Tesis de Doctorado en Informática

Investigating Evidence-based Software Engineering Training and Adoption

Sebastián Pizard

Directores de Tesis

Dr. Barbara Kitchenham ..... Keele University
Dr. Diego Vallespir ..... Universidad de la República

Tribunal

Dr. Carolyn Seaman (Revisor) ..... University of Maryland Baltimore County
Dr. Dag Sjøberg (Revisor) ..... University of Oslo
Dr. Daniel Calegari ..... Universidad de la República
Dr. Guilherme Horta Travassos ..... Universidade Federal do Rio de Janeiro
Msc. Omar Viera ..... Universidad de la República

Director Académico

Dr. Diego Vallespir ..... Universidad de la República

PEDECIBA Informática
Instituto de Computación – Facultad de Ingeniería
Universidad de la República
Montevideo, Uruguay
Setiembre 2024

Acknowledgments

My most sincere thanks to my supervisors and co-authors, Dr. Barbara Kitchenham and Dr. Diego Vallespir, for your support and guidance over this journey.

Thank you, co-authors, Dr. Silvana Moreno, Ximena Otegui, Fernando Acerenza, Joaquín Lezama, and Rodrigo García, for always being supportive.

Thank you, Natalia Milán, for advising us on the nuances of the English language.

Thank you, Dr. Guilherme Horta Travassos, Dr. Martín Solari, and Dr. Forrest Shull, for listening to the initial ideas of this research and for your helpful comments.

Thank you, Angie Lecot, Javier Barreiro, and Elena Machuca, for your support in certain activities of this research.

Thank you, colleagues at the Instituto de Computación of the Universidad de la República, especially Lucyla Alonso and Mónica Martinez, for your unconditional listening and companionship.

Thank you, family and friends, especially my parents, Elena and Luis; my brothers, Mauricio and Santiago; and my partner, Matías, for always being there for me. Over the years, you patiently listened to my research musings, even though you had no interest in them.

Abstract

Background: Evidence-based software engineering (EBSE) could bridge the gap between academia and industry by bringing together academic rigor and research of practical relevance. However, while Systematic Reviews (SRs) were initially introduced to support EBSE, their widespread adoption by researchers contrasts with the limited evidence about EBSE adoption in non-academic contexts. Moreover, the importance of training for embracing evidence-based practice has been recognized in other fields, yet remains relatively understudied within EBSE.

Objective: This research aims to investigate how the concepts of EBSE and SRs should be taught to future software engineering (SE) practitioners and to assess EBSE's role in supporting industry practice.

Methodology: The first objective was addressed by reviewing EBSE teaching studies, proposing an EBSE course, conducting an empirical evaluation of the course's effectiveness, and assessing its impact on the trainees' working practices. The second objective was approached by examining attitudes toward EBSE among members of a government agency and conducting an EBSE-focused project aimed at solving practical issues within a software company. This project allowed us to assess the outcomes of Rapid Reviews (RRs) —a streamlined form of SRs— in supporting SE practices and to evaluate the EBSE framework efficacy. Given its exploratory nature, the research is grounded on qualitative methods.

Results: Our course, that formalizes EBSE teaching through a set of learning outcomes, received favorable evaluations and demonstrated a beneficial impact on student work practices. Government agency staff acknowledged the utility of EBSE, while practitioners without SE research experience found RR results valuable and reliable than other sources. Nonetheless, using EBSE necessitated a combination of professional and research skills, and limitations in SE evidence required formulating broader research questions and tailoring results to actionable recommendations.

Conclusion: This thesis underscores the value of EBSE in fostering collaboration between academia, industry, and government stakeholders. It delineates strategies for teaching and adopting EBSE, while also highlighting challenges that required attention in its application.

Keywords: Evidence-based Software Engineering, Training, Rapid Reviews, Government Agencies, Industry–Academia Collaboration, Qualitative Research.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

This section introduces the motivation, objectives, and contributions of this thesis.

1.1. Motivation

Software engineering (SE) is an engineering field that focuses on every element of software development, starting from the initial phases of defining the system requirements to supporting the system once it's operational [108]. As in any discipline, knowledge in SE is mainly supported by evidence, defined by the Oxford Dictionary as "the available body of facts or information indicating whether a belief or proposition is true or valid." Evidence can take various forms, such as anecdotal evidence (based solely on personal observation, gathered informally and non-systematically) or scientific evidence[1] (generated through the use of hypotheses and investigated using the scientific method). Although all forms of evidence contribute to knowledge and are used in decision-making, scientific evidence is generally considered more reliable due to its rigorous methods aimed at minimizing bias [95]. The growing emphasis on evidence-based decision-making across various fields [70] suggests that the use of scientific evidence may be an indicator of a field's maturity.

