Training students in evidence-based software engineering and systematic reviews: a systematic review and empirical study

Article *in* Empirical Software Engineering · May 2021

DOI: 10.1007/s10664-021-09953-9

Sebastián Pizard
Universidad de la República de Uruguay

Fernando Acerenza
Universidad de la República de Uruguay

Ximena Otegui
Universidad de la República de Uruguay

Silvana Moreno
Universidad de la República de Uruguay

This is a preprint of an article accepted to be published in the Empirical Software Engineering Journal (Feb 2021). The final authenticated version will be available online at: https://doi.org/10.1007/s10664-021-09953-9

Training students in evidence-based software engineering and systematic reviews: a systematic review and empirical study

Sebastián Pizard · Fernando Acerenza · Ximena Otegui · Silvana Moreno · Diego Vallespir · Barbara Kitchenham

the date of receipt and acceptance should be inserted later

Abstract *Context:* Although influential in academia, evidence-based software engineering (EBSE) has had little impact on industry practice. We found that other disciplines have identified lack of training as a significant barrier to Evidence-Based Practice. *Objective:* To build and assess an EBSE training proposal suitable for students with more than 3 years of computer science/software engineering university-level training. *Method:* We performed a systematic literature review (SLR) of EBSE teaching initiatives and used the SLR results to help us to develop and evaluate an EBSE training proposal. The course was based on the theory of learning outcomes and incorporated a large practical content related to performing an SLR. We ran the course with 10 students and based course evaluation on student performance and opinions of both students and teachers. We assessed knowledge of EBSE principles from the mid-term and final tests, as well as evaluating the SLRs produced by the student teams. We solicited student opinions about the course and its value via a student survey, a team survey, and a focus group. The teachers' viewpoint was collected in a debriefing meeting. *Results:* Our SLR identified 14 relevant primary studies. The primary studies emphasized the importance of practical examples (usually based on the SLR process) and used a variety of evaluation methods, but lacked any formal education methodology. We identified 54 learning outcomes covering aspects of EBSE and the SLR method. All 10 students passed the course. Our course evaluation showed that a large percentage of the learning outcomes established for training were accomplished. *Conclusions:* The course proved suitable for students to understand the EBSE paradigm and to be able to apply it to a limited-scope practical assignment. Our learning outcomes, course structure, and course evaluation process should help to improve the effectiveness and comparability of future studies of EBSE training. However, future courses should increase EBSE training related to the use of SLR results.

S. Pizard · F. Acerenza · S. Moreno · D. Vallespir
Engineering School, Universidad de la República, Montevideo, Uruguay

X. Otegui
Teaching Department, Engineering School, Universidad de la República, Montevideo, Uruguay

B. Kitchenham
School of Computing and Mathematics, Keele University, Keele, Staffordshire ST5 5BG, UK

Keywords evidence-based software engineering · learning outcomes · training evaluation · systematic literature review

1 Introduction

Evidence-based software engineering (EBSE) aims to improve decision-making related to software development and maintenance by integrating the best current evidence of research with practical experience and human values (Kitchenham et al., 2004). This approach allows researchers to aggregate results from previous empirical studies and makes recommendations for professional practice. As main tools to achieve this, EBSE proposes secondary studies such as systematic literature reviews (SLR) and systematic mapping studies (Kitchenham and Charters, 2007).

At present, different initiatives suggest that EBSE can contribute to generating more applicable research results and improving the transfer of knowledge to the industry. Several authors argue that in practical areas like software engineering, general solutions are more likely to be obtained from bottom-up research and from a set of studies grounded in real development contexts than from top-down research (Basili et al., 2018). From this viewpoint, context-driven research and methods like EBSE, which allow the aggregation of empirical studies, become very relevant. In addition, some current research that attempts to address the challenge of achieving more impact with software engineering research proposes EBSE, among other approaches, to identify and select knowledge to transfer to practice (Badampudi et al., 2019a,b; Cartaxo et al., 2018).

However, fifteen years after introducing EBSE, there is little evidence of its adoption by industry (Cartaxo et al., 2016). For example, Hassler et al. (2014) found that lack of connection with industry is one of the most important barriers for using systematic reviews, while in a survey of Stack Exchanges users, Cartaxo et al. (2016) found that systematic reviews did not usually answer practitioners' questions. In a tertiary study of 120 systematic reviews, Da Silva et al. (2011) found only 32 that included recommendations for users. Subsequently, in a survey of 44 authors of 120 systematic reviews, Santos and Da Silva (2013) found that most SRs published before the end of 2013 had an academic motivation, and only six participants confirmed that their research had had a direct impact on industrial practices. In addition, Kitchenham et al. (2015) mentioned only a single report of direct application of EBSE in industry (Kasoju et al., 2013).

In other disciplines, in which the adoption of evidence-based practice (EBP) is also being studied, findings show the critical importance of appropriate training. In their systematic review, Upton et al. (2014) place the lack of knowledge and skills among the first five barriers that occupational therapists encounter when implementing EBP. Similarly, in another systematic review, Scurlock-Evans and Upton (2015) found training was in the top five facilitators for the adoption of EBP by social workers. The situation is also similar in the health area, where several systematic reviews also placed the lack of knowledge and skills as one of the most commonly reported barriers to adopting EBP by health professionals (Zwolsman et al., 2012; Sadeghi-Bazargani et al., 2014). Aglen (2016), meanwhile, focuses her review on pedagogical strategies to teach EBP in nursing. She found that much remained to be done, for example, teaching how evidence is used, and better adapting EBP teaching to students' learning prerequisites. More recently, studies have been carried out that seek to define and analyze competences in EBP for health professionals (Albarqouni et al., 2018; Saunders et al., 2019).

From the experiences of other disciplines, we conclude that it is important to provide EBSE training and the objective of this paper ***to develop and evaluate an EBSE training initiative appropriate for delivery in a university environment***. To achieve our objective, we have undertaken a series of three research activities:

1. We undertook an SLR aimed at assessing previous EBSE training initiatives which influenced both the development of the course and its evaluation.
2. We developed an EBSE course, with emphasis on SLR process, using the learning outcome approach aimed at codifying the knowledge and skill required of future EBSE users.
3. We delivered the EBSE course and evaluated it based on the students' performance and opinions.

This article is organized as follows. Section 2 presents the systematic review of EBSE teaching initiatives. Section 3 includes case study goals and context. In Section 4, we present the undergraduate EBSE teaching proposal. We explain the construction of the LOs for the course together with the course principles and structure. We regard our development, delivery, and evaluation of the EBSE course as a case study and have based our approach on Runeson and Höst (2009)'s guidelines. The case study is reported in Sections 5 to 9. In Section 5, issues related to participants selection and ethics are presented. Section 6 reports the data sources and the methods used for data collection and analysis. Case study results and discussion are included in Section 7 and Section 8, respectively. The threats to validity of our work are presented in Section 9. Finally, Section 10 presents the conclusions and future research.

2 SLR of Training Students in EBSE

In order to obtain a detailed understanding of previous research related to EBSE training, we conducted an SLR in July 2017. This exercise discovered 13 relevant articles relating to 11 unique research studies. After that, the SLR was updated two times: one in August 2018 (which discovered three more unique articles), and more recently, in December 2019 (no new studies were found). The first 11 unique studies were used as references to develop our teaching proposal. The background and discussion in this paper have been updated to include data from the new primary studies. We used Kitchenham et al. guidelines (Kitchenham and Charters, 2007; Kitchenham et al., 2015) for SLR planning and implementation.

2.1 Aim and Research Questions

The SLR aimed to determine how EBSE is taught and how EBSE teaching is evaluated. In order to achieve this, we defined the following research questions (RQ):

RQ1 Which EBSE teaching initiatives have been reported?

RQ2 In what context (academic program/courses/etc.) is it taught?
RQ3 What is the content taught and what are the methodologies used to teach it?
RQ4 What are the assessment tools used?
RQ5 What are the difficulties found and what are the recommendations provided?
RQ6 What are the benefits for students?

2.2 Methodology

The SLR protocol was developed by Pizard and Moreno and reviewed by Vallespir and Acerenza. The SLR was conducted by Pizard and Moreno. If differences were found during study selection or data extraction, Vallespir was consulted.

2.2.1 Search and selection process

In a first stage, Pizard performed automatic searches on selected scientific databases and Moreno validated all of them. The search string was first developed and agreed in the initial protocol and later updated to ensure that the maximum number of known studies were found. Even so, some known studies could only be found by snowballing because they were not indexed. The search terms are clustered in one bundle: title, abstract, and keywords for teaching, evidence-based or secondary studies, and software engineering. The search string presented in Table 1 was used in all of the searches, though some adaptations were made to it due to differences in the digital libraries. We supplemented the automatic searches with backward and forward snowballing and with manual searches in Google Scholar about of all the publications by the authors of the selected articles.

The selection process was carried out using the following criteria: Inclusion - articles that report on EBSE teaching initiatives (whether it is its main focus or not), and related to teaching SE/CS students; Exclusion - descriptions of keynotes, workshops, or articles that are not in English; articles whose full text is not available.

In a first stage, we independently read the titles and abstracts to discard those that did not meet the criteria. In a second stage, we read the complete text of the selected articles, in order to obtain the set of studies to be analyzed. Table 2 shows the results of both stages.

After completing the two stage search and selection process we identified 12 primary studies. In order to further reduce the probability of missing relevant studies we undertook two further search and selection procedures. Firstly, we performed backwards and forwards snowballing (Wohlin, 2014), where candidate articles were

Table 1 Search string

((teach OR learn OR education OR train OR students) AND (”evidence-based software engineering” OR ”evidence based” OR EBSE OR ”systematic literature review” OR ”systematic review” OR ”literature review” OR SLR OR ”systematic mapping” OR ”mapping study” OR ”scoping study” OR SMS) AND (”software engineering”))

	First Stage						Second Stage
Search	Papers Agreed Include	Exclude	Papers Disagreed	Papers Total	Kappa	Papers Selected	Papers Selected
2017	11	91	7	109	0.732	18	10
2018	2	162	2	166	0.661	4	0
2019	3	109	1	113	0.853	4	2

Table 2 Results of selection process

searched on the site where they were published (if available), and in both SCOPUS and Google Scholar. After completing the snowballing, we searched for other relevant papers published by the authors of the primary studies using Google Scholar.

Figure 1 presents a summary of the search and selection process for primary studies, not showing repeated studies from previous searches (by engines, from left to right, or by dates). The 16 selected publications included two examples of multiple publications related to the same study. Multiple reports were analyzed as a single study.

2.2.2 Data extraction and synthesis process

As a first step, Pizard and Moreno extracted data concerning the authors, title, publication venue, and publication date. Subsequently, Kitchenham proposed an extended categorization scheme (see Appendix I) and a synthesis method based on following the Miles and Huberman's Qualitative Data Analysis method (Miles et al., 2014). Pizard produced a revised data extraction form based on Google

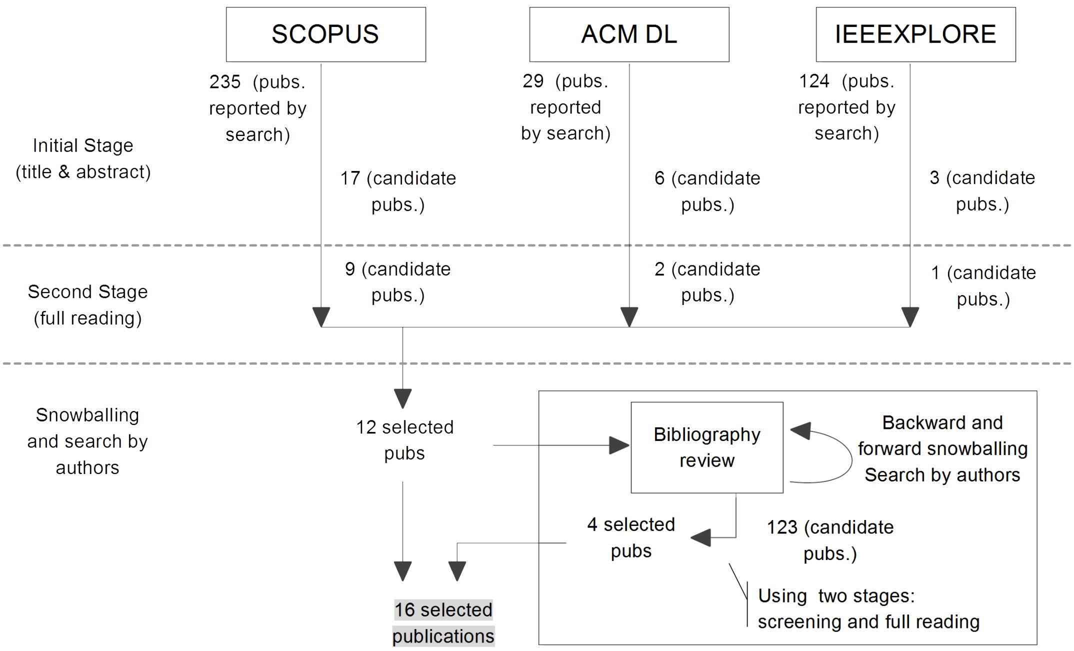

Fig. 1 Surveyed literature flowchart.

spreadsheets that was tested on some of the primary studies. Pizard then performed the data extraction and synthesis to present the results in tabular format. In addition, content analysis and open coding (DeFranco and Laplante, 2017; Elo and Kyngäs, 2008) were used to identify and categorize difficulties and benefits (RQ5 and RQ6 respectively).

To validate the extracted data, Moreno and Pizard performed a lean peer review as recommended by Garousi and Felderer (2017). This type of review involves selecting a random set of papers and reviewing them interactively by asking questions, while the other researcher explains the extraction. Reliability analysis of the data extraction process is included in Appendix I.

2.2.3 Quality assessment process

Kitchenham proposed a quality assessment of all the identified studies. Because the primary studies were of different types, we used the same questions as Kitchenham and Brereton (2013), which were originally used by Dybå and Dingsøyr (2008), see Appendix I. Pizard and Moreno independently assessed the quality assessment criteria for each primary study. In a meeting, all disagreements were resolved. Quality extraction was done in parallel to data extraction. Reliability analysis of the quality assessment process is included in Appendix I.

2.3 Reported Initiatives and their Context (RQ1, RQ2)

Table 3 presents the selected studies and their general characteristics and Table 4 presents the context of each EBSE's teaching initiative.

Half of the studies have a main objective related to the teaching of EBSE, while the rest seeks to study the EBSE process or study attitudes towards the approach. The studies were published between 2005 and 2018, but all the reported EBSE training courses took place prior to 2014. They were carried out by universities in seven countries with an important participation of the UK. The studies report experiences with postgraduates, both MSc and PhD candidates, and undergraduate students. They also present a diverse context of program areas and course focus in which these initiatives were carried out (see last two columns of Table 4).

The quality of the studies, with the exception of three of them, is above 60. When analyzing the quality by type of study (see Figure 2), Lessons learned scored worse (including the three cases below 60). We can assume that this is due to the lack of defined processes for such studies. The quality scores tend to favor papers that adhere to a well-defined process. We observe that Kitchenham and Brereton (2013) in their review of reports on the execution of secondary studies present a quality score by type of study somewhat higher than ours. We believe this is because software engineering education studies do not have as many guidelines as empirical studies do. Studies found in our review vary greatly in length, rigor, and the way they report their research. We also note that much of the information we required, and that we would suppose basic for a report of an educational experience (e.g., the number of students), was not included in some of the papers.

