (PnP)** is an educational card game that simulates the software engineering process and is designed to teach those process issues that cannot sufficiently addressed by lectures and educational projects (Baker, Navarro, & Van Der Hoek, 2005).
- **DesignMPS** is a computer game designed to support the teaching of software process modelling by reinforcing relevant concepts and providing software process modelling exercises (Oliveira Chaves, von Wangenheim, Costa Furtado, Ronaldo Bezerra Oliveira, Santos, & Favero, 2015).
- **SimSE** is an educational, interactive, fully graphical computer game that simulates software engineering processes, and is designed specifically to train students in situations that require an understanding and handling of software process issues (Navarro & van der Hoek, 2004).

Table 8 summarizes the information collected to answer this question.

Table 8. SGs.

Serious	Type	Application	Learning objectives	Primary studies

game		domain		
ProDec	Computer game	University / Industry	ISO/IEC 12207 and ISO/IEC 29110 standards	(Calderón, Ruiz, & O'Connor, 2017) (Calderón & Ruiz, 2016)
Floors	Computer game	University / Industry	ISO/IEC 12207 standard	(Aydan, Yilmaz, Clarke, & O'Connor, 2017)
Go for it!	Card game	Industry	ISO/IEC 29110 standard	(Sánchez-Gordón, O'Connor, Colomo-Palacios, & Sanchez-Gordon, 2016)
PnP	Card game	University	Software engineering process	(Baker, Navarro, & Van Der Hoek, 2005)
DesignMPS	Computer game	University / Industry	Software process modelling	(Oliveira Chaves, von Wangenheim, Costa Furtado, Ronaldo Bezerra Oliveira, Santos, & Favero, 2015)
SimSE	Computer game	University	Software engineering process	(Navarro & van der Hoek, 2004)

Regarding the type of SG, as shown in Table 8, we classified the SGs found in the primary studies in two categories: computer and card games. The category *Computer game* refers to SGs developed as computer software and the category *Card game* refers to SGs developed as a card game. Figure 8 presents the percentages of primary studies regarding the category of the SG that they introduced. As it can be seen, 71% of the primary studies introduce a SG developed as computer software, while the rest of primary studies (29%) introduce a card game.

If we further analyze each category, the majority of the computer games have been developed as simulation games in which players take the role of a project manager inside a business environment.

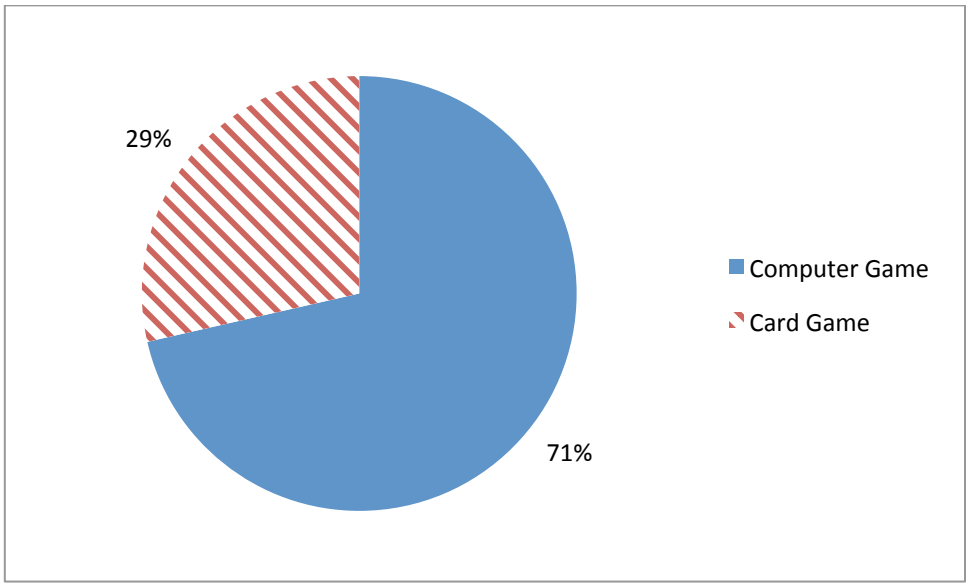

Figure 8. Type of game.