In SE, research involves the application of scientific principles, techniques, and methodologies to improve the understanding, development, and maintenance of software systems. The ultimate goal of SE research is to advance the field and provide practical insights and solutions to the challenges faced by software developers and organizations in creating high-quality software systems.

SE research is complex [7,39,88,111]. This is due to the diverse subject areas it covers, the constantly evolving technology landscape, the interdisciplinary nature requiring knowledge from multiple fields, the complexity of modern software systems, practical constraints like limited access to real-world projects, and the need to balance theoretical advancements with practical applicability in software development.

To address this complexity, empirical research, based on observation and experimentation, has been employed since the 1960s [34]. In brief, SE empirical research involves the collection and analysis of data from real-world software projects, experiments, surveys, case studies, and other empirical methods.

In the last decades, the use of empirical studies in software engineering research has seen significant growth. For example, a comparison between the IEEE/ACM International

[1] Analogous to the approach of Kitchenham et al. [54], this thesis focuses on considering scientific evidence. Therefore, unless stated otherwise, any reference to ´evidence' pertains to findings derived from scientific research.

Conference on Software Engineering (the most relevant SE conference) in 2002 and 2016 reveals a notable increase in the reporting of empirical studies and the utilization of empirical models [113]. In 2002, no papers reporting empirical studies were accepted, whereas, in 2016, this category comprised 30 % of the accepted papers. Empirical studies are now an integral part of the majority of papers published in prominent software engineering venues [112].

The importance of aggregating knowledge from the growing numbers of empirical studies has become increasingly significant in gaining a deeper understanding of SE [34]. As highlighted by Shull et al. [106], no individual empirical study can be deemed definitive or applicable to all contexts. In light of this need to aggregate evidence from research, evidence-based software engineering (EBSE) was introduced as an adaptation from the health field in 2004 by Kitchenham et al. [56]. EBSE emphasizes the use of evidence from research to inform decision-making, improve practices, and advance knowledge in the field of software engineering. It involves systematically gathering, analyzing, and applying relevant research findings and empirical data to address software engineering challenges and make informed decisions.

Systematic reviews (SRs) were introduced as a means to support EBSE. In various fields, evidence obtained from SRs is considered the most reliable and is used to inform state policies, regulations, and professional practice decisions. This is because SRs provide a summary of relevant scientific evidence that has been subject to critical assessment and provides an overall evaluation of the strength of the evidence. A notable example is in medicine, where evidence-based practice is widely recognized as successful [33].

In practice, adopting EBSE, that is, at a very high level, aggregating scientific results, interpreting them, and applying them in professional practice occurs through the interaction of various stakeholders, within different social processes (e.g., including the interpretation and discussion of evidence by professional networks and communities of practice) [35]. The complexity of these processes and the importance of considering different stakeholders were highlighted in the first EBSE paper, which indicated that it could only be addressed with "extensive collaboration and long-term commitment" among the various stakeholders of our field [56]. This collaboration should encourage the different types of stakeholders to contribute based on their talents and experience. For instance, researchers are better suited to conduct SRs and interpret the evidence collected, while industry and government practitioners may have a clearer understanding of the evidence needs that should motivate SRs and how to integrate their results into professional practice and policy making. Therefore, the degree of each individual's participation in the adoption of EBSE depends on their profile but always seeks to incorporate the best available research evidence into decision-making.

So far, we have more knowledge of EBSE adoption by researchers, who have conducted a vast number of SRs. To illustrate, Kamei et al. identified 446 SRs published solely in the top SE journals and conferences before 2019 [48]. Meanwhile, we know very little about the adoption by other types of stakeholders, especially regarding the incorporation of evidence into decision-making or policy development. The EBSE and SR book published a decade after its introduction [54] reported a solitary study on the practical application of EBSE [50]. This suggests that the central goal of EBSE is not being fully realized.

Recent research has been conducted to propose and evaluate mechanisms for transferring knowledge from SRs to SE practice, using rapid reviews (RRs) (i.e., expedited SRs designed to promptly provide research evidence) and evidence briefings (i.e., a succinct one-page document that summarizes relevant research findings) [21, 23]. However, our understanding of the extent to which EBSE has been adopted beyond academic circles is limited, and there is a need to explore ways to improve its adoption in practical settings. There is still a need to investigate what role, if any, EBSE should have in supporting non-academic stakeholders.