Id	Paper	Summary of aims of the study	Main motivation	Type of study	Overall quality (% of relevant questions)
S1	Ribeiro et al. (2018)	To investigate similarities and differences, and to characterize the challenges and pitfalls of the planning and generated outcomes of SLR research protocols dealing with the same research question and performed by similar teams of novice researchers	EBSE/SLR process issues	Case study	100 x 7.5/8 = 93.8
S2	Lavallée et al. (2014)	To present an iterative approach for conducting systematic literature reviews that addresses the problems faced by novices	EBSE/SLR process issues	Lessons learned	100 x 4.5/5 = 90
S3	Catal (2013)	To present the perspective on teaching EBSE as a single lecture within a course instead of an entire semester-long course	teaching EBSE/SLR	Lessons learned	100 x 0.9/5 = 17.5
S4	Castelluccia and Visaggio (2013)	To report experiences about teaching EBSE to master students	teaching EBSE/SLR	Lessons learned	100 x 3.1/6 = 52.1
S5	Carver et al. (2013)	To identify the most difficult and time-consuming phases of the SLR process	EBSE/SLR process issues	Case study	100 x 7.9/8 = 98.4
S6	Brereton (2011)	To explore the effectiveness of undergraduate students in carrying out a systematic review and identifying difficulties	teaching EBSE/SLR	Case study	100 x 7.1/8 = 89.1
S7	Kitchenham et al. (2010)	To assess the educational and scientific value of students undertaking a mapping study	attitudes to EBSE/SLR	Opinion Survey	100 x 7.8/8 = 96.9
S8	Oates and Capper (2009)	To teach SRs and EBSE topics, provide an experience report and empirical data, and investigate the results.	teaching EBSE/SLR	Case study	100 x 7.1/8 = 89.1
S9	Brereton et al. (2009); Turner et al. (2008)	To evaluate the applicability of an SLR by a master student in 13 weeks and to aggregate evidence about the effectiveness of pair programming for teaching introductory programming	attitudes to EBSE/SLR	Lessons learned	100 x 3/6 = 50
S10	Janzen and Ryoo (2009, 2008)	To report a course that incorporated EBSE topics and produced a community-driven Web database of study summaries	teaching EBSE/SLR	Opinion Survey	100 x 5.3/8 = 65.6
S11	Baldassarre et al. (2008)	To describe how students have been introduced and addressed to carrying out systematic reviews	teaching EBSE/SLR	Opinion Survey	100 x 5/8 = 62.5
S12	Rainer and Beecham (2008)	To empirically evaluate the use of EBSE by undergraduate students. To study how to apply findings on the practice of EBSE by students to professional practice. To obtain feedback in the use of EBSE guidelines and assessment schemes.	attitudes to EBSE/SLR	Case study	100 x 6/8 = 75
S13	Rainer et al. (2006)	To conduct a empirical investigation of the use of EBSE by undergraduate students	attitudes to EBSE/SLR	Lessons learned	100 x 4.3/6 = 70.8
S14	Jørgensen et al. (2005)	To report the Lessons learned from teaching EBSE	teaching EBSE/SLR	Lessons learned	100 x 4/5 = 80

Table 3 General characteristics of studies on EBSE teaching initiatives

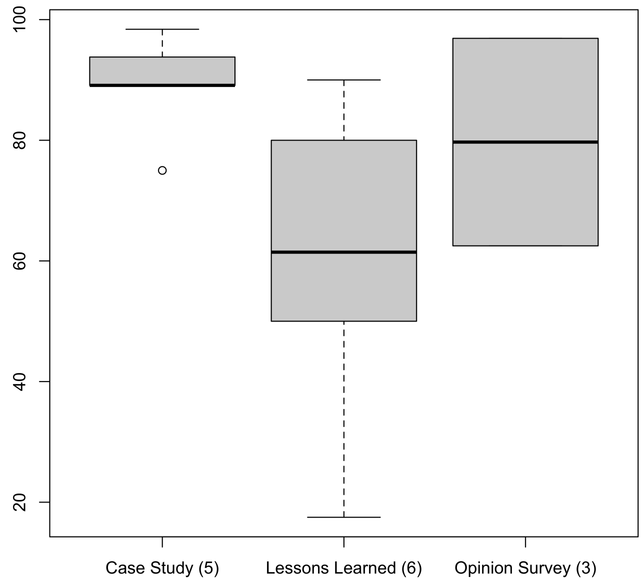

Fig. 2 Quality score for types of study (number of studies in parenthesis).

2.4 Content, Methodology and Assessment (RQ3, RQ4)

The most common educational approach was a brief introduction (1 to 3 classes) followed by a practical assignment (9 studies), although alternating introduction of concepts and practice or longer lessons and a practical assignment were also used (see Table 5). None of the studies identified any educational theory used to underpin their teaching approach.

All initiatives included a practical assignment (see fourth column in Table 5). In most cases, it involved participating in the execution of secondary study, i.e. an SLR, a limited SLR, or a mapping study (from now referred to as training studies). In some cases, it involved performing EBSE steps, that is, identifying a problem and trying to address it using scientific evidence, practical experience, and customer's values. In one study, the students wrote summaries of primary studies which they later arranged in a summary registration system (S10).

The training studies were conducted individually, in groups of students, or with the whole class working together. In half of the primary studies, the teachers limited the scope by setting a specific topic to study, while in others the scope was limited by omitting some stages (in S6 there is no quality assessment, in S4 the students worked on a subset of recovered articles). In another case (S1), a semi-built protocol with suggested questions and terms was used.

As presented in Table 6, evaluation approaches included marking student reports, teacher evaluation of EBSE or SLR outcomes, and giving students questionnaires to describe their experience. There is no indication that one method is inherently better.

Regarding evaluation approaches, the studies lack the following aspects:

- They did not include analysis of the evaluation methods or their limitations.

Id	University	Country	Year of the study	Number and type of students[b]	Program area	Course focus[e]
S1	Universidade Federal do Rio de Janeiro	Brazil	2010-2012	7 PhD, 14 MSc (PG)	CS and not CS	Empirical SE
S2	Polytechnique Montréal	Canada	2010-2012	24 PG	-	SE
S3	Istanbul Kültür University	Turkey	2013[a]	MSc	-	Software architecture
S4	University of Bari	Italy	2013	MSc	CS	EBSE/SLR
S5	The University of Alabama	USA	2012	8 PhD	CS and not CS	Empirical SE
S6	Keele University	UK	2008	44 UG	CS and not CS	Integrated modules
S7	Durham University	UK	2010[a]	3 UG, 3 PhD	-	Integrated modules
S8	Teesside University	UK	2008	52 MSc	not CS	Research methods
S9	Keele University	UK	2008	1 MSc	not CS	Individual projects
S10	California Polytechnic State University	USA	2007	13 MSc	CS	SE
S11	University of Bari	Italy	2008[a]	MSc	-	Empirical SE
S12	University of Hertfordshire	UK	2007	20\12 UG [c]	CS	Empirical SE
S13	University of Hertfordshire	UK	2005	15 UG[d]	CS	Empirical SE
S14	Hedmark University	Norway	2003-2005	30-60 UG	not CS	EBSE/SLR

[a] The authors do not specify the year of the study, so the paper publication year is included here.
[b] PhD: PhD candidate student, MSc: MSc degree student, UG: Undergraduate student, PG: Postgraduate student
[c] 37 students, 20 courseworks were studied and 12 students responded the feedback questionnaire.
[d] 39 students, 7 used to build checklist and 15 courseworks were studied.
[e] Integrated modules: modules that cover a variety of topics (usability, professional practice, teamwork and empirical methods in S6, or elements of physics and computer science programs in S7), Individual projects: individual work of medium and broad-scope (e.g., capstone projects).

Table 4 Context of EBSE teaching initiatives

– None of them included individual written tests, nor is it clear if any studies included theoretical and not only practical aspects in their evaluations.

2.5 Difficulties and Recommendations (RQ5)

The common issues (see Table 7) mentioned by at least two studies are:

The time and effort required are a limitation in the practical work of the students. The students' assignment generally involves carrying out an SLR or a mapping study (see previous section). This is not only a student issue, undertaking a secondary study is also time/effort consuming when done by non-students as reported by Kitchenham and Brereton (2013).

Students can do SLRs/Mapping studies. Although only 4 out of the 14 studies claim that novices can do secondary studies, only two of the rest of the studies include arguments that might indicate otherwise. In S1, the authors suggested that novices' inexperience generates inconsistencies in their protocol, and in the

Educational approach	Type of lessons	Study	Scope of the study	Classroom hours	Extra hours	Elapsed time	Partitipation criteria
Brief introduction (1 to 3 classes) plus practical assignment	Lectures	S1	SLR	-	Two months	Two months	Mandatory
		S3	Mapping Study	A 2-hrs lecture	Two months	Two months	-
		S7	Mapping Study	-	50 hrs	-	-
		S8	SLR limited	1-h lecture	-	Two semesters	Compulsory
	Lectures and tutorials	S6	SLR limited	3 hrs of lectures an tutorials, 6 hrs of timetabled access to teaching assitants	46 hrs approx.	6 weeks	Mandatory
		S12	EBSE steps	-	30 hrs approx.	6 weeks	-
		S13	EBSE steps	-	30 hrs approx.	6 weeks	-
	Seminars	S10	Other scope	-	-	-	-
	-	S4	Mapping Study	-	-	One semester	-
Alternating introduction of concepts and practice	Lectures	S2	SLR	A weekly 3-hrs lecture	One semester	One semester	-
		S14	EBSE steps	18 hrs	6-8 hrs per week	11 weeks	Mandatory
Longer lessons plus practical assignment	Lectures and tutorials	S11	SLR limited	10 lessons	-	-	Mandatory
-	-	S5	SLR limited	-	-	One semester	-
		S9	SLR limited	-	13 weeks	13 weeks	-

Table 5 Content and Methodology of EBSE teaching initiatives

Table 6 Evaluation approaches

Study	Student reports	EBSE/SLR outcomes	Student questionnaire	Not stated
S1	Team	Yes		
S2				Yes
S3	Individual			
S4	Ind. and Team			
S5	Individual			
S6	Ind. and Team	Yes	Yes	
S7			Yes	
S8	Ind. and Team	Yes		
S9				Yes
S10			Yes	
S11			Yes	
S12	Individual	Yes	Yes	
S13	Individual	Yes		
S14		Yes		

Common issues	Reported by
Time and effort required in practical assignments is a major problem	S5, S6, S7, S12, S14
Novices can do SLRs/Mapping studies	S3, S7, S8, S12
Search of studies is difficult for students	S1, S7, S9, S12
An iterative approach to conduct secondary studies can help students	S2, S5, S8
Value of teaching SLRs as a team project	S4, S6, S11
The research question of practical assignment should be focused	S5, S6

Table 7 Common issues and recommendations

execution of their review, they do unnecessary work and omit relevant information in their report. The authors even conjecture that SLRs are not reliable when carried out mainly by novices. This analysis seems harsh to us, given their report of the teaching process. The researchers in this study did not appear to monitor the novices during the process, nor did they offer advice or encourage iteration if processes were not properly completed. We believe that there needs to be a proper teaching method to make sure students do not compound misunderstandings or errors during the SLR process. In addition, the authors of S13, in what they call a preliminary investigation, obtained inconsistencies between their qualitative and quantitative results, and suggest that students tend to use EBSE superficially. However, in a continuation of their research (S12) two years later, they indicate that students managed to use EBSE effectively although it was a very challenging activity.

Searching for studies can be difficult for students. In this issue the researchers of the different studies include different stages of the SLR process, from the elaboration of the search string to the selection of articles, using inclusion and exclusion criteria. In addition to the clear inexperience of the students, the difficulty in searching could also be associated with how inappropriate the functionalities of digital libraries are (or were at the time) to undertake secondary studies, an issue also found by Kitchenham and Brereton (2013).

An iterative approach can help students. EBSE and domain novices can benefit from an iterative approach. The protocol can be adjusted as the review progresses and the students gain better domain perception and improve their EBSE knowledge. Instructors can also measure student progress and adjust their effort by removing or adding activities or iterations.

Value of teaching SLRs as a team project. Conducting a secondary study is challenging and time-consuming, due to this, several authors agree that teamwork seems like an appropriate approach. In fact, adopting team working is consistent with normal practice where SLRs require at least two-person teams to cater for search, select, and extraction validation processes. In addition, students may pay more attention when carrying out the SLR stages if they know that they have to present their results to the other members of their team or to the entire class.

Focusing the research questions is a key success factor. An adequate scope is very important so that the students can successfully complete the practical assignment, without requiring more effort than stipulated.

2.6 Benefits for Students (RQ6)

As shown in Table 8, on the benefits of an EBSE training there are more claims by the authors than results. The objectivity of the reported benefits worsens considerably if we consider that only the S7 study has sufficiently rigorous data collection and data analysis. Despite all this, the most common benefits are: learning how to search the literature and organizing results, learning about empirical studies, and learning how to assess the information on a topic.

2.7 Discussion of Findings

The 14 papers were extremely varied in their goals and methodology. This means that there is little to be gained by trying to aggregate the results into some overall model. Our approach has been to review the papers from the viewpoint of our research goal which is to develop a training initiative that can be delivered in a university environment. Despite this, in this subsection we include a very brief discussion of some important points.

Context of the training. Only two studies report courses specifically aimed at teaching EBSE. This may be because there is a lack both of detailed guidelines for conducting the EBSE steps, and of reports of EBSE use in industry, which makes EBSE training difficult. It is also the case that curricula guides for undergraduate students in CS and SE do not consider the issue of evidence-based practice (Joint Task Force on Computing Curricula - ACM and IEEE Computer Society, 2013, 2014).

Benefit	Claimed as possible benefits by	Reported as results by
Learn how to search the literature and organize results	S4, S12, S14	S7
Learn (more) about empirical studies	S4, S6, S10, S12	
Learn how to assess the relevance, validity or quality of the information on a topic	S6, S14	S3
Acquire or improve research skills		S3, S7
Become aware of the value of aggregating evidence	S6	S11
Practice the use of digital libraries	S6	S3
Improve critical and systematic evaluation of arguments	S6, S14	

Table 8 Common benefits

Scope of the training studies. Training was mostly based on giving students practical assignments, only in the three oldest studies did the student assignment include working on the EBSE steps. In the rest of the studies, the students participated in the execution of all, or part of, a secondary study. Again this might be due to a lack of detailed EBSE guidelines, but it may also be because many particpants were post-graduate students, and systematic reviews are a standard scientific research method which fits well into academic post-graduate training courses.

Benefits to students. Several studies include potential benefits of EBSE training, although very few of them are derived from the results obtained. The most reported benefits are: learning how to search the literature, learning about empirical studies, and learning how to assess information on a topic. These results are consistent with those of Aglen (2016) who reported that EBP training in nursing contributed to developing information literacy skills, i.e. the ability to identify the need for information, how to find relevant information and how to use it (Brettle and Raynor, 2013). More rigorous research on the real benefits to students after EBSE training would be very interesting, especially to motivate further training and to assess the possible inclusion of EBSE in CS and SE curricula.

Students' challenges and recommendations. Several studies mention difficulties encountered or recommendations for future initiatives. In this regard, the evidence seems to indicate that novice students can undertake secondary studies. However, the time and effort required are a limitation for the practical assignment, and searching for studies can be difficult for students. Using a project-based approach with iterations and well-focused research questions appear to help the teaching of EBSE.