Regarding the application domain in which the SGs are used, we classified them into two categories: industry and university. The category *University* refers to SGs developed for teaching future practitioners *at* University, while the category *Industry* refers to SGs developed for teaching practitioners in an industrial environment. Only one of the SGs analyzed, is specific for teaching in an industrial environment, two of them are used on university environment and the rest of the SGs (3 out of 6) are designed to teach learners in both, university and industrial domains.

Regarding the learning objectives of the SGs retrieved from the primary studies (see Table 8), we can conclude the following:

- Two primary studies introduce a SG for teaching the software lifecycle process of the ISO/IEC 12207 standard (Calderón & Ruiz, 2016) (Aydan, Yilmaz, Clarke, & O'Connor, 2017).

- Two primary studies introduce a SG for teaching the project management process of the Basic profile of the ISO/IEC 29110 standard (Calderón, Ruiz, & O'Connor, 2017) (Sánchez-Gordón, O'Connor, Colomo-Palacios, & Sanchez-Gordon, 2016).
- Two primary studies introduce a SG designed to teach general and specific lessons about the software engineering processes (Baker, Navarro, & Van Der Hoek, 2005) (Navarro & van der Hoek, 2004).
- A primary study presents a SG to support software process modeling education (Oliveira Chaves, von Wangenheim, Costa Furtado, Ronaldo Bezerra Oliveira, Santos, & Favero, 2015).
- A primary study introduced a SG that helps learners to take contact with the ISO/IEC 33014 and the ISO/IEC 15504 (Calderón & Ruiz, 2016).

As a conclusion, only three of the SGs found in the literature (scientific and grey) are directly related to the education of specific software process standards while the other three are related to software process education, in general terms.

6.2. RQ2. How have these SGs been evaluated?

Through this research question, our goal was to get information about the evaluation of the SGs retrieved from the primary studies in order to get an overview of the methods and techniques used to evaluate them. Additionally, the procedure to perform the evaluations, the population involved in the assessments and the assessment objectives have also been identified.

In Table 9, we show the objectives of the evaluations presented in the primary studies. In general terms, these evaluation were focused on: (i) assessing the feasibility of the SG presented as a learning resource to teach and support in the scope of software process education and (ii) getting feedback about the user experience in order to observe the strengths and weaknesses of the SG. In addition, two studies evaluated the learning effectiveness of the SG as compared with traditional learning approaches.

Table 9. Evaluation's objectives.

Primary studies	Serious game	Main evaluation' objectives
(Calderón & Ruiz, 2016)	ProDec	To assess the feasibility of using ProDec to teach the software lifecycle process of ISO/IEC 12207 standard.
(Aydan, Yilmaz, Clarke, & O'Connor, 2017)	Floors	To assess the benefits on the user experience when using virtual learning over the paper based learning one.
(Sánchez-Gordón, O'Connor, Colomo-Palacios, & Sanchez-Gordon, 2016)	Go for it!	To test the overall applicability of the game as a learning tool.
(Baker, Navarro, & Van Der Hoek, 2005)	PnP	To assess the game's feasibility and worth as a complementary teaching tool.
(Oliveira Chaves, von Wangenheim, Costa Furtado, Ronaldo Bezerra Oliveira, Santos, & Favero, 2015)	DesignMPS	To examine the effect of DesignMPS on the cognitive levels of remembering, understanding and applying of Bloom's taxonomy (Krathwohl, 2002), and to compare the learning effectiveness of DesignMPS and a project-based learning activity.
(Navarro & van der Hoek, 2004)	SimSE	The paper does not provide any evaluation of the SG.
(Calderón, Ruiz, & O'Connor, 2017)	ProDec	To get initial feedback of the player's experience and perceived learning with the goal to assess ProDec as a learning resource to support the training and understanding of ISO/IEC 29110 standard.

Only one primary study (Navarro & van der Hoek, 2004) did not provide any information related to the evaluation of the SG introduced. Nevertheless, we performed a manual search on Google in order to find some data to complement the information related to the SG SimSE, already located. In these searches, we found a study that presents and summarizes several evaluations of SimSE (Navarro & van der Hoek, 2007). In this study, the authors conducted two main evaluations: an initial evaluation to provide an overview of the effectiveness of SimSE and understand its strengths and weaknesses from multiple angles, and an experimental study to observe students' experience and learning effectiveness from using SimSE and to

compare it with a traditional learning approach. In the rest of this section, we will use the information retrieved from this complementary study in order to provide information related to the evaluation of SimSE.

Table 10 shows the techniques that the authors of our primary studies used to evaluate their SGs.