1.2. Objectives

The goal of my research program is to explore how to train future software engineering practitioners in EBSE and SRs and to evaluate the role of EBSE in supporting both the needs of practitioners and the needs of other non-academic stakeholders. To achieve this goal, the research was divided into three specific objectives:

1. **Investigate how to train future SE practitioners in EBSE and SRs.** Lack of adequate training is one of the main barriers to evidence-based practice in other disciplines [96, 104, 115, 118]. Therefore, the specific objective referred to this area is to investigate how to provide adequate EBSE training to future SE practitioners. Furthermore, an additional objective is the evaluation of potential effects of this training on the work practices of trainees.

2. **Evaluate the role of EBSE in supporting government agencies.** Government agencies are key players in the adoption of evidence-based practice (see, for example, [1]). The lack of studies that report the use or consideration of EBSE by government agencies or regulatory bodies indicates room for research. Thus, the specific objective in this area is to investigate whether EBSE could be considered for adoption by government agencies (e.g., by commissioning SRs or using existing SRs to inform the development of IT-related policies and regulations).

3. **Evaluate the role of EBSE in supporting software industry practitioners.** Until this thesis, very few SRs conducted in collaboration with the industry were known. In fact, my colleagues and I only knew two such reviews [21, 50]. Currently, several published SRs are practice-driven, taking various forms such as RRs or full SRs (see, for example, [12, 65, 109]). All these collaborative SRs seem to occur together with practitioners with knowledge in SE research or with experience in the topic investigated. Furthermore, while SRs are valuable in supporting EBSE, the scope of EBSE goes beyond mere evidence search and synthesis. It encompasses tasks such as formulating evidence-based questions, applying evidence within specific contexts while considering stakeholder input, and evaluating the overall use and effectiveness of such evidence. Considering this, the latter part of my research aimed to enhance the comprehension of EBSE applications in industry settings (i.e., the use of evidence from secondary studies to inform decision-making in professional practice). Specifically, this comprises: (1) evaluating the broader value of the EBSE framework within an industry application, and given the recent trend to employ RRs in collaborative academic-industry endeavors, (2) assessing the efficacy of RRs in supporting SE practice of industry practitioners.

1.3. Methods and Context

Considering my limited prior knowledge and experience in the three areas described above, the research strategy heavily relied on existing work in the field of SE when applicable, or on prior evidence-based practice (EBP) research in other fields. An exploratory research approach was predominantly employed, with a focus on qualitative methods. The achievement of the specific objectives was pursued through five empirical studies conducted in diverse settings (henceforth S1, S2, S3, S4, and S5).

Firstly, to investigate how EBSE and SRs could be taught to future SE practitioners the following activities were carried out: (1) Reviewing previous publications on EBSE teaching. (2) Proposing an undergraduate EBSE course that can lay the groundwork for future agreements. (3) Conducting an empirical evaluation of the course's effectiveness. (4) Assessing

the impact of EBSE training on the work practices of the trainees. This research on EBSE training was undertaken at the Faculty of Engineering of the Universidad de la República. In this context, I proposed an EBSE course for undergraduate students in the Computer Science curriculum and, along with other instructors, delivered it on several occasions. The SR of EBSE teaching reports, the proposal for an EBSE and SRs course, and its first empirical evaluation were reported in the study S1 (introduced in Chapter 3 and detailed in the first part of Appendix A). Meanwhile, the evaluation of three consecutive courses as well as an investigation of the effects of EBSE training on former students' work practices is presented in study S2 (introduced in Chapter 3 and detailed in the second part of Appendix A).

Secondly, I sought to explore whether EBSE could be considered for adoption by government agencies by studying the attitudes towards EBSE among members of a particular government agency and assessing the effects of learning about EBSE on their subsequent work practices. For this, I partnered with the Agency for the Development of Electronic Government and Information Society and Knowledge (AGESIC, by its Spanish acronym) of Uruguay to introduce EBSE to the agency's staff members and examine their attitudes towards EBSE. This research is reported in study S3 (introduced in Chapter 3 and detailed in Appendix B).

Thirdly, to improve the understanding of EBSE applications in industry settings I led an EBSE-based project to address a practical problem of a software company. This involved collaborating with practitioners to diagnose the problem, conducting an RR to gather evidence, and subsequently disseminating the findings to the company. I used this project to (1) replicate the original study of RRs in SE [21] and assess their utility in supporting SE practices in an industry setting (introduced in Chapter 3 and detailed in the first part of Appendix C) and (2) evaluate the value of the broader EBSE framework, identifying benefits, barriers and enablers (introduced in Chapter 3 and detailed in the second part of the Appendix C).