Negative effects of the training. Although, seven of the papers pointed out the difficulty of applying the technique (i.e., due to time and effort, or due to problems searching the literature), none of the studies suggested that EBSE or SLR training was harmful to students (e.g., causing them to doubt their ability if they had problems, or to miss the opportunity to take courses more directly related to developing CS/SE skills). Furthermore, five identified positive benefits. Thus, we were confident that undertaking a training initiative would not be detrimental to our students even if they were never in a position to undertake an SLR or personally adopt EBP.

Recommendations to researchers. Although the quality of most of the studies qualifies as good, much information necessary to understand the teaching initiatives, e.g. the number of students or details of the teaching method, was not included in the publications. We suggest that future studies should try to be clear about their aims and we also recommend researchers to adopt a well-defined strategy for evaluating the results of the study against those aims. In all cases, student participants should be asked to assess the value of the training they have received. Finally, we encourage researchers to consider the ethical aspects involved in research in educational settings. In fact, we recommend including a question about ethics to any quality evaluations of studies carried out in educational settings to ensure the educational experience (not solely the anonymity) of participants is properly safeguarded.

2.7.1 Impact of the SLR results on our case study

As a result of the information obtained from the SLR, we decided to undertake a case study to develop and evaluate a proposal for teaching EBSE. The case study has the following characteristics:

- *Program area and type of students.* In our own university we have a 5-year degree, this is quite different from other universities where a 3-year degree is more common. However, the results of the review suggest that both undergraduates and postgraduates can be trained in EBSE. So we decided to design an EBSE course (with emphasis on SLR process) for our undergraduates, who can take it optionally in the fourth or fifth year of their degree. In this way, we take advantage of our opportunity to provide more extensive and more intensive training than is normally possible for postgraduate courses. In addition, designing a course for our undergraduates allows us to ensure that students have similar prior knowledge, something that we could not verify for our graduate programs.
- *Educational theory.* Unlike previous studies, we decided to incorporate learning outcomes that allow better traceability of both the purpose of the training and the results of its execution. We present the basic theory and practice of LO in Appendix II and detail how we used it for our course in Section 4.1.
- *Educational approach.* Despite the fact that in previous initiatives the most used educational approach was a brief introduction followed by a practical assignment, we decided to use an alternating introduction of concepts and practice. We believe that this approach would better monitor student progress and their learning achievements. It would also allow students to iterate or re-run previous steps in their assignments if problems arose (an approach explicitly recommended by three studies, see Table 7).
- *Practical assignment.* The practical assignment element of our proposal is the execution of a limited SLR as a team project, a strategy used by several previous initiatives and also reported as a recommendation (see Table 7). Teams were allowed to choose their own topic. Working on a topic of their choice gives students additional motivation throughout the process. Our decision on this issue was influenced by the fact that the study that reported the worst outcome from student training imposed a specific topic (S1).
- *Teaching materials.* The main material of our proposal is the book by Kitchenham et al. (2015). It has not been used in any other reported initiative since they all were carried out prior to its publication.
- *Evaluation approach.* In previous initiatives, evaluation approaches included marking project reports, teacher evaluation of outcomes, and giving students questionnaires to describe their experience. We have incorporated all those approaches as well as written tests, a focus group moderated by the course teachers, and the results of a debriefing meeting held by the course teachers allowing triangulation of evaluation information from a variety of sources and perspectives.
- *Researcher/Teacher bias.* Two studies reported that researchers evaluating their own students or their own teaching methods was problematic. To minimize this bias in our study, Otegui and Vallespir, who were not teachers of the course, helped in preparing the learning outcomes, the course materials, and

the learning assessments. Additionally, Otegui carried out an analysis of the results of the learning assessments.

2.7.2 Threats to validity

Our systematic review was undertaken based on a protocol designed to reflect best practice in the conduction of systematic reviews and thus minimize standard threats to validity based on missing relevant sources and researcher bias or error. To mitigate the risk of the protocol being unsuitable, four researchers took part in its construction and validation. In the context of research bias, we also confirm that Kitchenham who was a co-author on several of the primary studies was not involved in either study selection or data extraction.

The only deviation from the initial protocol was the review of the extracted data and the data presentation tables, both suggested by Kitchenham, to improve traceability of the SLR results to the course design and case study design. This involved extracting additional data and classifying the information into new categories. The reliability analysis was updated taking these changes into account.

The decision to exclude papers written languages other than English could potentially have meant missing relevant papers. In practice, all candidate primary studies found by our search process were in English.

Our choice to use a lean peer review of textual data extraction based on a random selection of half primary studies is not the standard method data validation used for SLRs. This means that there is a potential threat to data validity. Although there was no disagreement in this review, the second reviewer asked several questions in order to understand each extraction performed and ensure that they were accurate.

3 Case Study Goals and Context

In order to continue our investigation of the skills needed to use EBSE and possible training methods, we carried out an embedded case study (Runeson and Höst, 2009). The case study involved the development, conduct, and evaluation of a teaching proposal for an undergraduate EBSE course with emphasis on SLR process and guided by learning outcomes (LOs).

3.1 Case Study Goals

The main goal of our course is to provide an effective method of EBSE training. This means that the main goal of our case study is to investigate whether the students have achieved the learning outcomes of the course. We also need to understand any problems the students had with the course content and structure, and their opinions about the relevance and value of what they learned.

Thus, our case study had the following research questions:

- RQI Does our training proposal enable undergraduate students to explain EBSE concepts and contribute to the conduct of an SLR?
- RQII How suitable were the method and materials used according to the students' perception?

RQIII What difficulties do students observe?
RQIV What benefits do students observe?

3.2 Case Description

The course we developed is optional for the Computer Science curriculum, of the Universidad de la República, which is a five-year degree similar to the IEEE/ACM's proposal for the Computer Science undergraduate curriculum (Joint Task Force on Computing Curricula - ACM and IEEE Computer Society, 2013). The program consists of 450 credits, and has certain minimums by areas, e.g. 70 credits in Mathematics, 60 in Programming, 30 in Computer architecture, operating systems and networks, 10 in Software Engineering, and 140 in non-mandatory courses. One credit is equivalent to fifteen hours of work required by a course for the adequate assimilation of its content, including classroom hours, assisted work, and personal student work. Also, there is a suggested course path but students have certain freedoms, for example, choosing some courses before others.

As a pre-requisite for entry to our EBSE course, students must have passed the undergraduate course on software engineering. This means that students would take this course during the fourth or fifth year of the degree and have approximately 270 credits.

The purpose of the course is to teach EBSE fundamental concepts and techniques for practical use. Once the course is completed, students are expected to understand basic EBSE concepts, identify professional activity issues that may be solved by searching for evidence in the literature, assess published secondary studies on software engineering, and participate in the planning and implementation of SLRs.

One important characteristic of our proposal is the definition of learning outcomes (LOs) to guide both the design of the entire course and the method we use to evaluate the course. Appendix II presents the theory of LOs together with Bloom's levels of cognition domain. Our use of these theoretical concepts is presented in Section 4.1.

4 EBSE Teaching Proposal for University Students

The course we developed is based on a teaching proposal with a high practical workload and an alternating introduction of theoretical and practical content. The design of the teaching proposal and its evaluation process were guided by the challenges and recommendations of previous research on EBSE teaching (see Section 2).

4.1 LOs Development for EBSE Training

A central aspect of our work was the creation of LOs that covered the purpose of the course following the theory outlined in Appendix II. The process used to create LOs statments included identifying, selecting, and putting them in writing. This task was performed by Pizard, who has experience using and conducting secondary studies, with the help of Otegui, whose research field is education. This process was iterative and required the involvement of the teachers who reviewed and adjusted the LOs during the course.

One of the main difficulties in identifying LOs is related to the fact that, as teachers, we generally think in terms of what students should know. Thinking in terms of student performance represents an important conceptual change (Barkley and Major, 2016).

To identify the course's LOs, we reviewed the results obtained from the SLR on EBSE teaching initiatives (see Section 2). As a result, the fact that students could participate in the planning and execution of a secondary study was set as the purpose of the course. Therefore, LOs that promote the practical application of EBSE were identified from secondary studies process activities presented in the reference book by Kitchenham et al. (2015). This selection also considered course characteristics, for example, the target audience, prior knowledge, and duration. In this regard, for example, it was necessary to include LOs to cover basic aspects of scientific publications.

In order to write the LOs, we used the recommendations by Kennedy et al. (2007) and Stanny (2016). Emphasis was placed on explaining LOs as observable behaviors, taking into account how it would be possible to assess them during or after the course. To write the LOs, we chose verbs following the recommendations of Barkley and Major (2016), and Kennedy et al. (2007) that provide practical advice. Among other recommendations, they suggest avoiding the terms 'know' or 'understand' or 'appreciate' because these words are open to many interpretations as well as potential misinterpretations. According to this, we set as a goal that all LOs correspond to specific things that students can do or achieve. In that way, we believe that the achievement of the course's LOs can be measured and compared.

Finally, we identified Bloom's taxonomy level for each LO in order to clarify the levels covered in the course, which allowed us to redefine and adjust some LOs when producing them. In this stage, teachers were able to identify that LOs linked to higher Bloom's levels were the most suitable for practical work. Therefore, most of the LOs corresponding to Bloom's levels 4 to 6 are linked to some team activity carried out during the course.

Table 9 shows the course's LOs with their Bloom's level, grouped by thematic unit. We used these LOs as a guide for all aspects of the course, from content and reading material selection, and classroom work methodology selection, to assignments and assessment methods.

Table 9: Learning outcomes for each of the course's Thematic units.

Id.	Learning outcome	Bloom's level
LO01	Plan and conduct an SLR on a specific topic of your choice as a team	3
	Basic aspects of scientific publications	
LO02	Interpret the different sections of a scientific paper	3
LO03	Access scientific papers through digital libraries and search engines	3
LO04	Distinguish between refereed scientific literature, grey literature, scientific communication publications, and opinion pieces	2
	Evidence-based paradigm	
LO05	Describe the role evidence has on knowledge acquisition	2
LO06	Present the benefits and limitations of the evidence-based paradigm	3
LO07	Explain the purpose and context of systematic reviews	2
LO08	Explain the five steps of the process of evidence-based software engineering	2
LO09	Explain the characteristics of the software engineering discipline that have an influence on the application of the evidence-based paradigm	2

Table 9: (continued)

Id.	Learning outcome	Bloom's level
LO10	Present the restrictions and limitations of evidence-based software engineering	3
	Systematic literature reviews (SLRs) in software engineering (SE)	
LO11	Describe the different stages of an SLR in evidence-based SE	2
LO12	Compare the different types of secondary studies (qualitative SLRs, quantitative SLRs, systematic mapping and tertiary reviews)	3
	Planning an SLR	
LO13	Interpret the aspects that influence the need and feasibility of an SLR	3
LO14	Analyze the role of research questions on an SLR	4
LO15	Participate in the identification of the need for an SLR	4
LO16	Participate in the design of the research questions for an SLR	4
LO17	Participate in the validation of the research questions for an SLR	6
LO18	Describe the protocol sections of an SLR	2
	Search for primary studies	
LO19	Describe the process to define a search strategy for primary studies, including resource identification	2
LO20	Analyze the different search methods for primary studies	4
LO21	Describe the search completeness criteria in the different types of secondary studies	2
LO22	Participate in the design and implementation of the search strategy for primary studies for an SLR	4
LO23	Participate in the definition and adjustment of a string for automatic search for an SLR	4
	Study selection	
LO24	Describe the characteristics of predatory publications	2
LO25	Characterize possible primary studies as predatory publications and describing their processing	3
LO26	Analyze how to process multiple relationships between scientific papers and studies	4
LO27	Describe the activities involved in the selection of primary studies	2
LO28	Participate in the selection stage of primary articles for an SLR with multiple reviewers	4
LO29	Participate in the definition of inclusion and exclusion criteria for an SLR	4
	Assessing study quality	
LO30	Analyze the need to assess the quality of primary studies	4
LO31	Explain the concepts and activities related to the quality assessment of primary studies	2
LO32	Participate in the definition of quality assessment criteria for primary studies for an SLR	2
LO33	Participate in the quality assessment of primary studies for an SLR	4
	Data extraction from the studies	
LO34	Explain the objective and methods of extraction for the different types of secondary studies	2
LO35	Participate in the creation of data extraction forms for an SLR	4
LO36	Participate in data extraction for an SLR	4
	Mapping study analysis	
LO37	Present the objectives and main characteristics of a mapping study (process stages, classification, presentation)	2
LO38	Analyze the differences between a mapping study and an SLR	4
LO39	Participate in the classification of primary studies and the presentation of the results of a mapping study	4
	Qualitative synthesis	
LO40	Describe the purpose of data synthesis	2
LO41	Describe the two main methods used for data synthesis	2
LO42	Analyze the context in which qualitative synthesis is used	4
LO43	Describe the objective and process of narrative synthesis	2
LO44	Describe the objective and process of thematic synthesis	2
LO45	Describe the objective and process of vote counting	2
LO46	Analyze the general issues of qualitative synthesis	4
LO47	Assess the use of a specific qualitative analysis technique in an SLR studied based on a scientific publication	6
LO48	Participate in the qualitative synthesis of the primary studies of an SLR	4
LO49	Participate in answering the research questions using the results of the synthesis	6
	Report a systematic review	

Table 9: (continued)

Id.	Learning outcome	Bloom's level
LO50	Describe the objective and model structure of an SLR report	2
LO51	Participate in the production of an SLR report	4
	Knowledge translation and diffusion	
LO52	Analyze the concept knowledge translation	4
LO53	Describe the knowledge translation activities to be performed in the context of SE	2
LO54	Describe the diffusion activities to be performed in the context of SE	2

4.2 Course Development Principles

The selection of teaching methods took into account the teachers' experience and the results of the previous SLR. In this regard, most of the previous initiatives reported an initial short instruction followed by the execution of a secondary study by the students. Our course is also based on the execution of a secondary study but with some variants. First of all, as it is intended to train novices, we propose an alternating introduction of theoretical and practical content and a weekly follow-up of the students' progress in the execution of their secondary studies. In addition, based on the difficulties and recommendations reported in previous studies (see 2.5), we organized the course so that:

– Students would be helped to choose their review topic (are assisted by teachers in choosing topics with enough published evidence).
– The workload would be limited in some stages of the process.
– Students who needed to perform iterations in one or more stages of the process would be supported.

4.3 Course Structure

The course is 14-weeks long and has one non-compulsory on-site class a week that is 3.5-hours long. Table 10 shows the course schedule.

Students are advised to study chapters of the reference book by Kitchenham et al. (2015) for almost all topics of the course. In addition, the materials on predatory publications (LO24 and LO25) are Beall's criteria and list (Beall, 2012, 2013); and to introduce a qualitative synthesis technique in a practical manner (LO47), we recommend the article on thematic synthesis by Cruzes and Dybå (2011).

Students have to work on a practical team assignment, which consists of defining and conducting guided activities for an SLR. Each student team chooses the topic of the SLR and the research questions according to their interests. The teachers guide this selection so that the scope and complexity of the work can be addressed in the available time. The practical assignment is carried out in teams of two or three students, starting on week 2 and finishing with the submission of the SLR's report on week 14. Classes are organized in such a way that, each week, a stage of the SLR is covered from the theoretical point of view, and the teams carry out that stage on their SLR.