Table 10. Evaluation techniques

Technique	Definition	SGs
Questionnaires	The method uses forms of questions to assess the SG. The answers are provided via written forms.	ProDec, Floors, Go for it!, PnP, DesignMPS, SimSE
Interviews	The method uses forms of questions to assess the SG in oral way. The answers are provided orally.	Floors, SimSE
Delphi	The Delphi method is used to perform the evaluation of the SG (Skulmoski, Hartman, & Krahn, 2007).	ProDec
Observations	The method is based on the observation of the session progress to assess the SG	SimSE, ProDec

In Figure 9, we show the number of SGs that have been evaluated through each of the identified techniques. We can observe that the technique based on questionnaires is the most used for assessing the SGs in the scope under review.

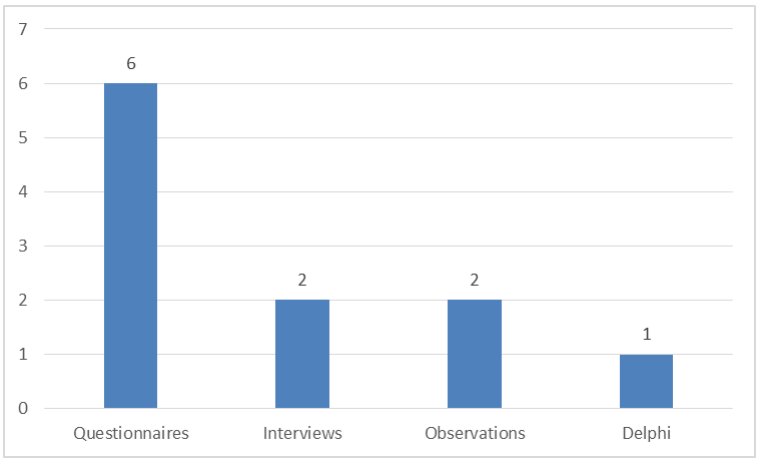

Figure 9. Number of SGs regarding the technique used to evaluate them.

Regarding the moment when the authors apply the evaluation methods during the process of the SG evaluation, we identified two main types of procedure: post or pre/post assessment. In a post assessment procedure, the authors carry out a session with the SG and, after playing the game, the evaluation methods are provided to the players. On the other hand, in a pre/post assessment, two evaluations are carried out during the session with the SG, one before using the SG and another after using the SG. In this context, 5 out of 6 of the SGs retrieved from the primary studies have been evaluated using a post assessment procedure, while 2 out of 6 were involved on a pre/post assessment procedure.

Moreover, we analyzed the experimental or non-experimental character of the procedure used to evaluate the SGs. We refer to an experimental procedure as the procedure where authors have used a control group to compare the results of the experimental group. Considering this classification, three of the SGs retrieved from our primary studies have been evaluated using an experimental procedure. These SGs are Floors (Aydan, Yilmaz, Clarke, & O'Connor, 2017), DesingMPS (Oliveira Chaves, von Wangenheim, Costa Furtado, Ronaldo Bezerra Oliveira, Santos, & Favero, 2015) and SimSE (Navarro & van der Hoek, 2007).

The last data of interest related to the evaluation of the SGs under review is the size of the population involved in the evaluation session. The size of the population varies between 10 and 41 people. All the evaluations were performed by undergraduate students, with the exception of the study (Calderón & Ruiz, 2016) in which the participants involved in the evaluation session were university professors. The analysis

data related to the size of the population reveals that the existing evaluations of the scope under review did not involve a large number of people.

Table 11 collects the information related to the population size, as well as the main information collected in order to answer RQ2. Each SG retrieved from the primary studies presents information related to an evaluation, with the exception of SimSE and ProDec that count with information regarding two different evaluations.

Table 11. Summary of the information about the evaluation features.

Serious game	Techniques	Assessment Procedure	Experimental	Population size
ProDec	Delphi	Post	Non-experimental	10
Floors	Questionnaire, Interviews	Post	Experimental	40
Go for it!	Questionnaire	Post	Non-experimental	33
PnP	Questionnaire	Post	Non-experimental	28
DesignMPS	Questionnaire	Pre/Post	Experimental	41
SimSE	Questionnaire, Interviews, Observations	Post	Non-experimental	29
SimSE	Questionnaire, Interviews, Observations	Pre/Post	Experimental	19
ProDec	Questionnaire, Observations	Post	Non-experimental	21

6.3. RQ3. What are the perceived benefits of using these SGs in software process standards education?

The main objective of this research question was to observe the perceived benefits of the use of SGs for teaching in the scope of software process standards. For this reason, we analyzed the primary studies in order to get information about the outcomes of the evaluations of the SGs, as well as the conclusions of the authors and their main insights.