1.4. Main Contributions

The contributions of this thesis are related to the study of the adoption of EBSE from the perspectives of its training, its use in government agencies, and its application in industry settings. These contributions can be summarized in the following points:

- Publication of an evidence-informed and validated EBSE and SR course suitable for Computer Science students, which is freely available in both English and Spanish.
- Providing evidence that an EBSE and SR related course can be of benefit to students in their subsequent careers.
- Reporting on the introduction of EBSE in an IT-related government agency, which confirmed its value for government agencies and regulatory bodies and also provided insights into the benefits of, and barriers to, its adoption.
- Offering an analysis of the RRs reported in SE and presenting evidence that RRs can support SE industry practice, even when the practitioners utilizing the evidence lack prior experience in SE research.
- Presenting a comprehensive evaluation of the EBSE steps in an industry application, identifying issues, barriers, and facilitators that arose during the process.

1.5. Publications

Part of the research and its results have been validated by several peer-reviewed publications. In particular, two articles have been published in international journals indexed in ISI-JCR and another article has been accepted and presented at the main software engineering conference. Below is a summary of the publications related to this thesis.

- The S1 study, which reports the SR of EBSE training reports, the proposal for an EBSE course and SRs for undergraduates, and the conducted case study for its evaluation, has been published as *"Pizard, S., Acerenza, F., Otegui, X., Moreno, S., Vallespir, D., Kitchenham, B. 2021. Training students in evidence-based software engineering and systematic reviews: a systematic review and empirical study. Empir Software Eng 26, 50."* [81]. The article, as published in the Empirical Software Engineering Journal, is presented in Appendix A.

- The S2 study, which sought to validate the effectiveness of our training and assess the impact of EBSE training on trainees' attitudes and behaviors, particularly in their work practices across various time periods, was published as *"Pizard, S., Vallespir. D., and Kitchenham, B.. 2022. A longitudinal case study on the effects of an evidence-based software engineering training. In ACM/IEEE International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET). Pittsburgh, PA, USA."* [86]. The article, as published in the International Conference on Software Engineering, is presented in Appendix A.

- The S3 study, which aimed to explore the attitudes of stakeholders within a government agency (GA) towards evidence-based software engineering (EBSE) and evaluate the influence of EBSE knowledge on their work practices, was published as *"Pizard, S., Acerenza, F., Vallespir, D., Kitchenham, B., 2023. Assessing attitudes towards evidence-based software engineering in a government agency, Information and Software Technology, 154"* [83]. The article, as published in the Information and Software Technology Journal, is presented in Appendix B.

- The S4 study, which reports a systematic review of published SE RRs and the replication of the original study proposing RRs for SE (i.e., [21]), has been accepted for publication in the Empirical Software Engineering journal. The preprint is available at [85].

Additionally, another article has been submitted to software engineering journal and is currently under review.

- The S5 study, which includes a report on the evaluation of the broader EBSE framework, identifying its benefits, barriers, and enablers, has been submitted to the IEEE Transactions on Software Engineering and is under review. The preprint is available at [87].

1.6. Other Outcomes & Dissemination Activities

- I made available the updated materials of our EBSE and SRs course both English and Spanish, which can be openly accessed and used [82].

- I recorded a talk for the 2022 Uruguayan Software Engineering Program (IS.Uy) online channel presenting two EBSE application cases conducted in Uruguay [79].

- Together with Juliana Herbert, I gave a talk at the 2018 Uruguayan Software Engineering Sessions (Jornadas de Ingeniería de Software del Uruguay) on EBSE and its application to improve agile practices in software development [84].
- I led a small group of researchers in the preparation of a technical report to bring the main concepts of EBSE and SRs closer to Spanish-speaking readers [80].

1.7. Authorship Statement

The papers included in this thesis are the result of collaborative work among researchers. All co-authors have reviewed and approved the final versions of these papers. The contribution of each author is described in detail within the report of each study, as required in empirical qualitative research. However, I provide a brief summary of my contribution below.

- Study S1: I participated in the conduct of the SR on EBSE training. Subsequently, I developed the learning objectives and designed the EBSE course for undergraduate students. I was one of the teachers of the EBSE courses taught in 2017, 2018 and 2019. I led the data analysis and wrote the first version of the report.
- Study S2: Together with the other co-authors, I designed and distributed a questionnaire to evaluate the impact of EBSE knowledge on former students of the 2017-2019 courses. I was primarily responsible for the data analysis and report.
- Study S3: To prepare the work with AGESIC, I planned and conducted a rapid review on the adoption of evidence-based practice in other fields. I designed and conducted a workshop to disseminate EBSE at AGESIC and also led the focus group to gather participants' opinions. Subsequently, I participated in the elaboration of a questionnaire to evaluate the impact of EBSE knowledge on the participants. I led the data analysis and reporting of the study.
- Study S4: As the first author, I led the project team in collaboration with a local company. I led the data collection and analysis. To study related work, I participated in the conduct of a mapping study of previous SE research using RR. I wrote the draft of the report.
- Study S5: As the first author, I proposed the idea for the empirical study. I conducted the preliminary data analysis. To assess the state of practice-driven adoption of EBSE, I reviewed and analyzed a subset of SE SRs. I wrote the first version of the report.