Week	Syllabus	Materials[a]	Assignments
1	Basic aspects of scientific publications, Evidence-based paradigm, SLRs in SE	1-3	Identify the sections of a scientific paper (LO02).
2	Planning an SLR	4	Define the objective and identify the need for the review to be performed by each team (LO15). Pose and validate the research questions (LO16, LO17).
3	Searching for primary studies	5	Define the search strategy for your review (LO22). Define and adjust the string for automatic search (LO23).
4	Study selection	6	Define inclusion and exclusion criteria for your review (LO29). Define the selection process and implement it, obtaining between 20 and 30 primary studies per student (LO28).
5	Study quality assessment	7	Define the quality assessment procedure (LO32) and implement it for the previously selected primary studies (LO33).
6	Data extraction from the studies	8	Define the extraction form for your review (LO35) and extract data from the primary studies obtained in the previous stage (LO36).
7	Mapping study analysis	9	Analyze the data from the primary studies and classify them according to commonly used schemes and schemes specific to your research questions (LO39).
8	Introduction to data synthesis, Qualitative synthesis	10	Individual midterm test.[b] Perform a type of qualitative synthesis on the extracted data (LO48) in order to answer the research questions established (LO49), taking into account the limitations of the review process performed.
9	Reporting a systematic review	12	Produce a report detailing the entire process and the decisions made (LO51).
10	Knowledge translation and diffusion	14	-
11-12	Work on team assignment		
13	Team assignment monitoring		
14	Deadline for team assignment and final individual test[c]		

[a] We suggest chapters from B. A. Kitchenham, D. Budgen, and P. Brereton, *Evidence-Based Software Engineering and Systematic Reviews*. Chapman & Hall/CRC, 2015.
[b] The individual midterm test consists of a written test with open-ended questions on the use of the thematic synthesis technique established in D. S. Cruzes and T. Dybå, 'Recommended Steps for Thematic Synthesis in Software Engineering', *International Symposium on Empirical Software Engineering and Measurement*, no. 7491, pp. 275–284, 2011.
[c] The final individual test consists of a written test with open-ended questions on topics discussed throughout the course, chosen by the teachers.

Table 10 Course timetable, including syllabus, materials and student assignments.

Each class has two parts: a lecture by the teachers and some assigned time for teamwork. During the first hour, teachers explain briefly the main concepts of the stage of the SLR process to be covered in that class. Students are asked to read the material for each class (see 3rd column of Table 10) beforehand, which is then summarized including, for example, questions for the students. Then, the weekly task assignment is presented, which consists of completing the SLR activities of the stage of the process discussed.

In the second part of the class, which lasts approximately two hours, each team works on their SLR, and teachers guide them according to each team's specific needs. Teams can ask questions regarding their projects whether about their current weekly task assignment or about previous ones. The two teachers in charge of the course are in the classroom during the entire class. Each team is expected to devote four hours a week outside the classroom in order to be able to follow the process.

Students can discuss their problems at any time during the course through a Moodle platform site (Rice, 2006). Teachers also use this site to publish material and answer questions.

5 Subjects Selection and Ethical Issues

Our main unit of analysis was each student, although on some occasions it was necessary to study student teams, e.g. while they were carrying out their practical assignment.

Regarding the selection of subjects, students were encouraged to take the course by a typical course information entry on the institutional website of the university. The course had a maximum capacity of 30 students. In the event that more students enrolled, the selection would be by lottery, something that in practice was not necessary.

During the first class of the course, the teachers explained to the students that they were going to conduct research on the EBSE training, and described the purpose and data collection procedures that they planned to use. In addition, the students were told that the information collected was to be treated confidentially (i.e., reports of the course assessment would not link grades or test scores to individual students or teams), that participation or not in the study would not influence their learning experience or evaluation, and that they could withdraw from the study at any time without leaving the course. All students agreed to participate voluntarily and signed an informed consent form.

6 Data Sources, Data Collection, and Data Analysis

In order to answer the case study research questions, we collected quantitative and qualitative data. We particularly considered data triangulation, both regarding collection methods and the observers' point of view. The data include the students' opinion —collected through a survey and a focus group— as well as the marks each student obtained in the course tests and SLR project. In addition, in order to present another point of view, the data include a summary of the teachers' debriefing meeting.

The survey and the focus group were carried out in the last class of the course. This class was organized in stages. In the beginning, the teachers gave each team detailed feedback on their work. The feedback for each LO included the following: the grade achieved by the team, the teachers' comments on these achievements, and also the assessment criteria. The teachers answered the teams' questions. Then, the individual test was carried out. In the end, although participation was optional, all the students participated in the survey and in the focus group (Kontio et al., 2008; Bernard et al., 2015).

6.1 Opinion Survey

To design and implement the survey, we followed the recommendations by Kasunic (2005) and Torchiano et al. (2017). The purpose of the survey was to collect the students' opinions on learning acquisition and the suitability of the content and method of the course. We identified two units of analysis: course students individually and student teams.

The target population was the group of students taking the course. Participation in the survey was optional and students were reminded that survey reports would not link scores to individual students or teams. Students signed an informed consent form when completing the survey.

6.1.1 Survey form construction

We used a form for each unit of analysis as a tool for the survey, all of them presented in Appendix III. The individual opinion form had four sections (A, B, C, and D). The first section included each student's prior experience and knowledge; the second one included general questions about the course; the third section grouped questions to survey the opinions on the achievement of the course's LOs; and the fourth introduced questions on the benefits and difficulties of the course. The opinion form for the teams had only one section (E) that included questions on motivation and difficulties, and each team filled a single form after having reached an agreement. In those questions, where it was possible, we used the Likert item format with a five-point agreement scale (Likert, 1932). From now on, the questions in the forms will be referred to with the SN format, where S corresponds to the section and N to the question number. Under this nomenclature, B1 is the first question of section B.

In the interest of improving survey participation, we used the principles listed by Smith et al. (2013). We applied the principle of reciprocity by offering 10 additional points on top of the 100 points total from the course assessments. We applied brevity by including questions that avoided open-ended questions and were as specific as possible. We applied the authority and credibility principles, along with social benefit, by having the course teachers introduce the survey and explain to students the benefits that reporting their opinions would bring to future courses.

6.1.2 Survey form analysis

Section A was meant to relate to the students' knowledge before the start of the course. Asking students to complete it at the end of the course was a mistake on our part, so we did not analyze this Section. To analyze responses to section B, we counted the number of responses in each of the five levels of agreement categories. For responses to Section C, we used response categories that identified the extent to which students felt that they had achieved each LO. We counted the number of responses in each response category. Section C asked students to comment on any of the LO's they had not achieved. These comments were collated and listed. The textual comments were not subjected to any formal content or thematic analysis. Section D asked students to rate the phases of the SLR in terms of difficulty using a five-point scale with 1 meaning easy and 5 meaning very difficult. For each phase, we counted the number of students assessing the phase in each of the response categories. The participants were also asked to identify benefits of the course in free text format. The textual responses were collated and organized by the first author into three related concepts: responses that related to learning SLR skills, responses that related to improving professional practice, responses related to improving research skills.

For Section E, relating to the team assessments, textual responses were extracted and listed. The textual comments were not subjected to any formal content or thematic analysis.

6.2 Focus Group

In the last class, we also carried out a focus group (Kontio et al., 2008; Bernard et al., 2015). In the focus group, all students talked about and discussed with the teachers the course's difficulties, as well as their opinions and suggestions for improvement. The teachers prepared a list of all the specific issues raised by the students. The comments were not subjected to any formal content or thematic analysis.

6.3 Learning assessment

We used the course learning assessments to add a more objective perspective on student achievement to the opinion survey and the focus group that covered a more subjective perspective, all within the students' point of view. During the course, there were three learning-assessment instances: an individual midterm written test, a final individual written test, and the practical team assignment.

6.3.1 Course assignments

The midterm test assessed the ability to understand and evaluate the use of a qualitative synthesis technique (LO47) (Cruzes and Dybå, 2011). A week before the test, students were asked to read a paper about qualitative synthesis technique (Cruzes and Dybå, 2011) and were advised on the details of the assessment. The test included theoretical questions and others regarding the application of the qualitative synthesis technique to the practical team assignment. For example, two questions in the midterm test were:

- What do you understand by thematic analysis?
- Do you think it is applicable to your current review? If yes, indicate how you would apply it.

The final test took place once the course had concluded and consisted of four open-ended questions on some of the topics covered in the course, including the practical team assignment. The final test assessed six LOs chosen by the teachers (LO01, LO12, LO30, LO37, LO38, and LO52). For example, two questions in the final test were:

- Explain the concept of knowledge translation in EBSE.

– Discuss the limitations of the SLR undertaken by your team during the course. You can add reflections that have not been included in the SLR report submitted.

Finally, the practical team assignment was assessed based on the SLR report submitted by each team at the end of the course. Both teachers separately marked (using a five-point achievement scale, from Not achieved to Completely achieved) each team's SLR reports using the list of LOs covered by the practical assignment. For each team, the evaluations were then unified using averages and gathering the teachers' comments on the achievement of each LO. In addition, a general average of all LOs was calculated as the final score of the practical assignment.

All three assessments sought to cover the LOs we consider most relevant to the course purpose, which is to teach EBSE fundamental concepts and techniques for practical use. Assessments covered all LOs that correspond to the practical team assignment, which allows students to learn how to participate in the conduct of secondary studies. In addition, we randomly selected other LOs that we included in individual tests to have a sample of the learning achievement of more theoretical aspects of the course. The individual midterm test had a maximum score of 10 points, the final individual test, 40 points, and the practical team assignment, 50 points. A minimum of 60 points was required to pass the course.

6.3.2 Learning assessment analysis

Each written test question related to a single LO and the mark obtained by each subject on each LO was allocated to a five-point achievement scale (from Not achieved to Completely achieved) so that it could be compared the assessment of personal achievement made by each subject. We counted the number of students marked in each category for each LO.

For the team assignment, the stages of the SLR were linked directly to specific LOs and were marked against each of those LOs. The mark obtained by each team on each LO was again allocated to a five-point achievement scale. We counted the number of teams on each scale point for each relevant LO.

6.4 Teachers' Debriefing Meeting

Finally, and with the purpose of having an additional point of view, we present the opinion of the course teachers. To collect it, the teachers held a meeting a few weeks after finishing the course in which they discussed the experience and identified things that could be changed to improve future similar courses.

6.5 Case Study Participants' Roles

The roles of the authors in the teaching process were as follows: Pizard and Otegui elaborated the LOs, Acerenza validated them. Pizard and Acerenza designed the course and learning assessments. They both taught the course. Otegui validated the course design and learning assessments before its execution.

In the case study: Pizard defined the research objectives and methods with Vallespir's validation. The data collection (student learning assessments, survey, focus groups) was carried out by Pizard and Acerenza. Pizard and Otegui did the data analysis and Acerenza validated it. All authors participated in the discussion of the results and limitations of the study.

6.6 Course Validation

Our study has some characteristics of the participant observation method. In this method, the researcher, with the aim of gaining in-depth knowledge of a topic or situation, is both participant and observer in an activity over time (Emerson et al., 2001). This method, however, though applicable for studying and describing contexts like ours, has certain limitations (Ko, 2017): it requires introspection, which can be subjective, it presents a single perspective, and it is also possible that the observed subjects modify their behavior. To mitigate the first two limitations, a researcher (Otegui) who does not share our line of research, and therefore, neither our expectations of success, collaborated in our study. Otegui validated the course design and learning assessments prior to its execution. She helped with the data analysis, giving an external perspective. She also formally validated the assessment of the students' practical work. In this validation, she found the assessment satisfactory, with the following strengths: the definition of the evaluation criteria for each LO prior to the start of the evaluation process, 80% agreement on assigned scores, and the feedback to students, that included giving and discussing with them the evaluation criteria used. The most important points to improve included: (1) creating more detailed criteria, like the LOs, so they can be clearer and more specific, and (2) transforming the evaluation guideline into an evaluation matrix or rubric, see for example Venning and Buisman-Pijlman (2013).

7 Case Study Evaluation Results

This section presents the results of the case study evaluation process. Ten students, organized into four teams, enrolled in the course and all passed; all of them participated in the survey and the focus group. The LO achievement levels derived from analyzing the student opinion forms and marking the student tests are reported in terms of percentages rather than counts. This means that for student opinions and test results 10% corresponds to one student, while for team-based opinions and project assessments, 25% corresponds to one team.

The next three subsections present the answers to the research questions using the students' point of view, and for clarity, the teachers' opinion is included separately in the last subsection.

7.1 RQI - Teaching Proposal Assessment

To assess whether it was possible to train undergraduate students on EBSE using our proposal, student achievement of LOs was evaluated by both students and teachers. In Section B and section C of the opinion form, the students stated their perception of their achievement using a five-point agreement scale, both individually and in groups. The teachers assessed the teaching proposal through the

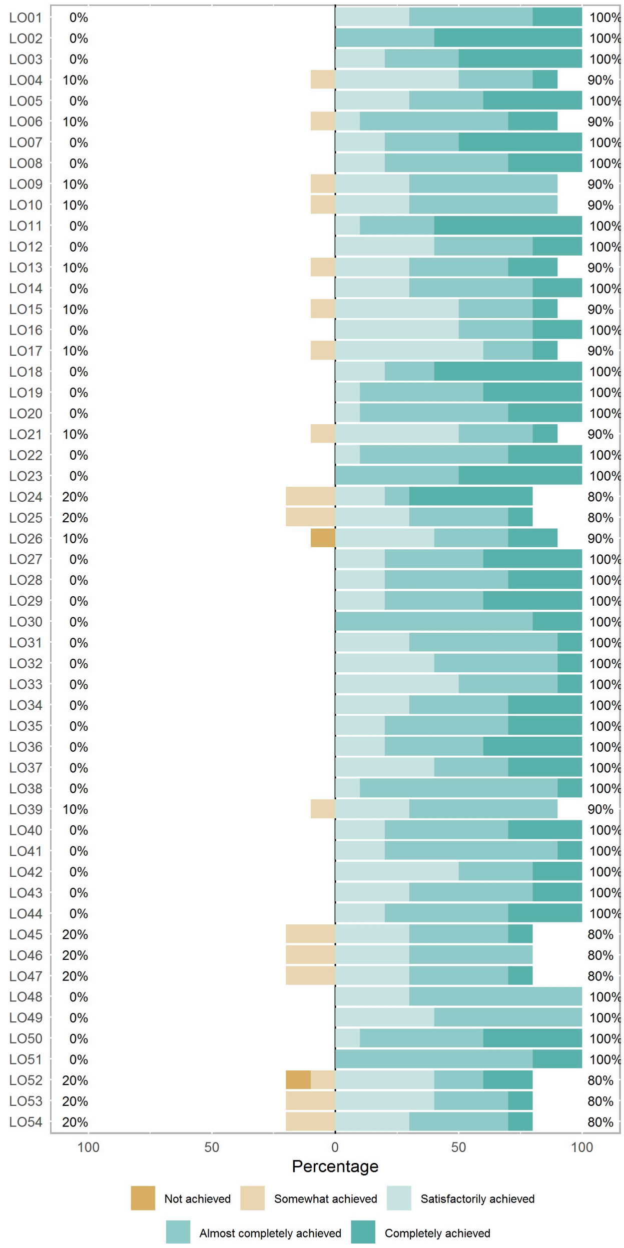

Fig. 3 LO achievement according to student opinion survey.

analysis of the learning assessments which were graded on a five-point achievement scale.

