Información para entender el protocolo y la ejecución de la revisión.	Hay que pararse en el primer semestre de 2017. Los revisores son Silvana Moreno y Sebastián Pizard. Pizard tiene interés en hacer la revisión pues obtendrá información valiosa para crear el curso de EBSE que vos estás cursando ahora. Además, Moreno y Pizard son supervisados por Diego Vallespir y Barbara Kitchenham y también responden dudas sobre la revisión y su ejecución.
1. Contexto	
1.1. Objetivo	Obtener conocimiento de investigación previa relacionada al entrenamiento de EBSE.
1.2. Necesidad	Antes de dar por primera vez un curso de EBSE consideramos importante tener en cuenta iniciativas similares previas. Al momento no hemos encontrado una revisión o mapeo sistemático adecuado.
1.3. Preguntas de investigación	
RQ1	¿Qué iniciativas de enseñanza de EBSE se han reportado?
RQ2	¿En qué contexto (grados/cursos/etc.) es enseñado?
RQ3	¿Qué contenido es enseñado y qué metodología es usada?
RQ4	¿Qué mecanismos de evaluación utilizan esas iniciativas?
RQ5	¿Cuáles son las dificultades encontradas y qué recomendaciones se dan?
RQ6	¿Qué beneficios obtienen los estudiantes?
2. Proceso de Búsqueda	
2.1. Estrategia	Búsqueda automática por título, abstract y keywords. A partir de los artículos identificados como relevantes se agrega: backward y forward snowballing y búsqueda manuales en Google scholar de otras publicaciones de los autores.
2.2. Snowballing	Backward y forward snowballing
2.3. Términos	
Iniciativa de Enseñanza	teach, learn, education, train, students
ESBE/SLRs	evidence-based software engineering, evidence based, systematic literature review, systematic reviews, literature review, slr, systematic mapping, mapping study, sms
Ingeniería de Software	software engineering
2.4. Cadena de Búsqueda	((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"))
2.5. Motores y Cadenas de Búsqueda	
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"))
2.6. Fuentes a considerar	-
2.7. Período a tener en cuenta (justificar)	Sin períodos específicos. Como es un tema relativamente nuevo, es mejor abarcar la mayor cantidad de estudios posibles.

2.8. Procedimientos auxiliares	-
2.9. Evaluación del Proceso de Búsqueda	Se deberían cubrir los artículos encontrados en una búsqueda preliminar y que se listan a continuación: - Jørgensen M., Dybå T., Kitchenham B., Teaching evidence-based software engineering to university students, 2005, International Software Metrics Symposium - Baldassarre M.T., Boffoli N., Caivano D., Visaggio G., A hands-on approach for teaching systematic review, 2008, Lecture Notes in Computer Science - Janzen D.S., Ryoo J., Seeds of Evidence: Integrating Evidence-Based Software Engineering, 2008, Conference on Software Engineering Education and Training (CSEET) - J. C. Carver; E. Hassler; E. Hernandes; N. A. Kraft, Identifying Barriers to the Systematic Literature Review Process, 2013, International Symposium on Empirical Software Engineering and Measurement - Cagatay Catal, Teaching Evidence-based Software Engineering to Master Students: A Single Lecture Within a Course or an Entire Semester-long Course?, 2013, SIGSOFT Softw. Eng. Notes
3. Proceso de Selección de Estudios Primarios	
3.1. Criterios de Inclusión	
I1	Se incluyen artículos que reporten iniciativas de enseñanza de EBSE, SLRs a estudiantes de ciencias de la computación o ingeniería de software.
3.2. Criterios de Exclusión	
E1	No se toman en cuenta descripciones de keynotes, workshops o artículos que no estén en inglés.
E2	El artículo no está en inglés.
E3	El artículo no está disponible en texto completo
3.3. Roles de los revisores	Un revisor ajustará las cadenas a cada motor. Luego ambos revisores leerán títulos y abstracts de los artículos. Definirán aceptaciones, rechazos y una lista de artículos en duda (a leer en forma completa para decidir si serán rechazo o aceptaciones). Luego se hará una reunión para llegar a un acuerdo. Si hay artículos para leer de forma completa se leerán y se hará una segunda reunión para definir.
3.4. Cómo se evaluará el acuerdo entre revisores	Se calculará el coeficiente kappa.
3.5. Cómo se resolverán diferencias	Ante un no acuerdo sobre un artículo en particular se incluirá en la lista de artículos a leer en forma completa. Luego de la lectura completa se aceptará o rechazará cada artículo sin artículos con dudas. Ante un no acuerdo en ese nivel se incluirá el artículo para la siguiente etapa de la revisión.
4. Proceso de Evaluación de la Calidad de los Estudios	
4.1. Se evaluará la calidad de los estudios (justificar)	Si
4.2. Checklist propuesta	Ver en hoja "Evaluación de calidad"
4.3. Cómo se evaluará el acuerdo entre revisores	Lo harán los dos revisores. Algunas preguntas pueden aplicar y otras no, para estas últimas decir explícitamente "N/A".
4.4. Cómo se resolverán diferencias	Para las preguntas numéricas se hará promedio. Diferencias en los N/A serán analizadas utilizando correlación de Pearson y discutidas, en caso de dudas se pondrá N/A. Como la pregunta 2 es medio crítica, se usará Kappa para ver el nivel de acuerdo.
4.5. Cómo se usarán las checklists	Ver en hoja "Evaluación de calidad"
5. Proceso de Extracción de Datos	
5.1. Formulario de extracción	Ver en hoja "Formulario de extracción"