1.8. Document Structure

As stated above, the thesis is organized as a compilation of studies, with some already published. It consists of two main parts. The first part provides a comprehensive overview, presenting the background, research objectives, methods employed, a summary of the conducted studies within the broader strategy, and a general conclusion. The three appendices contain detailed information about the compiled studies included in this thesis.

Below is a brief overview of each chapter.

Chapter 1 introduces the work done in this thesis. It presents the motivation, the objectives of the research, and the structure of this document.

Chapter 2 gives an overview of evidence-based software engineering. It describes its foundations, its role in software engineering research, and some known insights into its adoption status.

Chapter 3 presents a summary of the research conducted in this thesis. It provides an overview of the studies included, outlining their objectives, methods employed, and key findings and reflections.

Chapter 4 includes a general discussion of the studies of this thesis.

Chapter 5 includes final remarks and establishes lines of future research.

Appendix 1 presents the studies carried out to investigate the training in EBSE and SRs. It encompasses an SR of teaching reports on EBSE and SRs, a comprehensive description of our proposed course for EBSE and SRs, including an evaluation of its adequacy. Furthermore, it includes a study on the impact of training on the work practices of former students.

Appendix 2 presents the study conducted to assess the possibility of adopting EBSE within government agencies. The study took place in a Uruguayan government agency and involved introducing EBSE, conducting a focus group to explore the initial attitudes of participants, and subsequently circulating a survey to understand the changes in their work practices after being exposed to EBSE concepts.

Appendix 3 presents the research carried out to evaluate the application of EBSE in an industry setting. It presents two studies conducted within the context of an EBSE-based project aimed at resolving an industry problem. Firstly, it provides an in-depth examination of the RR conducted during the project, including collaboration with practitioners and their feedback. As related work its include an analysis of the RRs reported in SE. Secondly, it offers an evaluation of the broader application of the EBSE framework, encompassing each of the EBSE steps.

Chapter 2

Background

This chapter introduces two relevant areas that underpin this thesis. Firstly, Section 2.1 presents key aspects of evidence-based software engineering and its adoption. Secondly, Section 2.2 provides a brief description of the most significant methodological approaches and lessons learned from applying them.

2.1. Evidence-based Software Engineering

This section provides an overview of the motivation behind introducing evidence-based software engineering (EBSE), its key characteristics, and a brief discussion of its adoption by researchers and other stakeholders.

2.1.1. Motivation

Software engineering (SE) stands out from other engineering disciplines due to its unique characteristics [8]. It involves creative development processes rather than manufacturing, with each software product being a creation rather than a replication. Software's intangibility poses challenges in understanding its structure and processes, requiring learning from diverse situations and abstract thinking. Developing models for future use requires additional resources and, without explicit support, quality improvements in the development process may not occur.

In the early days of SE research, the predominant focus was on theoretical models and frameworks for software development [13]. However, as software systems became more complex and larger in scale, for some researchers it became evident that understanding and improving SE practices required a more empirical approach [34]. Empirical refers to a methodology or approach that relies on observation, experimentation, or data collection to gain knowledge and draw conclusions. It emphasizes the use of evidence from real-world observation or experiences as the basis for understanding and analyzing a subject or phenomenon.

Although many believe that SE is still governed by folklore that has been accepted as facts [13], empirical research is now widely adopted by researchers and has confirmed or refuted part of the previous knowledge. As pointed out by Méndez Fernéndez et al. [69], one example is grounded in the well-known essay by Dijkstra titled "Go To Statement Considered Harmful"from 1968 [30]. This essay sparked a public exchange among scholars who published notes expressing their views on the topic. However, this debate relied primarily on argument-based reasoning and remained within the scholarly community. In 2015, almost 50 years later, Nagappan et al. conducted a large-scale study analyzing C code from GitHub

repositories [72]. Their study concluded that the practical use of goto statements does not appear to be harmful, providing evidence from research to support their findings. However, this does not imply that Dijkstra was wrong about coding problems in the 1960's. It is possible that Dijkstra's article influenced programming practices in the subsequent years, leading programmers to use goto statements less frequently and more carefully. Nonetheless, evidence from research has allowed us to conclude that there is now no systematic problem with the use of the goto command.

Although the first empirical studies on SE date back to the 1970s, it was in the mid-1990s that there was a growing trend in its use [8, 54]. By this time, some researchers began to consider how to find, select, and aggregate the results of various empirical studies to obtain evidence that could be used to improve practice. This was prompted by various concerns, e.g. the difficulty or inadequacy of applying results from a single empirical study due to its specific context or limitations, or the bias in the selection of studies in classic reviews.