The students' perceptions of their LOs' achievement (C1-C54) are summarized in Figure 3. Sixty-seven percent of the LOs had only positive scores (3 to 5 points), 18% were scored negatively (1 to 2 points) by only one student and 15% were scored negatively by 2 students.

The LOs with two lack-of-achievement perceptions correspond to: predatory publications, some topics within qualitative synthesis, and knowledge translation and diffusion. The LOs with one lack-of-achievement perception correspond to: basic aspects of scientific publications, introduction to the evidence-based paradigm, planning an SLR, study search, study selection, and classification of studies and presentation of the results of a mapping study.

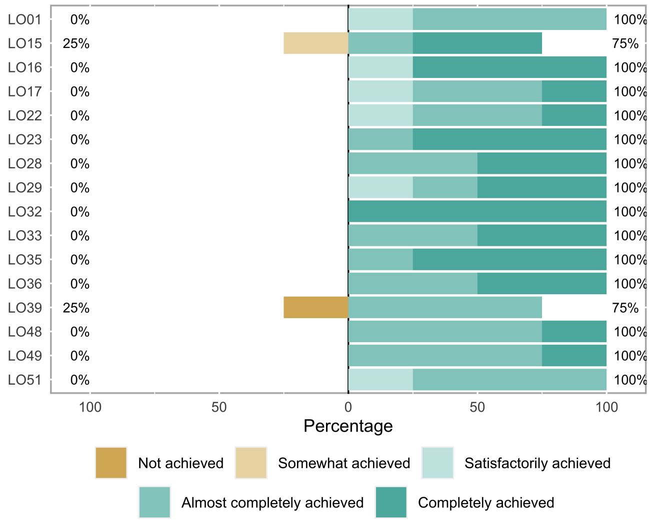

Fig. 4 LO achievement according to learning assessments from team assignment.

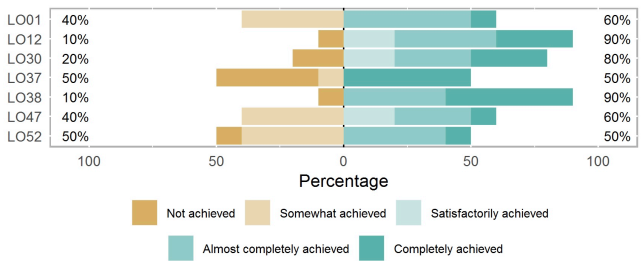

Fig. 5 LO achievement according to learning assessments from individual tests.

In the team assignment assessment performed by the teachers, there is a high rate of LOs achieved. Figure 4 shows a diverging stacked-bar graph with the learning assessment of the team assignment.

Only 2 out of the 16 LOs covered in the team assignment have negative scores. Firstly, one team performed poorly when asked to identify and explain the need for the SLR (LO15). Secondly, one team completely failed when asked to classify primary studies and present the results as a mapping study (LO39).

Individual tests assessed different LOs. The midterm test assessed LO47 and the final test assessed LO01, LO12, LO30, LO37, LO38, and LO52. Figure 5 shows the results of the scores normalized according to a 5-grade scale to make it easier to compare to the preceding figures.

Only half the students properly explained the concept of knowledge translation (LO52) and also only half the students were able to present the objectives and characteristics of mapping studies (LO37). We asked students about their SLR's limitations (in the context of LO01) and six were able to answer correctly. Also, six students were able to assess the use of a qualitative synthesis technique on their SLR (LO47).

7.2 RQII - Course Method and Materials

We analyzed the course method and materials using the opinion survey and the focus group. The summary results for all Section B questions are shown in Fig. 7. All comments students wrote on the opinion form relating to Section C are reported in Appendix IV. Regarding the overall opinion on the course, all students were satisfied: on a scale of 1 to 5, six gave it a rating of 4, and four gave it a rating of 5 (B11). In the focus group, everyone expressed a positive opinion about the weekly work dynamic, i.e., the introduction in class to theoretical concepts on a topic followed by the team assignment on the same topic.

In the survey (B6, B1) and in the focus group, students emphasized how useful the classes were. They also felt the site available on the Moodle platform was a valuable asset.

Both in the survey (B19) and in the focus group, the students noted that the book was too complicated or boring to read during the course and that this often caused them not to follow the readings recommended by the teachers.

All students stated that the assessment process focused on the understanding of the course (B10) and, during the focus group, they all positively emphasized the fact that half the course grade rested upon the team assignment assessment.

Regarding the content of the course, only one student included a negative comment. According to him (C55): *'the topic of knowledge translation wasn't enough for our understanding'*. This may also be related to the students' perceptions about their understanding of this concept and the fact that only half of the students were able to explain it in the tests.

The topics chosen by the students for their team assignment were: estimations in agile development, automatic testing, deep learning in information systems, and market value prediction using neural networks. In general, the research questions were quite open and sought to learn more about the chosen topic or make simple comparisons.

Four students considered the topics of the team assignment related to an engineer's professional life (B15). On this matter, a student stated the following (C55): *'I would've liked one of the classes to have included a guest that strongly or partially applies EBSE in their professional practice.'*

7.3 RQIII, RQIV - Difficulties and Benefits of the Course

We also assessed the difficulties and benefits of the course based on the opinion survey and the focus group. All comments students wrote on the opinion form relating to Section D and Section E are reported in Appendix IV. The students considered performing an SLR (simplified) to be difficult or moderately difficult. They also stated their greatest challenges were data synthesis and the iterative dynamics of the process. All teams highlighted that the results of their SLR seemed to be useful. Regarding the benefits of the course, three believed they are better prepared for their professional practice, whereas three others believed the course is only beneficial on an academic level.

Difficulties

Figure 6 shows the results of the survey regarding the difficulty of the EBSE process (D1).

According to the individual opinion survey (C1), half the students found it difficult to conduct an SLR while the other half found it neither difficult nor easy. Regarding the SLR stages, the stage that most students scored as difficult

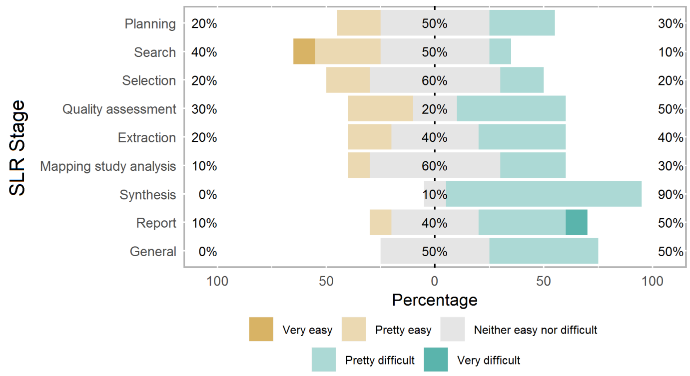

Fig. 6 Opinion survey on the difficulty of conducting an SLR.

to carry out was synthesis (nine students). Secondly, the quality assessment and review report stages were ranked as difficult by half the students.

Three teams agreed that the greatest challenge was the data synthesis (E3). Some believed it was due to a lack of experience, others that the primary studies were very different.

All teams thought the results of the SLR conducted could be useful (E5). In fact, one of the teams stated some members had already used knowledge from their review results in their professional practice.

As a course organization issue, the students indicated the limit imposed by the teachers on the number of studies to select presented a problem. Half the teams reported the need to return to previous review stages in order to obtain further primary studies for the following stages as a challenge. Also, half the teams considered there were strong limitations in the usefulness of their SLR results due to the low number of primary studies considered.

Other course organization issues were the SLR topic and the primary studies' language. In this regard, all teams agreed that getting to choose the subject of their SLR was motivating (E1). However, half the teams reported on the complexity added by choosing a subject they were interested but not experts in. Regarding language, half the teams found working with primary studies and bibliography in English difficult, but the other half said this had no influence on their work (E2).

Course benefits

The students' opinion on the course benefits (D2) includes three different points of view. Firstly, four students had a very pragmatic vision and believed the course taught them how to conduct an SLR or gain knowledge of EBSE. A student summarized the benefits of the course in the following way: *'Becoming aware of a systematic method, i.e., one that includes steps and procedures that were reviewed by experts in order to search and synthesize material that can answer questions.'*

Three students considered that, after the course, they were better prepared for their professional practice. One student stated the following as a benefit of the course: *'I gained the ability to quickly assess and absorb studies of all kinds, search and find scientific data that I am already using in other courses and at work.'*

Lastly, another three students reported that they had acquired a useful tool to conduct more reliable research after the course. For example, they considered using it for their capstone project. One of these students added: *'It is eventually beneficial for a change in technology in the professional practice, though I believe most local companies in the industry don't consider this methodology important (yet).'*

7.4 Course Assessment by Teachers

From the teachers' point of view, the course met its objectives: the students were able to understand the fundamental concepts of EBSE and participate in a limited SLR. Even so, we noticed that in some cases their performance in practice was superior to their theoretical learning. Perhaps this is due to the course's strong emphasis on practical activities or the difficulties reported when using the book. In general, it was difficult for students to understand primary studies. Most of the students had no experience in reading scientific articles or knowledge of empirical research methods in software engineering. Although the course has LOs that allow them to be introduced to scientific reading, we do not have any on empirical software engineering, something that can be added in future courses. In addition, we believe it is useful to carry out an initial survey to find out more about the knowledge that students have before starting the course.

We believe it is necessary to improve communication with students in the following two points. First, in the first class we must explain better how the course works, including how the evaluation is carried out and the score assigned to each activity. Secondly, some students complained that we sometimes gave different answers to their questions about the practical assignment. We believe that this is due to the fact that there is no single criterion to address the challenges that may arise when applying EBSE and furthermore, students are more used to CS courses where single or standard solutions are more common. This situation could be improved by explaining that both teachers will answer their doubts and questions from their own perspective and that some of their answers could be different since there are no perfect solutions.

Some topics were difficult for students to understand: mapping studies, predatory publications, and knowledge translation. Regarding the first one, we could try to present and discuss during class the analysis carried out by one of the groups. Seeing and discussing an example of how to perform the classification of papers and their presentation could help them better understand that topic. On the other two topics, perhaps including practical exercises could contribute to increased learning.

Regarding the practical assignment, we should provide more guidance to the students when they make their topic choice, so they do not select a topic completely unknown to them. This should make it easier for them to read and understand the primary studies. We might also consider giving the students advice on what to read first such as overview papers published in the IEEE and ACM magazines and existing systematic reviews on their SLR topic. Even if excluded from the SLR they can give students useful insights into the topic.

Other minor changes could be made: using a tool to facilitate communication through forums (e.g., Slack), checking that students read the textbook, and submitting the final report on Moodle.

8 Discussion

In this section, we include an additional discussion about the results obtained and their possible meaning in relation to the previous work.

Overall evaluation of our course. Our case study results suggest that our EBSE teaching proposal is suitable for preparing students with more than 3 years CS/SE training at university level to participate in the undertaking of secondary studies. Although the students did not execute a complete SLR during the course, they showed that they had acquired skills to perform the different activities of the process. Their performance in the assigned practical work (see Figure 4) and their opinion on the achievement of objectives (see Figure 3) are quite similar and give an account of the acquired skills. However, the evaluation of the individual tests (see Figure 5) shows only a minor acquisition of EBSE theoretical elements. This result seems to be in accordance with two aspects of our proposal. Firstly, the course approach is more practical than theoretical and mainly seeks to train students in SLRs activities. On the other hand, the students responded to the survey before knowing the grade of their final individual exam. This might have made their opinion on the achievement of the most theoretical objectives somewhat optimistic. This result might also indicate that students may need more time or instruments to reflect and assimilate the theoretical foundations that support the activities carried out.

The proposed LOs a basis for future agreements. We believe that the proposed LOs are suitable for training future participants in the undertaking of secondary studies. The LOs were prepared using the reference book by Kitchenham et al. (2015), previous EBSE training initiatives, and our experience conducting and using secondary studies. This set of LOs is far from being a result of consensus, as the core competencies in EBP for health professionals proposed by Albarqouni et al. (2018) actually are. Despite this, we hope that, together with their rigorous elaboration and validation, they will contribute to further discussions about EBSE training as a facilitator for its adoption.

About teaching methods. The teaching method used —based on high practical workload and an alternating introduction of theoretical and practical content— seems quite adequate. Unlike most previous studies that included a brief theoretical introduction and then a practical assignment based on the execution of a secondary study (see 2.4), our course is based on a weekly theoretical-practical advance. This approach allowed detailed monitoring of the progress of each team. Some of which had to perform iterations at some stages of their review. It is noteworthy that the choice of the SLR theme by the students was motivating for many of them, although it demanded an additional effort from the teachers associated with understanding and following the different topics chosen. Although the students were motivated to choose a topic or problem that could arise in their professional practice, it was not explicitly intended that they seek for relevant problems in the industry. However, the teachers found the topics quite close to professional practice.

About teaching materials. All students had difficulties following the textbook. It is extremely likely that the fact the book is available in English and not the students' mother tongue had an influence in its perception as half the teams stated that using bibliography in English was somewhat difficult. This might also be due to the fact that, as it is an EBSE reference book, it contains highly technical language and advanced content, thus not making it an appropriate introductory book to novices. More research is needed to study the challenges students face in using the book. Also, it may be possible that the text by Kitchenham et al. (2015) is too focused on obtaining evidence and not sufficiently concerned with using evidence for decision making. This could also be a barrier to its use by industry practitioners.

The challenge of learning about SLRs. Half of the students found undertaking an SLR difficult. In this regard, and like Rainer and Beecham (2008), we believe it is a very challenging activity and we also suspect that practitioners may find themselves in a similar situation. We believe that our methodology, and mainly the fact of limiting the workload in different stages of the process, gave the students sufficient time to execute the SLR. Even so, teachers sometimes had to help students manage their frustration. For some topics, such as data synthesis, we believe that it is necessary to find new teaching strategies and tools that allow a better understanding of the process.

EBSE Training issues and recommendations. After the first experience of running our course, we agree with most of the reported issues in previous EBSE training initiatives (see 2.5): conducting a secondary study requires substantial effort; students can do secondary studies; an iterative teaching approach with the help of teachers can help students; students may find learning secondary studies as a team project very valuable; and that the focus and scope of the research questions must address topic areas where there is a sufficient level of research. However, we found no evidence that the search for articles was difficult for the students; in fact, according to the opinion survey, it was the stage that the students found the least difficult. This could be partly explained by the advances made in search engines and digital libraries.

EBSE Training benefits. There seems to be a certain degree of consensus among students on the fact that training allowed them to obtain a different perspective (see 7.3); some see that the acquired skills can help them in their professional life in general, although others only see a certain benefit for their academic life. Despite their opinions, it is difficult to assess what actual benefits the students obtained from the course. It would be necessary to wait some time and have other instruments to evaluate, for example, if the course helped them to improve their critical thinking or if they are more reflective when searching and consulting scientific literature. In this regard, a few months after completing the course, two students contacted one of the teachers to tell him that they were doing a limited SLR to address the state of the art of their capstone project. Their project aimed to use sensor tools and software to help patients with the freezing of gait (FOG) in Parkinson's disease. The students themselves decided to conduct this review on similar previous initiatives and were already making decisions on the design of their solution based on the evidence found. This can be taken as a positive contribution of EBSE training. From the results of the case study and the experience of these two former students, we could agree with Aglen (2016) that training in EBP seems to contribute to developing information literacy skills, i.e., the ability to identify the need for information, how to find it, and use it (Brettle and Raynor, 2013).