The use of SGs to teach the ISO/IEC 12207 standard has been evaluated by (Calderón & Ruiz, 2016) and (Aydan, Yilmaz, Clarke, & O'Connor, 2017). In the first, a group of professors evaluate the idea of using the SG ProDec to teach the software lifecycle processes of the ISO/IEC 12207 standard. As a result, the experts involved in the assessment concluded that: (i) the use of ProDec can help learners to acquire practical experience in the software process field and (ii) ProDec can be only considered as a support learning resource because it does not teach the main principles of software processes, but does not put them into practice. On the other hand, Aydan et al. (Aydan, Yilmaz, Clarke, & O'Connor, 2017)investigate the usefulness of the SG Floors to teach the basics of ISO/IEC 12207 standard. For this end, they conducted two different training sessions, one using Floors and another with a traditional paper-based learning approach. The outcomes of the evaluation showed that the user experience of the participants involved in the game learning session was more positive and beneficial than the one of those involved in the traditional learning session.

Regarding the ISO/IEC 29110 standard, the primary study (Sánchez-Gordón, O'Connor, Colomo-Palacios, & Sanchez-Gordon, 2016) presents the results of a pilot study conducted to assess the overall applicability of the SG Go for it! as a learning tool and (Calderón, Ruiz, & O'Connor, 2017)shows a preliminary evaluation to assess the educational potential of the SG ProDec as a learning resource to support the understanding of the standard. The results of the evaluations showed that the participants involved in the evaluation considered that: (i) Go for it! has potential as a learning game to support gaining awareness and understanding of the ISO/IEC 29110 standard, (ii) Go for it! seems to be fun, immersive and certainly involves the participants, (iii)ProDec helps to understand and put into practice the concepts and practices related to software project management process of the standard, and (iv) ProDec can be a potential learning resource not only in terms of learnings goals but also in terms of usability, relevance, confidence, challenge, satisfaction, immersion, social interaction and fun that the game is able to provide.

An initial evaluation of the SG PnP to get feedback about the students' thoughts, feelings and opinions about the educational effectiveness of the game is presented in (Baker, Navarro, & Van Der Hoek, 2005). As a result, the students qualified their experience as a useful and enjoyable one. They also thought that the

integration of the game in the course's curriculum could have positive benefits. The same authors introduce the SG SimSE, in the primary study (Navarro & van der Hoek, 2004), whose evaluation is shown in (Navarro & van der Hoek, 2007). In this study, the authors conducted several evaluations to assess the educational effectiveness of SimSE. In their results, they concluded that: (i) SimSE is most educationally effective when used as a complementary component to other teaching methods, and (ii) students who play SimSE seem to successfully learn the concepts it is designed to teach, find the experience relatively enjoyable and think the game is repetitive when played for extended periods.

Finally, the primary study (Oliveira Chaves, von Wangenheim, Costa Furtado, Ronaldo Bezerra Oliveira, Santos, & Favero, 2015) presents a formal experiment to: a) examine the effects of the SG DesignMPS on the cognitive levels of remembering, understanding and applying, b) to assess the learning effectiveness of DesignMPS, and c) to compare it with a project-based learning activity. The outcomes showed that the use of DesignMPS is more effective than the project-based learning at the applying level and leads to improvement at the understanding level. Nevertheless, the authors admitted that the significance of the results to be weak taken into account the threats of validity of their work.

On the other hand, we also analyzed the conclusions, of the primary studies related to the use of SGs for software process standards education. In Table 12, we show the main findings.

Table 12. Conclusions of the primary studies.