2.1.2. Evidence-based Software Engineering & Systematic Reviews

In 2004, Kitchenham et al. introduced the concept of evidence-based practice from the healthcare field to SE. [56]. Evidence-based software engineering (EBSE) was proposed to improve decision-making related to software development and maintenance by integrating the best current evidence from research with practical experience and human values.

The five steps of EBSE are the following.

1. Converting the need for information into an answerable question.
2. Finding the best evidence with which to answer that question.
3. Appraising evidence validity, impact, and applicability.
4. Integrating the appraised evidence with expertise and stakeholders' values and circumstances.
5. Evaluating effectiveness and efficiency in executing previous steps and seeking ways to improve.

Medical researchers recommend using systematic reviews to support evidence-based practice, whether by utilizing existing systematic reviews or by applying the systematic review method to aggregate information from individual empirical studies. Systematic reviews (SRs) are a form of secondary study that aggregate evidence from primary studies, including controlled experiments, case studies, and surveys, among others. SRs provide a rigorous and transparent approach to searching and synthesizing the existing research evidence on a particular topic or research question. SRs allow to collect and synthesize evidence from different sources. They are the core tool of the evidence-based approach, as they usually support steps 1-4 of the EBSE process. The key feature that distinguishes them from traditional (or classical) narrative reviews is their explicit attempt to minimize the chances of reaching wrong conclusions, which may result from the bias in the primary studies or in the review process itself [32]. To achieve this, a protocol must be established with all the activities to be carried out and criteria to be used specified prior to the conduct of the SR. Figure 2.1 summarizes the stages of an SR.

The stages of conducting an SR in SE typically include [55]:

[1]By Centre for Health Communication and Participation La Trobe University, Australasian Cochrane Centre [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons

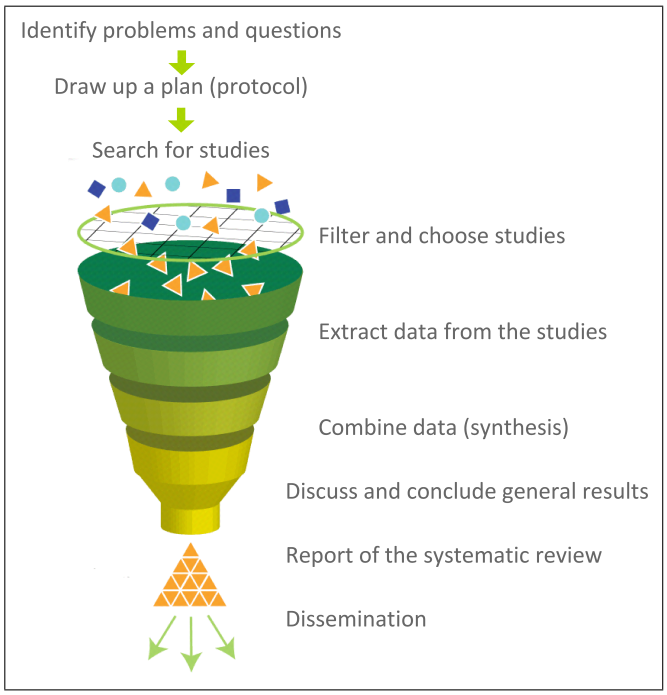


Figure 2.1: Process of a systematic literature review[1].

1. Formulating Research Questions: Clearly define the research questions or objectives that the SR aims to address.

2. Planning and Protocol Development: Develop a detailed plan and protocol for the SR. This includes defining the inclusion and exclusion criteria for selecting studies, determining the search strategy, specifying the data extraction process, and outlining the synthesis and analysis techniques.

3. Literature Search: Conduct a comprehensive search to identify relevant studies. This may include automatic searching across multiple databases, such as IEEE Xplore, ACM Digital Library, Scopus, and others, manual searching in one or more journals or conferences, or a combination of both.

4. Study Selection: Apply the predefined inclusion and exclusion criteria to screen the retrieved studies. This could include a multi-stage selection procedure. Typically including, first, screening of titles and abstracts to eliminate irrelevant studies and, second, conducting a full-text review of the remaining studies to determine their eligibility for inclusion in the SR.

5. Data Extraction: Extract relevant information from the selected studies, such as study characteristics, research methods, software engineering topics, and key findings. This process involves designing data extraction forms and ensuring consistency in data extraction across the included studies.

6. Quality Assessment: Evaluate the quality and validity of the included studies. This step may involve using specific assessment criteria or tools to assess the methodological rigor and relevance of each study.