On the motivation to teach EBSE. In one of the first EBSE-teaching reports, when reflecting on whether EBSE should be taught, Jørgensen et al. (2005) said: *'We cannot claim that we have demonstrated that teaching EBSE has a significant* *positive effect on real-world software development work (though it is our hope that it does)'*. Much research has been done since then in both EBSE teaching and in EBP teaching in general. Currently, several authors and results of systematic reviews support two statements: the importance of the evidence-based approach in support of professional practice and in the knowledge transfer from academia to industry; and the use of EBP training as a facilitator of its adoption. These statements could account for the positive effect of teaching EBP, and specifically EBSE, in professional practice.

Training as a facilitator to EBSE adoption. We acknowledge that changing industry practice is a different problem than teaching students good practice. However, it is not possible for practitioners to adopt techniques that they do not know. In addition, one issue with respect to EBSE adoption is the extent to which the individual software engineers have control over the techniques and methods they use for software development. This is possible in small start-up companies but not usually for new graduates working in companies with established quality assurance and development practices. However, software engineering methods still change as technology changes, so even in established companies, the ability to identify information about new methods and tools may be useful.

EBSE training in industrial settings. We believe that the evidence-based approach can be very beneficial for the software industry but it needs diffusion both in academic and industrial settings. It is also important to consider advances and initiatives already carried out in other areas. A very interesting study is the one carried out by Vachon et al. (2010) in which a workshop on EBP for practitioners (occupational therapists) was carried out. These authors used (with good results) reflective learning, including critical incident analysis and journal writing, to empower attendees in the use of EBP in their professional practice. The use of these techniques, along with short workshops, could be an alternative to explore when training software engineering practitioners. Also using a different or broader perspective could help in the training-adoption relationship of EBP, for example, Beidas and Kendall (2010) use a systems-contextual approach to study the effects of training not only taking into account the students but also examining the quality of the training, organizational support, and other additional variables.

9 Threats to validity

The work reported in this article has some limitations. First, since our case study can also be viewed as a type of field study, it has the limitations of this type of study (according to the categorization and analysis of Stol and Fitzgerald (2018)): results that could be strongly linked to the context and may not be generalizable, there may be no control of events and low precision in measurements. These limitations can be improved by using other strategies in subsequent complementary studies. There are also threats to validity in the review of related work submitted in 2.7.2 and in the authors' participation in the case study presented in 6.5. We discuss other limitations relevant to our specific study below.

We carried out only one EBSE course, which limited data collection. Moreover, we had a small sample size since we only had ten students. We hope that the qualitative results of this study serve as input for more generalizable reflections and future studies. Meanwhile, we are repeating the course and collecting data for future analysis.

The course was not compulsory and students were encouraged to participate in a survey and a focus group by receiving a bonus. There is some risk that the course has been taken only by students who like research in SE and EBSE. At the moment, we cannot mitigate this risk by making the course mandatory, although in future versions it will be possible to better characterize the students. To minimize the risk of students not being honest in the survey or focus group, teachers explained their purpose and the importance of giving accurate and honest answers. We also confirmed that all the information would be treated as confidential and individual anonymity in any external research reports would be maintained. We also made it clear that students were allowed to leave the final course session at any time with or without completing the surveys or participanting in the group session and that leaving the session would have no impact on their course mark.

Another specific limitation related to volunteer participation is the possibility of cooperative student behavior. Rosnow and Rosenthal (1997) refer to this behavior as the 'good subject effect'. According to them, voluntary participants tend to be motivated and willing to support the goals of the study in which they agreed to participate. In our study, this behavior might explain the discrepancy between the results of the written tests and the results of the subjective assessment

The learning assessments could include the bias of the course teachers. This type of bias was reduced by defining assessments based on LOs, using previously-defined qualification criteria, and with both teachers correcting all assessments.

The reference material of the course is not available in the students' mother tongue. Thus, we reduced the risk that students did not understand it by including activities and materials, such as slides and an introductory EBSE report, in Spanish.

Finally, another limitation that our course has is the strong emphasis on SLRs rather than EBSE. The practical assignment consisted of conducting the steps of an SLR and, in fact, the tests did not evaluate any of the EBSE LOs (LO05 to LO10). Consequently, our students may be better prepared to aggregate studies and obtain evidence than to identify problems that can be addressed with EBSE and use already aggregated evidence. We discuss this issue further in Section 10.

10 Conclusions and Future work

The work reported in this article includes the following novel aspects: a systematic review of previous EBSE training initiatives, a LOs proposal for EBSE teaching, the development of a course based on those LOs that incorporates a large practical content related to undertaking an SLR, and the course's assessment taking into account the students and teachers' perspectives. The LOs include a guide as to which EBSE concepts and skills are needed to train future users and, although debatable, they form a basis for future research initiatives.

The evidence collected from the students' opinions, the learning assessment, and teachers' opinion suggests the LOs and the teaching methodology enabled students to understand EBSE and to apply it through the execution of the steps of the SLR process. The evidence also shows that, in agreement with the majority of the previous research, a teaching approach with a strong practical workload gives good results. We used an iterative teaching method with each theoretical class being followed by a practice session using the introduced topic. It was used only once among the primary studies in our SLR (Lavallée et al., 2014), but our results confirm it to be both beneficial for, and liked by, students.

Future work should consider how to improve EBSE teaching at university level. For example, students found data synthesis difficult (this arose both in perception of achievement and difficulty by students, and in the teachers' evaluation), which implies that training on this topic needs to be improved. Improved training could include teaching students how to tabulate their results before trying thematic analysis. It is often not clear whether thematic analysis is viable without a good overview of the primary studies. Additionally, in order to improve knowledge translation learning, which was another topic found difficult by students, learning outcomes related to more practical aspects should be included. As a concrete way to approach that, students could be asked to elaborate the results of their practical assignment through a one-page summary as suggested by Cartaxo et al. (2016) and Budgen et al. (2020).

Our course needs a greater emphasis on EBSE and not so much on the SLR process. This would improve both the learning of topics such as use of evidence and knowledge translation, and the students' perception of the usefulness of evidence-based practice. As a way to achieve this it could be interesting:

1 To include the paper by Kasoju et al. (2013) in the reading list for the EBSE training.
2 To test knowledge of EBSE (i.e. LO5 to LO10) in one (or both) of the written examinations.
3 To consider both of the two final EBSE steps at the end of the course covering issues such as what types of contextual information affect the use of knowledge (i.e. company size, experience of staff, type of applications etc.) and reflections on how the EBSE process worked, and what this means in practice.
4 To set up an assignment based on some scenario such as starting up a new company and deciding whether to use test-driven development or conventional testing, then asking students to find one or more SLRs on the topic and identify what decision they might make and why.

However, some issues require more EBSE research-based in an industry context. For example, few students believed the course prepared them better for their professional practice, and many students found knowledge translation difficult to understand. Without input from industrial case studies or reports from industry practitioners, it is difficult to improve training on these topics.

Future courses should be assessed in order to obtain perspectives from a wider range of students and teachers. If some of the improvements presented in the previous paragraphs were carried out, detailed evaluations should be added to determine their impact.

Another future line of work is EBSE training in the industry, something that was not been addressed in any of our SLR primary studies. Although adaptations of the content, methods, and materials of our course could be used, it is also necessary to include more emphasis on EBSE and the use of evidence by practitioners, as we discuss above.

11 Declarations

– Funding: All authors apart from Kitchenham are employees of the Universidad de la República funded by the government of Uruguay. Kitchenham is an unpaid Emeritus Professor and received no funding.
– Conflicts of interest/Competing interests: The authors declare that they have no conflict of interest.
– Availability of data and material: Not applicable.
– Code availability: Not applicable.
– Authors' contributions: Author's contributions are explained in different sections of the paper to discuss possible research bias. Pizard wrote the first draft of the paper and thereafter Kitchenham and Vallespir assisted with restructuring and refining the structure and contents of the original draft and its subsequent revision. All authors reviewed the original draft and the final revision.

Acknowledgements We thank the reviewers for their helpful comments, which have allowed us to substantially improve our manuscript.

References

Adam S (2004) Using learning outcomes: A consideration of the nature, role, application and implications for european education of employing learning outcomes at the local, national and international levels. Conference background report on United Kingdom Bologna seminar. Tech. rep., University of Westminster
Adelman C (2015) To imagine a verb: The language and syntax of learning outcomes statements about the author. Tech. rep., National Institute for Learning Outcomes Assessment
Aglen B (2016) Pedagogical strategies to teach bachelor students evidence-based practice: A systematic review. Nurse Education Today 36:255–263
Albarqouni L, Hoffmann T, Straus S, Olsen NR, Young T, Ilic D, Shaneyfelt T, Haynes RB, Guyatt G, Glasziou P (2018) Core competencies in evidence-based practice for health professionals. JAMA Network Open 1(2)
Badampudi D, Wohlin C, Gorschek T (2019a) Contextualizing research evidence through knowledge translation in software engineering. In: International Conference on Evaluation and Assessment in Software Engineering, pp 306–311
Badampudi D, Wohlin C, Gorschek T (2019b) An evaluation of knowledge translation in software engineering. In: International Symposium on Empirical Software Engineering and Measurement, pp 1–6
Baldassarre M, Boffoli N, Caivano D, Visaggio G (2008) A hands-on approach for teaching systematic review. In: International Conference on Product-Focused Software Process Improvement, pp 415–426
Barkley EF, Major CH (2016) Learning Assessment Techniques: A Handbook for College Faculty. John Wiley & Sons
Basili V, Briand L, Bianculli D, Nejati S, Pastore F, Sabetzadeh M (2018) Software engineering research and industry: A symbiotic relationship to foster impact. IEEE Software 35(5):44–49

Beall J (2012) Criteria for determining predatory open-access publishers by Jeffrey Beall. URL https://scholarlyoa.files.wordpress.com/2012/11/criteria-2012-2.pdf, (last accessed on Dec 2019)

Beall J (2013) Potential, possible, or probable predatory scholarly open-access publishers. URL https://web.archive.org/web/20170112125427/https://scholarlyoa.com/publishers/, (last accessed on Dec 2019)

Beidas RS, Kendall PC (2010) Training therapists in evidence-based practice: A critical review of studies from a systems-contextual perspective. Clinical Psychology: Science and Practice 17(1):1–30

Bernard B, França ND, Ribeiro TV, Sérgio P (2015) Using focus group in software engineering: lessons learned on characterizing software technologies in academia and industry. In: Ibero-American Conference on Software Engineering, pp 351–364

Biggs JB (2011) Teaching for quality learning at university: What the student does. McGraw-hill education (UK)

Bloom B (1956) Taxonomy of educational objectives, the classification of educational Goals, handbook I: Cognitive domain. D. McKay

Boggs GR (1999) What the learning paradigm means for faculty. American Association for Higher Education Bulletin 51

Brereton P (2011) A study of computing undergraduates undertaking a systematic literature review. IEEE Transactions on Education 54(4):558–563

Brereton P, Turner M, Kaur R (2009) Pair programming as a teaching tool: a student review of empirical studies. In: Conference on Software Engineering Education and Training, pp 240–247

Brettle A, Raynor M (2013) Developing information literacy skills in pre-registration nurses: An experimental study of teaching methods. Nurse Education Today 33(2):103–109

Britto R, Usman M (2015) Bloom's taxonomy in software engineering education: A systematic mapping study. In: IEEE Frontiers in Education Conference, pp 1–8

Budgen D, Brereton P, Williams N, Drummond S (2020) What support do systematic reviews provide for evidence-informed teaching about software engineering practice? E-Informatica Software Engineering Journal 14(1):7–60

Cartaxo B, Pinto G, Vieira E, Soares S (2016) Evidence briefings: Towards a medium to transfer knowledge from systematic reviews to practitioners. In: International Symposium on Empirical Software Engineering and Measurement, pp 57:1–57:10

Cartaxo B, Pinto G, Soares S (2018) The role of rapid reviews in supporting decision-making in software engineering practice. In: International Conference on Evaluation and Assessment in Software Engineering, pp 24–34

Carver J, Hassler E, Hernandes E, Kraft N (2013) Identifying barriers to the systematic literature review process. In: International Symposium on Empirical Software Engineering and Measurement, pp 203–212

Castelluccia D, Visaggio G (2013) Teaching evidence-based software engineering: Learning by a collaborative mapping study of open source software. ACM SIGSOFT Software Engineering Notes 38(6):1–4

Catal C (2013) Teaching evidence-based software engineering to master students: A single lecture within a course or an entire semester-long course? ACM SIGSOFT Software Engineering Notes 38(2):1–2

Cruzes D, Dybå T (2011) Recommended steps for thematic synthesis in software engineering. In: International Symposium on Empirical Software Engineering and Measurement, pp 275–284

Da Silva F, Santos A, Soares S, Frana A, Monteiro C, Maciel F (2011) Six years of systematic literature reviews in software engineering: An updated tertiary study. Information and Software Technology 53(9):899–913

DeFranco JF, Laplante PA (2017) A content analysis process for qualitative software engineering research. Innovations in Systems and Software Engineering 13(2-3):129–141

Dybå T, Dingsøyr T (2008) Empirical studies of agile software development: A systematic review. Information and Software Technology 50(9-10):833–859

Elo S, Kyngäs H (2008) The qualitative content analysis process. Journal of Advanced Nursing 62(1):107–115

Emerson RM, Fretz RI, Shaw LL (2001) Participant Observation and Fieldnotes, SAGE Publications Ltd, chap 24, pp 352–368

Feinstein AR, Cicchetti DV (1990) High agreement but low kappa: I. the problems of two paradoxes. Journal of Clinical Epidemiology 43(6):543 – 549

Garousi V, Felderer M (2017) Experience-based guidelines for effective and efficient data extraction in systematic reviews in software engineering. In: International Conference on Evaluation and Assessment in Software Engineering, pp 170–179

Hassler E, Carver J, Kraft N, Hale D (2014) Outcomes of a community workshop to identify and rank barriers to the systematic literature review process. In: International Conference on Evaluation and Assessment in Software Engineering, pp 31:1–31:10

Havnes A, Prøitz TS (2016) Why use learning outcomes in higher education? exploring the grounds for academic resistance and reclaiming the value of unexpected learning. Educational Assessment, Evaluation and Accountability 28(3):205–223

Janzen D, Ryoo J (2009) Engaging the net generation with evidence-based software engineering through a community-driven web database. Journal of Systems and Software 82(4):563–570

Janzen DS, Ryoo J (2008) Seeds of evidence: Integrating evidence-based software engineering. In: Conference on Software Engineering Education and Training, pp 223–230

Joint Task Force on Computing Curricula - ACM and IEEE Computer Society (2013) Computer Science Curricula 2013: Curriculum Guidelines for Undergraduate Degree Programs in Computer Science. Association for Computing Machinery, New York, NY, USA

Joint Task Force on Computing Curricula - ACM and IEEE Computer Society (2014) Software engineering 2014: Curriculum Guidelines for Undergraduate Degree Programs in Software Engineering. Association for Computing Machinery, New York, NY, USA

Jørgensen M, Dybå T, Kitchenham B (2005) Teaching evidence-based software engineering to university students. In: International Software Metrics Symposium, pp 24:1–24:8