Primary studies	Author's conclusions
(Calderón & Ruiz, 2016)	▪ The use of games and simulation-based experiences help trainers to teach software process in a practical way within a risk-free environment. ▪ The use of ProDec is beneficial for learners and helps them to consolidate their knowledge in software process and software project management.
(Aydan, Yilmaz, Clarke, & O'Connor, 2017)	▪ Games can offer genuine and lasting improvements over classical teaching techniques. ▪ Using games as teaching tools is ultimately serves for improving student's understanding of the ISO/IEC 12207 processes. ▪ More research is needed to better understand Floors' potentials.
(Sánchez-Gordón, O'Connor, Colomo-Palacios, & Sanchez-Gordon, 2016)	▪ Go for it! could be used as a strategy to promote the ISO/IEC 29110 standard and show practitioners that learning a standard does not have to be boring and painful. ▪ Games have potential to support education.
(Baker, Navarro, & Van Der Hoek, 2005)	▪ When used in conjunction with lectures and projects, PnP will allow students to gain a thorough understanding of real-world lessons that might otherwise have been poorly understood or overlooked altogether.
(Oliveira Chaves, von Wangenheim, Costa Furtado, Ronaldo Bezerra Oliveira, Santos, & Favero, 2015)	▪ The little evidence available suggests that games for teaching software processes are usually more effective when used to support other educational method. ▪ Simpler models for game assessment should be investigated to obtain more reliable evidence of the impact of SGs on software engineering education.
(Navarro & van der Hoek, 2004)	▪ SimSE is most effective when used with other teaching methods. ▪ SimSE has the potential to be and educationally effective tool in teaching students process concepts. ▪ Learning through visual clues has proven to be far more effective than simply studying textual output.
(Calderón, Ruiz, & O'Connor, 2017)	• ProDec has the potential to be a learning resource for supporting software process education and help practitioners of VSEs understanding and practicing the Project Management process of the Basic profile of the ISO/IEC 29110 standard.

Taking into account the outcomes of the evaluations and the author's conclusions, the following insights can be summarized:

- SGs have the potential as learning resources for teaching software processes and helping to understand the concepts and practices of the standards.
- SGs seem to be most effective as supporting tools that complement other teaching methods.

- Learners' experience is more positive in game based learning approaches than in traditional learning approaches.
- More research in game' assessment and methods for game assessment are required.

7. Threats to validity

As with all literature reviews, this study has several factors that may affect its validity. Regarding the search process, our review is limited by the search terms used, the search strings defined, the journals included and the time period of studies published, because they are the factors that limit the work performed during the search process. To reduce the influence of these factors, we developed the review protocol following the guidelines proposed by (Kitchenham & Charters, 2007) and we decided to perform a MLR instead of a SLR to include the grey literature within our review. Hence, we also analyzed other SLRs and MLRs in the literature in order to complement our review protocol.

The inclusion of the grey literature in our review helped to overcome the scarse works retrieved by the automatic searches in the digital databases of scientific literature. In order to analyze the grey literature in a systematic way, we defined a new selection process that allowed us to obtain not only the primary studies from the scientific literature but also from the grey literature.

The process for selecting the final primary studies described could not be feasible when the number of papers retrieved from the scientific literature is too high. In this situation, researchers must include constraints to meaningfully restrict the number of papers retrieved, such as limiting the year of publication within the inclusion/exclusion criteria or reduction techniques within the selection process in order to reduce the amount of retrieved studies. For instance, in (Calderón & Ruiz, 2015), the authors reduced from 102 to 13 the number of studies that were analyzed to retrieve the grey literature by focusing only in the best primary studies based on their score in the quality assessment. Another example is shown in (Garousi, Felderer, & Hacaloğlu, 2017) where the searches for the grey literature in Google's regular search engine were restricted by using the relevance ranking of the search engine (Google's PageRank algorithm). The inclusion of constraints and reduction techniques within the MLR selection process helps to decrease the number of studies involved in the process and therefore, the time required for conducting the review at the expense of accuracy in the results. In our particular MLR, we retrieved a total of 20 primary studies from the scientific literature and, consequently, we did not have to deal with this problem.

Regarding the human resources involved in the MLR, a single researcher was the responsible for performing the selection process of the primary studies. This could be a threat to the validity of the review because of possible human errors or subjectivity opinions. In order to overcome this limitation, we decided to perform each selection process (grey and literature) based on a test-retest approach through which studies had to pass two evaluation stages before being selected as primary studies. Moreover, this threat to the validity was present during all the procedure of the MLR, to avoid the possible human errors during it, two expert researchers reviewed the work performed during all the complete procedure of the MLR.

Finally, regarding the information extracted from the primary studies, it can occur that a primary study does not provide enough information to answer all the research questions defined by the review. In this case, due to the small number of primary studies found to perform manual searches via Google to find studies that could complement the information of each SG.