7. Data Synthesis and Analysis: Analyze and synthesize the extracted data from the included studies. This can be done using various techniques, such as qualitative analysis, quantitative meta-analysis, or mixed-methods approaches, depending on the nature of the research questions and the available primary studies.

8. Reporting and Dissemination: Document the entire SR process, including the search strategy, study selection process, data extraction forms, analysis techniques, and findings. Write a comprehensive report and establish mechanisms to transfer knowledge to practice. Additionally, following established guidelines can help ensure a high-quality SR report [57].

These stages can be iterative, and adjustments might be necessary based on the outcomes of each stage.

This report of the stages of the SR process and their descriptions are based on earlier SE guidelines [54, 55], and were the foundation of the research in this thesis.

However, the SE guidelines for reporting secondary studies (SEGRESS) [57], published this year, can be considered as the latest iteration of guidance for the SR process. The SEGRESS guidelines introduce several changes aligned with the evolving standards of evidence-based practice (EBP) and SRs in medicine. Two important modifications refer to certainty assessment in SRs, a concept initially introduced to SE by Dybå and Dingsøyr [32]. These changes involve: (1) adopting the terminology "Risk of Bias" in place of "Quality assessment", and (2) emphasizing the evaluation of the "Certainty of evidence", which relates to a set of studies rather than individual primary studies. This is particularly relevant for large SRs that encompass multiple findings, each of which may rely on a distinct set of individual studies. These changes appear highly pertinent for enhancing the maturity of SRs in SE. They also underscore the need to consider current medical terminology and changes to the medical SR process. It is both reasonable and essential to monitor changes in SRs proposed in the health field, where EBP is more developed, and integrate significant ones into the SE community to improve our research and practice.

Systematic reviews can be classified into the following different categories, which seek to answer different types of questions (descriptions of the first three were adapted from [18]).

1. **Mapping studies.** They seek to answer questions such as "What studies have investigated...?". These reviews not only seek to identify relevant studies but also offer some level of analysis, often categorizing the studies (e.g., by types of research methods, participant demographics, and other relevant criteria). Mapping studies provide a valuable overview of a particular subject area and can assist in determining gaps in research on which further work can be done.

2. **Qualitative Systematic Reviews.** Its primary purpose is to gather and synthesize data from studies that examine the advantages, disadvantages, and stakeholder perspectives concerning the utilization of SE technologies. Furthermore, it aims to prioritize the most critical issues identified in these studies. This type of review often involves the aggregation of findings derived from thematic analyses of case studies and/or industry opinion surveys. Consequently, it can provide valuable guidance on matters to contemplate when embracing new technologies, including the identification of potential obstacles.

3. **Quantitative Systematic Reviews.** As the name suggests, these reviews aggregate quantitative findings, often presented as statistical data. Whenever possible, they employ statistical techniques like meta-analysis to provide numerical information. For instance, a quantitative SR may yield outcomes such as a ranking of various techniques based on their effectiveness.

4. **Rapid Reviews.** They are designed to expedite the traditional SR process by simplifying or eliminating certain steps, enabling the efficient generation of evidence with limited resources [41]. Some key characteristics of RRs in SE, slightly adapted from [21], are: timely results and reduced costs, collaboration with practitioners, and appropriate dissemination of its results.

Although SRs support EBSE, they are not synonymous [9]. As stated above, EBSE encompasses more than just the process of searching for and synthesizing evidence. To recap, it involves: (1) Transforming practical problems into well-formulated questions that can be addressed using evidence, (2) applying the evidence derived from SRs while taking into account the specific context, preferences, and expertise of stakeholders, and (3) evaluating both the utilization of evidence and the overall performance of the EBSE process itself.

2.1.3. Adoption of SRs and EBSE

Ever since the introduction of EBSE and SRs in 2004, researchers have extensively embraced SRs as a valuable tool for exploring and synthesizing research on numerous SE subjects. It is estimated that in only the first ten years, over 200 systematic reviews have been published [54]. More recently Kamei et al. identified 446 SRs published before 2019 only in the top SE journals and conferences (i.e., a minimum h5-index of 20 for conferences and 25 for journals) [48].

Several SRs yielded highly significant results that have influenced research to the point of reshaping common understanding within the field. For example, in 2004, Jørgensen [44] found no evidence that estimation models are better than experts' estimation (a common assumption at the time). They noted that there are situations in which models do not include important information about the application domain that is actually considered by experts.