5.2. Estrategia de extracción	Los datos categorizados serán extraídos por ambos revisores y se utilizará Kappa para evaluar el nivel de acuerdo. Los datos de texto libre serán extraídos por Pizard. Luego se utilizará una lean peer review recomendada por Garousi (2017). Este tipo de revisión involucra seleccionar una muestra randomica del conjunto de papers y revisarlos haciendo preguntas, mientras el otro investigador explica su extracción.
5.3. Consideraciones adicionales (datos calculados, subjetivos, etc.)	-
6. Proceso de Síntesis de Datos	
6.1. Tipo de Síntesis	Los datos categorizados serán presentado tabularmente. Qué datos mostrar en qué tablas será decidido luego de tener los resultados. Los datos sobre desafíos y beneficios (RQ5 y RQ6) serán analizados utilizando content analysis y open coding. La sintesis la hará Pizard y será revisada por Moreno.
6.2. Forma de validación de la síntesis	Moreno revisará la síntesis preliminar y Pizard ajustará según sus comentarios. Luego, Barbara Kitchenham revisará los resultados de la síntesis.
7. Limitaciones del Estudio	
7.1. Limitaciones (de constructo, internas y externas)	Kitchenham se unió tarde al equipo y sus recomendaciones sobre la extracción y síntesis fueron realiza No tuvimos en cuenta papers en otro idioma diferente al inglés. Usamos lean peer review para validar algunos datos extraídos, esto no es algo usual en las SRs.
7.2. Otras limitaciones (conflictos de intereses, etc.)	-
8. Informe	
8.1. Consideraciones (público objetivo, estrategia de difusión, etc.)	Se publicará un paper sobre la revisión. El formato y contenido está sujeto a la elección de la conferencia o revista a la cual enviar el paper.
9. Cronograma	
9.1. Cronograma con principales etapas	-

1 EBSE SLR example report - 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[1]. 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.

1.1 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?

1.2 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.

1.2.1 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

[1] This example was extracted from: Pizard, S., Acerenza, F., Otegui, X. et al. Training students in evidence-based software engineering and systematic reviews: a systematic review and empirical study. *Empir Software Eng* 26, 50 (2021). https://doi.org/10.1007/s10664-021-09953-9

in all of the searches, though some adaptations were made to it due to differences in the digital libraries. Table 2 shows the search strings used in the different search engines, which were adapted from the original string presented in Table 1. 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 3 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 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.

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”))

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 2 Adapted search strings

Search	First Stage: Papers Agreed Include	First Stage: Papers Agreed Exclude	First Stage: Papers Disagreed	First Stage: Papers Total	First Stage: Kappa	First Stage: Papers Selected	Second Stage: 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 3 Results of selection process

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 4 Papers by search

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.

Table 4 presents the studies obtained from each search carried out (studies details will be presented Table 6).