8. Conclusions and future works

The contributions of this paper are the following:

- A new procedure for performing MLR. For that, we have defined a new procedure for executing MLRs that establishes all the steps involved in the search and evaluation strategies to get the primary studies from both the scientific and grey literature sources. The procedure is based on the best practices and guidelines for conducting SLRs in software engineering proposed by Kitchenham et al. (MacDonell, Shepperd, Kitchenham, & Mendes, 2010) (Kitchenham & Charters, 2007) and the

methods and procedures of SLRs that take into account the grey literature (Garousi, Felderer, & Hacaloğlu, 2017) (Calderón & Ruiz, 2015).

- The state of the art of SG for software project standards education. We have applied the defined procedure to perform a MLR that collects the current serious games within this scope.
- An overview of the features and perceived benefits of the current serious games for software process standards education. By conducting the MLR, we identified 7 primary studies, analyzed them and classified the information retrieved to answer each of the three research questions defined. The current SGs for software process standards education and its main features were identified. In addition, the evaluations of these SGs, their main results obtained, as well as, the perceived benefits related to their use were analyzed and synthesized. This analysis provides an overview of the SGs for software process standards education and establishes the analyzed information as a starting point for further research.

Six different SGs were identified within the topic under review: ProDec, Floors, Go for it!, PnP, DesignMPS and SimSE, but only three of them aim to teach specific software process standards such as ISO/IEC 12207 and ISO/IEC 29110. We found that the majority of these SGs have been developed as computer games and there is a high interest in using them to teach in an industrial environment, since most of the SGs analyzed have been specifically created for this kind of environment. Regarding the assessment procedure of these SGs, all the studies applied questionnaires as the main evaluation technique, and most of them conducted non-experimental procedure using post-session evaluations to assess the educational effectiveness of the SGs as learning tools for software process standards education. Although the main target of these games is training practitioners, all the evaluations have taken place in academic environment involving professors and undergraduate students.

The results of the different SGs' evaluations, the conclusions of the analyzed studies and the perceived benefits of their authors allow us to conclude that SGs have potential as supporting tools for software process standards education. Nevertheless, the use of the SG for software process standards education is an emerging research scope in which several studies have been published in the recent years. Hence, more research and experimental outcomes are needed in order to observe the full potential of SGs as learning resources in the scope of software process standards education.

Acknowledgements

This work was funded by the Spanish National Research Agency (AEI) with ERDF funds under project BadgePeople (TIN2016-76956-C3-3-R) and the Andalusian Plan for Research, Development and Innovation (grant TIC-195).

References

Abt, C. (2002). *Serious Games.* Lanhan, MD: University Press of America.

Auger, P. (1998). *Information Sources in Grey Literature* (Fourth ed.). London, UK: Bowker Saur.

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.

Baker, A., Navarro, E. O., & Van Der Hoek, A. (2005). An experimental card game for teaching software engineering processes. *Journal of Systems and Software, Software Engineering Education and Training , 75* (1), 3-16.

Brereton, P., Kitchenham, B., Budgen, D., Turner, M., & Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software , 80* (4), 571-583.

Calderón, A., & Ruiz, M. (2015). A systematic literature review on serious games evaluation: an application to software project management. *Computers & Education , 87*, 396-422.

Calderón, A., Ruiz, M., & O'Connor, R. V. (2018). A multivocal literature review on serious games for software process standards education. *Computer Standards & Interfaces*, *57*, 36-48.

Calderón, A., & Ruiz, M. (2016). Coverage of ISO/IEC 12207 software lifecycle process by a simulation-based serious game. *Proceedings of the 16th International Conference on Process Improvement and Capability dEtermination (SPICE)*, (pp. 59-70). Dublin, Ireland.

Calderón, A., Ruiz, M., & O'Connor, R. (2017). Coverage of ISO/IEC 29110 Project Management Process of Basic Profile by a serious game. *Proceeding of the 24th European System, Software & Service Process Improvement & Innovation Conference (EuroSPI)* (pp. 111-122). Ostrava, Czech Republic: Springer-Verlag.

Clarivate. (2017). *Clarivate Analytics*. Retrieved September 7, 2017 from Clarivate Analytics: https://clarivate.com/products/journal-citation-reports/

Computing Research and Education Association of Australasia. (2016). *CORE*. Retrieved September 7, 2017 from CORE: http://www.core.edu.au/

Garousi, V., Felderer, M., & Hacaloğlu, T. (2017). Software test maturity assessment and test process improvement: a multivocal literature review. *Information and Software Technology , 85*, 16-42.