Juristo N (2007) Empirical Software Engineering Issues. Critical Assessment and Future Directions: International Workshop, Dagstuhl Castle, Germany, June 26-30, 2006. Revised Papers, Springer Berlin Heidelberg, chap Educational objectives for empirical methods, pp 143–150

Kasoju A, Petersen K, Mäntylä M (2013) Analyzing an automotive testing process with evidence-based software engineering. Information and Software Technology 55(7):1237–1259

Kasunic M (2005) Designing an effective survey. Tech. Rep. CMU/SEI-2005-HB-004, Carnegie Mellon University Software Engineering Institute

Kennedy D, Hyl Á, Ryan N, Hyland Á, Ryan N (2007) Writing and using learning outcomes: A practical guide, European University Association, chap C 3.4-1, pp 1–30

Kitchenham B, Brereton P (2013) A systematic review of systematic review process research in software engineering. Information and Software Technology 55(12):2049–2075

Kitchenham B, Charters S (2007) Guidelines for performing systematic literature reviews in software engineering. Tech. Rep. EBSE-2007-01, Keele University

Kitchenham B, Dybå T, Jørgensen M (2004) Evidence-based software engineering. In: International Conference On Software Engineering, pp 273–281

Kitchenham B, Brereton P, Budgen D (2010) The educational value of mapping studies of software engineering literature. In: International Conference on Software Engineering, pp 589–598

Kitchenham B, Budgen D, Brereton P (2015) Evidence-based software engineering and systematic reviews. Chapman & Hall/CRC Innovations in Software Engineering and Software Development Series, CRC Press

Ko AJ (2017) A three-year participant observation of software startup software evolution. In: International Conference on Software Engineering, pp 3–12

Kontio J, Bragge J, Lehtola L (2008) The focus group method as an empirical tool in software engineering, Springer London, London, chap 4, pp 93–116

Lavallée M, Robillard P, Mirsalari R (2014) Performing systematic literature reviews with novices: An iterative approach. IEEE Transactions on Education 57(3):175–181

Likert R (1932) A technique for the measurement of attitudes. Archives of Psychology 22(140):1–55

Miles MB, Huberman AM, Saldana J (2014) Qualitative Data Analysis: A Methods Sourcebook, 3rd edn. Sage Publications

Mulder AE (2019) 13 ideas that are transforming the community college world. Routledge

Oates B, Capper G (2009) Using systematic reviews and evidence-based software engineering with masters students. In: International Conference on Evaluation and Assessment in Software Engineering, pp 79–87

Rainer A, Beecham S (2008) A follow-up empirical evaluation of evidence based software engineering by undergraduate students. In: International Conference on Evaluation and Assessment in Software Engineering, pp 78–87

Rainer A, Hall T, Baddoo N (2006) A preliminary empirical investigation of the use of evidence based software engineering by undergraduate students. In: International Conference on Evaluation and Assessment in Software Engineering, pp 91–100

Ribeiro T, Massollar J, Travassos G (2018) Challenges and pitfalls on surveying evidence in the software engineering technical literature: An exploratory study with novices. Empirical Software Engineering 23(3):1594–1663

Rice W (2006) Moodle E-Learning Course Development. Packt Publishing

Rosnow R, Rosenthal R (1997) People studying people: artifacts and ethics in behavioral research. WH Freeman

Runeson P, Höst M (2009) Guidelines for conducting and reporting case study research in software engineering. Empirical Software Engineering 14(2):131–164

Sadeghi-Bazargani H, Tabrizi JS, Azami-Aghdash S (2014) Barriers to evidence-based medicine: A systematic review. Journal of Evaluation in Clinical Practice 20(6):793–802

Santos R, Da Silva F (2013) Motivation to perform systematic reviews and their impact on software engineering practice. In: International Symposium on Empirical Software Engineering and Measurement, pp 292–295

Saunders H, Gallagher-Ford L, Kvist T, Vehviläinen-Julkunen K (2019) Practicing healthcare professionals' evidence-based practice competencies: An overview of systematic reviews. Worldviews on Evidence-Based Nursing 16(3):176–185

Schoepp K (2019) The state of course learning outcomes at leading universities. Studies in Higher Education 44(4):615–627

Scurlock-Evans L, Upton D (2015) The role and nature of evidence: A systematic review of social workers' evidence-based practice orientation, attitudes, and implementation. Journal of Evidence-Informed Social Work 12(4):369–399

Smith E, Loftin R, Murphy-Hill E, Bird C, Zimmermann T (2013) Improving developer participation rates in surveys. In: International Workshop on Cooperative and Human Aspects of Software Engineering, pp 89–92

Stanny C (2016) Reevaluating bloom's taxonomy: what measurable verbs can and cannot say about student learning. Education Sciences 6(4):37:1–37:12

Stol KJ, Fitzgerald B (2018) The ABC of Software Engineering Research. ACM Transactions on Software Engineering and Methodology 27(3):1–51

Torchiano M, Fernández D, Travassos G, De Mello R (2017) Lessons learnt in conducting survey research. In: International Workshop on Conducting Empirical Studies in Industry, pp 33–39

Turner M, Kaur R, Brereton P (2008) A lightweight systematic literature review of studies about the use of pair programming to teach introductory programming. In: Annual Meeting of the Pschology of Programming Interest Group, pp 21:1–21:15

Upton D, Stephens D, Williams B, Scurlock-Evans L (2014) Occupational therapists' attitudes, knowledge, and implementation of evidence-based practice: A systematic review of published research. British Journal of Occupational Therapy 77(1):24–38

Vachon B, Durand MJ, Leblanc J (2010) Empowering occupational therapists to become evidence-based work rehabilitation practitioners. Work 37(2):119–134

Venning J, Buisman-Pijlman F (2013) Integrating assessment matrices in feedback loops to promote research skill development in postgraduate research projects. Assessment and Evaluation in Higher Education 38(5):567–579

Viera AJ, Garrett JM (2005) Understanding interobserver agreement: the kappa statistic. Family Medicine 37(5):360–3

Wohlin C (2014) Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: International Conference on Evaluation and Assessment in Software Engineering, pp 1–10

Zwolsman S, Te Pas E, Hooft L, Wieringa-De Waard M, Van Dijk N (2012) Barriers to gps' use of evidence-based medicine: A systematic review. British Journal of General Practice 62(600):511–521

I Complementary Information on the SLR

For the purpose of increasing traceability and reproducibility, this section includes additional information on the systematic literature review on EBSE and SLR training presented in Section 2.

A Search Strings by Search engines

Table 11 shows the search strings used in the different search engines, which were adapted from the original string presented in Table 1.

Search Engine	Search String
SCOPUS	TITLE-ABS-KEY((teach OR learn OR education OR train OR students) AND ("evidence-based software engineering" OR "evidence based" OR ebse OR "systematic literature review" OR "systematic review" OR "literature review" OR slr OR "systematic mapping" OR "mapping study" OR "scoping study" OR SMS) AND ("software engineering"))
ACM DL	(acmdlTitle:(teach learn education train students) AND acmdlTitle:("evidence-based software engineering" "evidence based" ebse "systematic literature review" "systematic review" "literature review" slr "systematic mapping" "mapping study" "scoping study" SMS) AND acmdlTitle:("software engineering")) OR (recordAbstract:(teach learn education train students) AND recordAbstract:("evidence-based software engineering" "evidence based" ebse "systematic literature review" "systematic review" "literature review" slr "systematic mapping" "mapping study" "scoping study" SMS) AND recordAbstract:("software engineering"))
IEEExplore	((teach OR learn OR education OR train OR students) AND ("evidence-based software engineering" OR "evidence based" OR ebse OR "systematic literature review" OR "systematic review" OR "literature review" OR slr OR "systematic mapping" OR "mapping study" OR "scoping study" OR SMS) AND ("software engineering"))

Table 11 Adapted search strings

B Papers Obtained by each Search

Table 12 presents the studies obtained from each search carried out. The identifiers of the studies are those previously presented in Table 3.

Search	Total	Papers selected
2017	10	S3, S4, S5, S6, S7, S9(2009), S10(2008, 2009), S11, S14
2018	0	-
2019	2	S8, S12
Snowballing & search by authors	4	S1, S2, S9(2008), S13

Table 12 Papers by search

C Categorization Scheme and Quality Assessment

The categorization scheme included:

- Main motivation: EBSE/SLR process issues (e.g. analysis of EBSE execution -reproducibility, effort required, etc.- or proposals for new variants to the EBSE process) / teaching EBSE/SLR (e.g. EBSE teaching proposals and their results) / attitudes to EBSE/SLR (e.g. research on whether practitioners perceive EBSE useful or what stages they find most challenging to execute)
- Summary of aims of the study
- Number of student participants
- Student type: Undergraduate / MSc / PhD / Under and postgraduate / Not stated
- Program area: Computer Science / Another field (not CS) / CS and another field / Not stated
- Course focus: Integrated modules (i.e. modules that cover a variety of topics) / Empirical SE / EBSE or SLR / SE / Research methods / Individual projects (i.e. individual work of medium and broad-scope) / Software architecture / Experimental SE
- Scope of the study (i.e. type of student practical assignment): SLR limited / SLR / Mapping Study / Other scope / Not stated
- Educational methodology: Brief introduction (1 to 3 classes) plus practical assignment / Longer lessons plus practical assignment / Alternating introduction of concepts and practice / Not stated
- Type of lessons: Lectures / Lectures and tutorials / Tutorials / Not stated
- Type of training
  - Number of classroom hours
  - Number of extra hours required of participants
  - Proportion of total training time dedicated to practical work
  - Elapsed time
  - Participation criteria: Mandatory / Optional / Not stated
- Evaluation process used
  - Written Tests: Yes/No
  - Teacher evaluation of EBSE or SLR outcomes: Yes/No
  - Student questionnaire: Yes / No
  - Student reports (i.e. reports that describe the experience of students during their participation in the practical assignment of the course): Individual / Team / Individual and Team / No
  - Not stated: Yes / No
- EBSE/SLR training problems and difficulties
- EBSR/SLR training benefits
- Study limitations

We extracted the data independently using an extraction form, created in Google spreadsheets, and tested previously with some articles. In a subsequent meeting, we reached an agreement for each item of data. Each conflict was discussed and an agreement was reached.

Textual data was extracted by Pizard. To validate the extraction Moreno and Pizard performed a lean peer review as recommended by Garousi and Felderer (2017). This type of review involves selecting a random set of papers and reviewing them interactively by asking questions, while the other researcher explains the extraction. We reviewed half of the papers randomly using this method.

As the primary studies were of different types, for the quality assessment we used the same questions as Kitchenham and Brereton (2013). This set of generic questions, originally used by Dybå and Dingsøyr (2008), can be applied to different types of studies. Pizard and Moreno extracted the quality data of each primary study independently. In a meeting, the disagreements were resolved. Quality extraction was done in parallel to data extraction. The set of questions was: (questions 3 through 12 admit the following answers: Yes / Partly / No / Not applicable. Score as 1, 0.5, 0. Interpolation is permitted for numerical values).

1 Is the paper based on research (or is it a discussion paper based on expert opinion)? Yes / No.
2 What research method was used: Experiment, Quasi-Experiment, Lessons learned, Case study, Opinion Survey, Other (specify)? Note: This is to be based on paper reading, not the method claimed by the authors.

3 Is there a clear statement of the aims of the study?
4 Is there an adequate description of the context in which the research or observation was carried out?
5 Was the research method appropriate to address the aims of the research?
6 Was the recruitment strategy (for human-based experiments and quasi-experiments) or experimental material or context (for Lessons learned) appropriate to the aims of the research?
7 For empirical studies (apart from Lessons learned), was there a control group or baseline with which to evaluate SLR procedures/techniques?
8 For empirical studies (apart from Lessons learned), was the data collected in a way that addressed the research issue?
9 For empirical studies (apart from Lessons learned), was the data analysis sufficiently rigorous?
10 Has the relationship between researcher and participants been considered to an adequate degree?
11 Is there a clear statement of findings?
12 Is the study of value for research or practice?

To study the reliability of the initial agreement in the quality assessment, and again in a similar way to the study of Kitchenham and Brereton (2013), Pizard calculated the Kappa coefficient for Question 2 and the Pearson correlation coefficient between the values for each reviewer both for the number of relevant questions and for the average quality score for each study.

D Reliability of Data Extraction and Quality Assessment

The extraction agreement with respect to the categories assigned by each author was evaluated using Kappa statistic (see Table 13).

Disagreement on Educational methodology was due to the fact that reviewers had different criteria during the individual extraction. This happened only for papers that reported courses of a different focus than EBSE, for example, Empirical Software Engineering. In these cases, for example, if a paper reported many classes but only one on EBSE, one author classified it as a 'brief introduction' while the other as 'longer lessons'. At the meeting, reviewers agreed to use only the information on EBSE teaching to classify the studies.

The zero values of Kappa in the Written tests and Not stated categories of Process evaluation are due to the fact that the Kappa is affected by the prevalence of the findings under consideration and strongly depends on the marginal distributions (Viera and Garrett, 2005; Feinstein and Cicchetti, 1990). In both cases, the number of observed agreements and the number of agreements expected by chance coincide in 13 of 14 classified studies.

During the final agreement meeting the following categories were added: 'seminars' for type of lessons (to classify study S10); 'postgraduate' for student type (to classify studies S1 and S2), and 'EBSE steps' for scope of the study (to classify studies S12, S13, and S14).

Regarding quality assessment, the initial agreement for question 2 about the type of study was 11 out of 14 studies with a Kappa coefficient of 0.659. The major disagreements were due to the fact that one author classified two studies as case studies when they should have been classified as opinion surveys using the Kitchenham and Brereton criteria (they correspond to case studies based only on opinion surveys).

The Pearson correlation between the number of questions each of us believed to be relevant was 0.73 with p=.003. We believe this level of disagreement in the number of questions is related to the level of disagreement in the classification of article types. In many cases, we considered the type of study when identifying the relevant questions. Reliability was better for the average scores for each study, where the correlation was 0.96 with $p<0.00001$. Both were statistically significant ($p < 0.05$).

II Learning Objectives Theory and Practice

In 1991, the Learning paradigm emerged in California (Mulder, 2019). It was a shift from identification with processes to identification with results or outcomes. Under this approach, edu-

Data extracted	Categories	Agreement (out of 14 assessed)	Kappa
Main motivation	EBSE or SLR process issues / teaching EBSE or SLR / attitudes to EBSE or SLR	11	0.650
Student type	Undergraduate / MSc / PhD / Under and posgraduate / Not stated	13	0.890
Program field	Computer Science / Another field (not CS) / CS and another field / Not stated	12	0.810
Course focus	Integrated modules / Empirical SE / EBSE or SLR / SE / Research methods / Individual projects / Software architecture / Experimental SE	11	0.736
Scope of the study (i.e. type of practical assignment)	SLR limited / SLR / Mapping Study / Other scope / Not stated	11	0.722
Educational methodology	Brief introduction (1 to 3 classes) plus practical assignment / Longer lessons plus practical assignment / Alternating introduction of concepts and practice / Not stated	7	0.246
Type of lessons	Lectures / Lectures and tutorials / Tutorials / Not stated	12	0.774
Evaluation process used - Written Tests	Yes / No	13	0.000
Evaluation process used - Teacher evaluation of EBSE or SLR outcomes	Yes / No	11	0.588
Evaluation process used - Student questionnaire	Yes / No	13	0.859
Evaluation process used - Student reports	Individual / Team / Individual and Team / No	11	0.700
Evaluation process used - Not stated	Yes / No	13	0.000

Table 13 Initial agreement in the categorization of studies

cational institutions must focus their mission on student learning instead of teaching. Learning outcomes represent, or maybe catalyze, the learning paradigm (Schoepp, 2019). They can be used to evaluate the effectiveness of the teaching initiatives instead of measuring the resources or processes (Boggs, 1999).