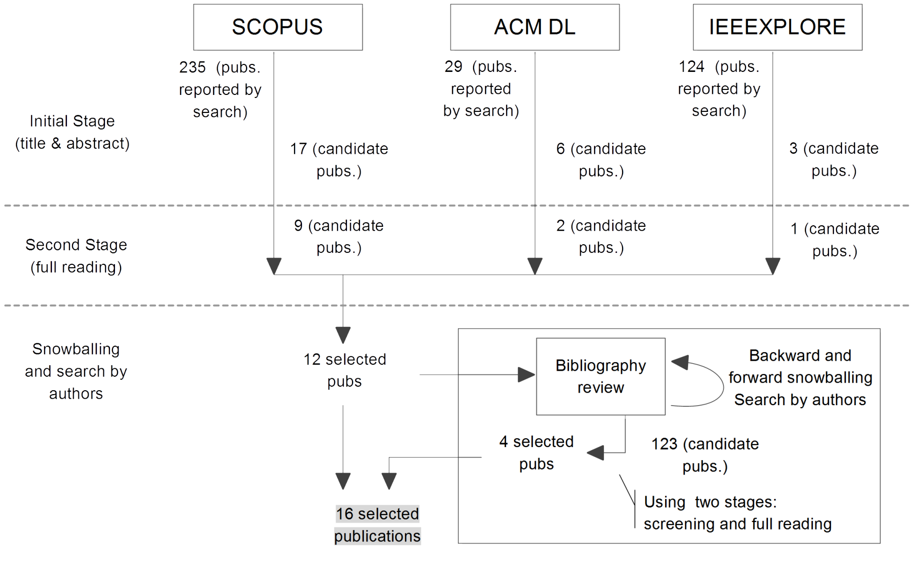

Fig. 1 Surveyed literature flowchart.

1.2.2 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 and a synthesis method based on following the Miles and Huberman's Qualitative Data Analysis method (Miles et al., 2014).

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.

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).

1.2.3 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). 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.

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

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 5 Initial agreement in the categorization of studies

coefficient between the values for each reviewer both for the number of relevant questions and for the average quality score for each study.

1.2.4 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 5).

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 ($p < 0.05$).

1.3 Reported Initiatives and their Context (RQ1, RQ2)

Table 6 presents the selected studies and their general characteristics and Table 7 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 7).

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.

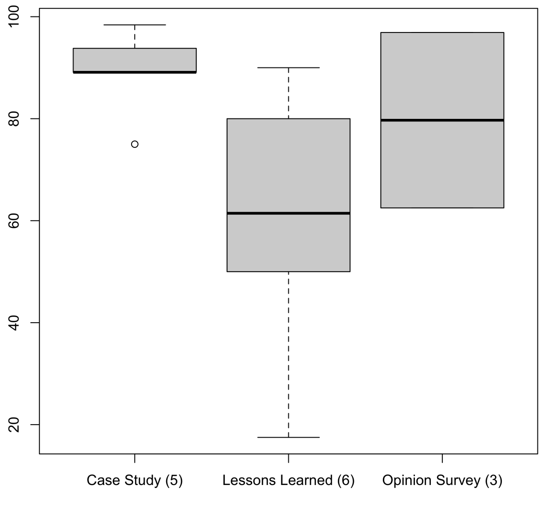

Fig. 2 Quality score for types of study (number of studies in parenthesis).

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 6 General characteristics of studies on EBSE teaching initiatives

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 7 Context of EBSE teaching initiatives

1.4 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 8). 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 8). 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

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 8 Content and Methodology of EBSE teaching initiatives

Table 9 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 10 Common issues and recommendations

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 9, 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.
- None of them included individual written tests, nor is it clear if any studies included theoretical and not only practical aspects in their evaluations.

1.5 Difficulties and Recommendations (RQ5)

The common issues (see Table 10) 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 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.

1.6 Benefits for Students (RQ6)

As shown in Table 11, 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.

1.7 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).

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 assign-

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 11 Common benefits

ment, 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.

1.8 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.

References

Aglen B (2016) Pedagogical strategies to teach bachelor students evidence-based practice: A systematic review. Nurse Education Today 36:255–263

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

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

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

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

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

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

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, 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

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

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

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

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

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