Garousi, V., Felderer, M., & Mäntylä, M. (2016). The need for multivocal literature reviews in software engineering: complementing systematic literature reviews with grey literature. *Proceeding of the 20th International Conference on Evaluation and Assessment in Software Engineering (EASE)*. Article n. 26: ACM.

Heredia, A., Colomo-Palacios, R., & Amescua-Seco, A. (2015). A systematic mapping study on software process education. *Proceedings of the 1st International Workshop on Software Process Education, Training and Professionalism (SPETP)*, (pp. 7-17).

Humphrey, W. S. (1995). *A Discipline for Software Engineering*. Boston, MA, USA: Addison-Wesley Longman Publishing Co.

ISO/IEC. (2008). *ISO/IEC 12207:2008 - Systems and software engineering — Software life cycle processes*.

ISO/IEC. (2004). *ISO/IEC 15504-1:2004 Information Technology – Process Assessment – Part 1: Concepts and Vocabulary*.

ISO/IEC. (2016). ISO/IEC TR 29110-1:2016 - Systems and software engineering -- Lifecycle profiles for Very Small Entities (VSEs) -- Part 1: Overview.

Kitchenham, B., & Charters, S. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering*. Keele University and Durham University Joint Report.

Kosa, M., Yilmaz, M., O'Connor, R., & Clarke, P. (2016). Software engineering education and games: a systematic literature review. *Journal of Universal Computer Science , 22* (12), 1558-1574.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: an overview. *Theory into Practice , 41* (4), 212-218.

MacDonell, S., Shepperd, M., Kitchenham, B., & Mendes, E. (2010). How reliable are systematic reviews in empirical software engineering? *IEEE Transactions on Software Engineering , 36* (5), 676-687.

Marques, M. R., Quispe, A., & Ochoa, S. F. (2014). A systematic mapping study on practical approaches to teaching software engineering. *Proceeding of the Frontiers in Education Conference (FIE)* (pp. 1-8). IEEE.

Navarro, E. O., & van der Hoek, A. (2007). Comprehensive evaluation of an educational software engineering simulation environment. *Proceedings of the 20th Conference on Software Engineering Education & Training (CSEET)*, (pp. 195-202).

Navarro, E. O., & van der Hoek, A. (2004). SimSE: an interactive simulation game for software engineering education. *Proceedings of the 7th International Conference on Computers and Advanced Technology in Education (CATE)*, (pp. 12-17). Kauai, Hawaii.

Calderón, A., Ruiz, M., & O'Connor, R. V. (2018). A multivocal literature review on serious games for software process standards education. *Computer Standards & Interfaces*, *57*, 36-48.

Oliveira Chaves, R., von Wangenheim, C. G., Costa Furtado, J. C., Ronaldo Bezerra Oliveira, S., 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.

Pedreira, O., García, F., Brisaboa, N., & Piattini, M. (2015). Gamification in software engineering - A systematic mapping. *Information and Software Technology , 57* (1), 157-168.

Petersen, K., Feldt, R., Mujtaba, S., & Mattson, M. (2008). Systematic mapping studies in software engineering. *Proceeding of the 12th International Conference on Evaluation and Assessment in Software Engineering (EASE)* (pp. 68-77). UK: British Computer Society.

Petri, G., & von Wangenheim, C. G. (2017). How games for computing education are evaluated? A systematic literature review. *Computers & Education , 107*, 68-90.

Sánchez-Gordón, M.-L., O'Connor, R. V., Colomo-Palacios, R., & Sanchez-Gordon, S. (2016). A learning tool for the ISO/IEC 29110 standard: understanding the Project Management of Basic Profile. *Proceedings of the 16th International Conference on Process Improvement and Capability dEtermination (SPICE)* (pp. 270-283). Dublin: Springer International Publishing.

Skulmoski, G., Hartman, F., & Krahn, J. (2007). The Delphi method for graduate research. *Journal of Information Technology Education , 6*, 1-21.

Tom, E., Aurum, A., & Vidgen, R. (2013). An exploration of technical debt. *Journal of Systems and Software , 86* (6), 1498-1516.

Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. *Proceeding of the 18th International Conference on Evaluation and Assessment in Software Engineering (EASE)*. Article n. 38, London, England, UK: ACM.