LOs are statements that express what students are expected to know, understand, and/or be able to demonstrate at the end of the learning period (Kennedy et al., 2007). An example of LO for software design is: *'describe a form of refactoring and discuss when it may be applicable'* (Joint Task Force on Computing Curricula - ACM and IEEE Computer Society, 2013). LOs can be seen as basic educational building blocks because of their impact on other educative tools (Adam, 2004). They can be used to identify learning achievements but, if well designed, can also encourage alignment between learning, teaching, or educational activities and evaluation (Biggs, 2011). Its use motivates curricula development with more content-based practices since to use LOs it is necessary to specify the expected results before designing a course.

LOs' adoption has received strong international support and its defenders find it has several advantages (Kennedy et al., 2007). Furthermore, some authors consider its adoption as an international de facto standard (Schoepp, 2019). By focusing on the student, this approach promotes the idea of teachers as facilitators of the learning process and also recognizes that much of it occurs outside the classroom (Adam, 2004). The use of LOs helps teachers to: communicate to students precisely what is expected of them, design materials more effectively, select the most appropriate learning strategies for each objective, and help to develop assessments based on delivered materials (Kennedy et al., 2007). Students often have less anxiety because they have a clear direction, they know the priorities of their instruction, and they can perceive that the grading process is fair (Barkley and Major, 2016). The adoption of LOs also contributes to objectives' transparency and their compatibility with standards, the consistency between courses and educational programs, and the mobility of students between educational institutions by facilitating the recognition of their qualifications (Adam, 2004).

In philosophical terms, the main objection to the adoption of LOs may be that they do not facilitate an open-ended approach to academic study (Adam, 2004). Another risk is the oversimplification of the learning process (Havnes and Prøitz, 2016). This happens, for example, by simplifying the concepts to model programs with LOs too quickly or by carrying out the LO writing process too mechanically. Many authors agree that the adoption of this approach takes considerable effort and time.

One of the main success factors in LOs' adoption is their correct written specification (Kennedy et al., 2007). To achieve this, there are various guides and recommendations. All of them agree on the importance of the verb used in each LO. In this regard, Adelman (2015) says *'the verb is the center, fulcrum, engine of a learning outcome statement'*. Different studies have been carried out to propose and evaluate verbs to be used. Most of the initiatives are based on Bloom's taxonomy as it provides a structure and a list of verbs. It is also recommended that LOs have a single verb and are simple and concrete, observable, and measurable (Kennedy et al., 2007).

Each LO can be associated with one of Bloom's levels of cognitive domain (Bloom, 1956). The cognitive domain has six levels whose description is as follows:

1 *Knowledge.* The student can remember or recognize information, concepts, and ideas on a subject.
2 *Comprehension.* The student can comprehend, interpret, organize, and relate the general idea of a topic.
3 *Application.* The student can use what they have learned to solve a new problem or situation.
4 *Analysis.* The student can examine information on a topic, identify causes and infer in order to substantiate generalizations.
5 *Synthesis.* The student can find new patterns or combine information to create new proposals.
6 *Evaluation.* The student can evaluate and validate ideas and make an assessment on a topic.

An appropriate design of LOs together with their categorization according to Bloom's levels allow teachers to better select content, teaching methodology, teaching resources, and assessment tools for their courses. In particular, they are very useful for guiding the design of proposals focused on learning, which aim to make student learning more effective (Kennedy et al., 2007).

Recently, several authors have promoted the use of LOs for the design and teaching of courses related to software engineering (Joint Task Force on Computing Curricula - ACM and IEEE Computer Society, 2013; Britto and Usman, 2015) and to empirical software engineering (Juristo, 2007). In practice, between 2000 and 2014, 26 studies were conducted on the application of Bloom's taxonomy in areas of software engineering education (Britto and Usman, 2015). None of them reported applications related to EBSE or SLR training.

III Student Opinion Survey Form on the EBSE Course

This section presents the survey form used to collect students opinions and described in section 6.1.

A Regarding your Experience and Previous Knowledge

1 If you have a job, indicate the role you occupy:
2 Indicate your level of experience in the following areas (1 = None, 5 = Expert)
   – Software Engineering Area
      – Requirements Engineering
      – Software Design

    - Software Construction
    - Software Testing
    - Software Maintenance
    - Configuration Management
    - Project Management
    - Software Process
    - Software Quality
- Other areas
    - On the topic of your SLR
    - Reading Comprehension in English
    - On scientific articles and, in particular, on the area of software engineering or on the topic of your SLR (primary studies)
    - About Software Engineering Based on Evidence or Systematic Reviews of Literature (secondary studies)

B About the Course

Indicate your level of agreement with the following statements about the course and about teamwork (1 = Strongly disagree, 5 = Strongly agree).

1 The spaces for consultations (face-to-face and/or virtual) are useful.
2 Recommended study materials are useful.
3 It is possible to access the recommended study materials.
4 The course's Moodle website is useful.
5 There is coordination between theoretical and practical classes.
6 Attending class favors the understanding of course topics.
7 The evaluation criteria of the subject were clearly explained.
8 The evaluation process (mid-terms, submissions, etc.) could be carried out with the knowledge developed during the course.
9 The evaluation proposals made were clear and unambiguous.
10 The evaluation process focused on the understanding of the subject.
11 Overall opinion on the course.
12 The techniques provided to develop the practices are clear and unambiguous.
13 There is integration between the theoretical classes and teamwork.
14 Teamwork allows us to integrate knowledge of different subjects.
15 The topics developed in teamwork are linked to the professional life of an engineer.
16 The time available for the performance of each practice is adequate.
17 The evaluation process (submissions, presentations, mid-terms, etc.) could be carried out with the knowledge developed during the course.
18 Overall opinion on teamwork.

C Learning Outcomes of the Course

Below are the learning outcomes (LOs) proposed by the teachers for the course. They are organized according to each thematic unit worked. We ask you to indicate the level of achievement that you believe you had in each LO during the course. To complete this section you must work individually and you can consult the course materials.

The scale corresponds to: 1. Not achieved at all — 2. Very little achieved — 3. Successfully achieved — 4. Almost completely achieved — 5. Completely achieved

- *In the original form, all the learning outcomes presented in Table 9 are listed here.*

  Below is a space available for comments on learning outcomes not achieved.

D Regarding Teamwork and the Benefits of the Course

1 Indicates the level of difficulty of the different stages of an SLR (1 = very easy, 5 = very difficult).

  - SLR planning
  - Search for primary studies
  - Study Selection
  - Study quality evaluation
  - Data extraction from studies
  - Mapping studies analysis
  - Qualitative synthesis
  - Report of a systematic review
  - Overall difficulty of the entire SLR

2 Indicate according to your criteria what are the main benefits of the course.

E Questions to Answer as a Team

1 Indicate how it influenced your work to have chosen your subject (motivation, difficulty, etc.).
2 Indicate how it influenced your work that the bibliography was written in English (1. Very negatively — 2. It was something that caused some difficulty — 3. It did not influence — 4. It was something that caused some benefit — 5. Very positively).
3 Indicate which were the two biggest challenges or difficulties you had to face during teamwork and how you overcome them.
4 Indicate what comments you can make about the reviewed primary studies (for example, about quality, completeness, complexity, terminology, etc.).
5 Indicate whether the results of its SLR seem useful, if not explain why.

IV Case Study Additional Data

This section includes additional data collected in the case study described in this paper.

A Student Opinion Survey - Individual Sections

Part B. General questions about the course

- Figure 7 presents the results of the survey on students' opinions about the course and teamwork.

Part C - 55. Comments on LOs achievement perceptions

- I think the topic of knowledge translation wasn't enough for our understanding.
- I would've liked one of the classes to have included a guest that strongly or partially applies EBSE in their professional practice.
- The objectives of the course are satisfactorily achieved. I believe that with the practice and execution of another SRL we would cement our knowledge and see in which case each concept applies.

Part D - 2. Benefits of the course
A pragmatic vision

- Becoming aware of a systematic method, i.e., one that includes steps and procedures that were reviewed by experts in order to search and synthesize material that can answer questions.
- I gained knowledge about evidence-based engineering, procedures, and scientific papers.
- I learned how a scientific article is composed and how to do an SRL in IS.
- I learned how to perform a complete SRL, became aware of each step, when to apply what type of review, and also which templates or forms to use.

To improve professional practice

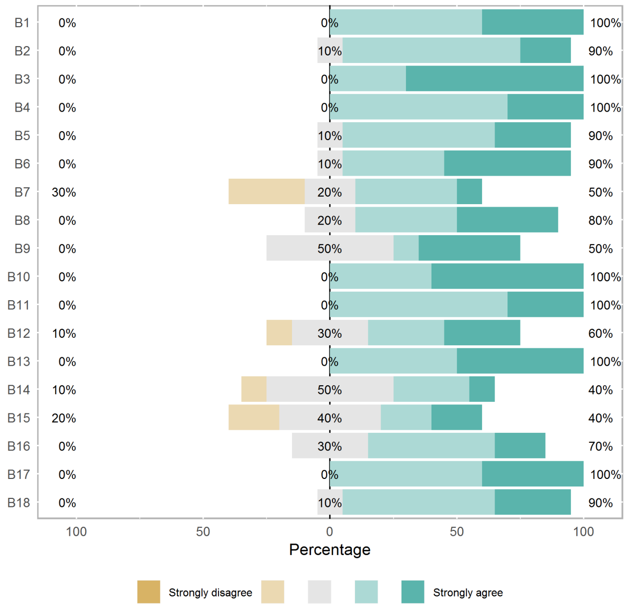

Fig. 7 Opinion survey on general aspects of the course and teamwork.

– I learned a lot about literature in our profession, how to look for it, where to find it, the format it should have. Before the course, I had no knowledge of the area. I believe that the methodology taught can be applied to each report that I have to deliver to someone else.
– I learned a new tool that can be used to compare practices, techniques, and processes in software engineering.
– I gained the ability to quickly assess and absorb studies of all kinds, search and find scientific data that I am already using in other courses and at work.

To improve research skills

– The main contribution is for academic training. It's a great help for doing the capstone projects required to complete our degree. It is eventually beneficial for a change in technology in professional practice, though I believe most local companies in the industry don't consider this methodology important (yet).
– Mainly, I believe it helps to carry out an investigation and to be based on scientific and truthful evidence. It will be useful for the capstone project.
– Now I have the necessary tools to carry out research on a topic related to software engineering

B Student Opinion Survey - Team Section (E)

Question 1 - SLR topic selection by students

– It reduces the difficulty since a familiar theme is selected and/or known by the members. It is motivating since a topic of interest is chosen [by us] and not imposed by others.
– Although it was interesting, we didn't take into account the difficulty and complexity of the subject. However, the choice of topic provides motivation. It would be good if [the teachers] could warn about the complexity and difficulty.
– It had a great influence since, in the professional practice, [the subject of our project] is being used in a great way, thus strengthening the knowledge that already exists in this regard.
– We were motivated because it is a subject that we could see at work. On the difficulty of not having too much experience on the subject, it was a limitation.

Question 1 - SLR topic selection by students

– Regarding the bibliography being in English, two teams indicated that it was something that caused them some difficulty (2) and two teams indicated that it had no influence (3).

Question 3 - Biggest challenges during teamwork

- 1) The selection stage was difficult as we had some difficulties in reaching the number of items to select. Some [papers] were discarded in advanced stages that led to redoing previous steps. 2) Data extraction and quality assessment. Both due to our lack of practice in applying these methods.
- 1) The complexity of the subject. 2) Performing the synthesis was one of the most complex tasks.
- 1) The teachers seemed not to agree on the suggestions they made. On some occasions, it was necessary to carry out a rework due to their different opinions. This was even noticed in the evaluation of the SLR, whereby it is concluded that the problem was not overcome. 2) [We found it difficult to] reach conclusions about the synthesis carried out, data crossing, how it affected quality. There were a few examples, it was necessary to read other reviews and discuss possible options in the team.
- 1) Synthesis was the most complex [part] since the primary studies were very different, we solved it by doing the whole synthesis together. 2) Many of the studies selected by title and abstract did not pass the second stage (full text) and we had to process new studies.

Question 4 - Comments on reviewed primary studies

- We had expected to obtain articles on tools that are better known to us and in none of the cases did the analyzed tools coincide. We would have expected more studies on tools.
- Some [papers] required expert knowledge in certain areas.
- One problem encountered was [the difficulty of] being able to understand/find the proposed methodologies and algorithms in the articles, since they were not explicitly mentioned.
- Many [papers] didn't comply with basic quality aspects, didn't evaluate the models presented (metrics) and it happened to us that we discarded some due to their complexity.

Question 5 - Usefulness of SLR results

- We believe that yes, it provided us with the knowledge that some members have already used in practice.
- We believe that useful recommendations were obtained in the SLR, but it should be taken into account that few articles were analyzed.
- The SLR results allowed us to have a general idea about the topics related to the chosen topics, and to see what is on the market and in what context they are applied.
- Due to time constraints, we processed 10 articles (out of 1700 found) this means that it is not very valid outside the context of the course.

C Focus Groups Teachers' Notes

After filling out the course survey, students were asked to discuss among themselves what improvements or changes could be made to the course. The following ideas emerged from one or more students:

1. More students per team in order to be able to process more articles.
2. Selection of topics for practical work by teachers in order to minimize the number of articles and be able to follow the stages more easily.
3. The previous discussion continued and students reported that choosing the practical assignment topic was very motivating. But perhaps it would be good to recommend students to choose topics that they already know something about, in order to avoid upcoming difficulties
4. The weekly meetings worked well. The modality of the course seems good to them.
5. Some students indicate that it would be good for teachers to promote the reading of the technical report on the subject in Spanish because they think it was useful, but they read it too late in the course.
6. It's proposed (4 out of 11 students raise their hands when asking for confirmation) to include reading controls and to make it mandatory to read the book chapters before classes. The proponent argues that he read chapters just before the individual assessment and they would have helped him much earlier. This reading control could be done on the Moodle platform before or during the class on the subject to be evaluated.
7. There is a discussion about the use of the book and the students conclude that it was difficult for them to follow and somewhat boring.

8 Teachers are asked to indicate the homework on the Moodle platform earlier and not too near the dates of the next meetings, something that happened a couple of times.
9 Students say it would be nice to have a forum with more participation, although it is not clear what it would be for.
10 The individual test seemed somewhat ambiguous to some of them, perhaps something more concrete would have been better. Apparently they didn't like question 1 very much.
11 They agree with the distribution of the course score, with 50% for practical assignments.
12 Students recommend that teachers post previous tests (in the coming years) or at least indicate the format that the evaluation will have.
13 A team states that they had to re-run previous steps in their assignments due to a disagreement between teachers when they asked us questions on different occasions.
14 Students also recommend improving the dissemination of the call to enroll in the course.
15 They also indicate that they find it better to submit the practical assignments on Moodle.