Within the comprehensive definition of its five steps, EBSE aims to enhance decision-making processes in software development and maintenance by integrating the best available research evidence, practical experience, and human values. This ambitious objective recognizes the significant gap that often exists between research and practice. To bridge this gap, EBSE promotes a stronger emphasis on methodological rigor while prioritizing the relevance of research to practical applications [33]. Considering this, several authors suggest that EBSE has the potential to yield more practical outcomes and enhance collaboration between academia and industry [29, 62].

Thus far, there is a dearth of evidence showcasing the implementation and acceptance of the broader framework of EBSE. Moreover, the lack of EBSE adoption has been acknowledged by multiple authors [22, 23, 28, 42, 54]. Some of these authors have specifically pointed out that the limited adoption could be attributed to SRs not adequately addressing practice-relevant issues [23, 101] or lacking practical recommendations for practitioners [28].

While these studies offer valuable insights, they primarily focus on examining SRs (or RRs) and their characteristics rather than directly exploring the adoption of EBSE. It is worth considering the significance of investigating barriers to evidence-based practice in other domains. For instance, the role of appropriate training, potentially formalized like in the field of medicine [3], emerges as a crucial factor [96, 104, 115, 118]. Additionally, Kitchenham et al. suggest that the widespread implementation of EBSE necessitates interest and broad collaboration from not only academics but also industry practitioners, government agencies, and regulatory bodies [56]. Research that considers these factors can provide further insights into facilitating EBSE adoption.

2.2. Key Methodological Aspects and Takeaways

This thesis summarizes multiple studies, and each study includes a report of the adopted research methodology. In this section, I discuss specific methodological aspects which influenced much of my research program,specifically: the use of RRs and SRs as a research starting point, the application of qualitative research methods, and my role in managing the research teams.

2.2.1. Using RRs and SRs as Starting Points for Research Activities

Secondary studies (i.e., RRs or SRs) were used as the starting point of all the research that is part of this thesis. In particular:

- With Silvano Moreno we conducted an SR to gather previous experiences of teaching EBSE and SRs. With Fernando Acerenza and Ximena Otegui we used the SR outputs to design our EBSE and SRs course. For example, to choose suitable teaching methods or considering the common challenges beforehand. Both SRs and the course are reported in study S1 (reported in Appendix A).
- I conducted an RR to better understand the adoption of EBP in other disciplines. In particular, three aspects were analyzed: the perceived value of EBP, the main barriers to the adoption of EBP, and the mechanisms to improve its adoption. Understanding the challenges and facilitators of EBP supports our analysis and interpretation of the results of our assessment study of attitudes toward EBSE in a government agency. Both the RR and the study with the government agency are reported in study S3 (reported in Appendix B).
- In the research carried out to support a software company improving its knowledge management practices, together with Joaquín Lezama and Rodrigo García I conducted an RR to gather information and approaches to support the company. This collaboration with a software company served as the basis for two studies, as mentioned before, one to study in greater detail the value of RRs and another to evaluate the application of EBSE and all its steps, which are reported in S4 and S5 respectively (reported in Appendix C).

Regarding the use of RRs and SRs to support research, my experience confirms that:

- SRs and RRs are very useful as starting points for research activities. Their results are useful especially when researchers do not have much knowledge or experience on the subject. However, there is some inherent difficulty in reviewing studies from other domains (knowledge of those disciplines is often required). As a way of addressing this problem, S3 including the conduct of an RR of secondary studies in EBP adoption. This strategy allowed for learning about the topic without the necessity of reading primary studies from unfamiliar areas.
- RRs are useful for providing relatively quick answers to specific problems (e.g., S3, and S4-S5). However, researchers must evaluate whether they are an appropriate method for their information needs and fully identify the limitations imposed by the decisions made in their conduct.
- Researchers can also use existing catalogs of SRs. Currently, in SE we do not have a recognized catalog of secondary studies as happens, for example, in Medicine[2]. As a

[2]In healthcare, the Cochrane Database of Systematic Reviews (CDSR) is considered by its proponents to be the foremost database for systematic reviews in the field. https://www.cochranelibrary.com/cdsr/about-cdsr

Table 2.1: Several characteristics of Focus Group (FG) application in SE [60]

Research Objectives	- Is a good method for gathering preliminary input on new concepts, crafting questionnaires, brainstorming ideas, identifying and ranking potential issues, receiving input on the presentation of models, and uncovering underlying motivations. - Not suitable for hypothesis testing, making conclusive decisions, gathering quantitative assessments, addressing sensitively charged topics, or examining complex subjects that cannot be adequately explored in one session.
Planning	The focus group sessions follow a predetermined schedule and structure, aiming to cover a manageable number of topics (allowing participants sufficient time to understand the issues and engage in meaningful discussions and interactions). However, the time constraint also limits the depth at which complex issues can be explored.
Participant Selection	The effectiveness of this method depends significantly on the experience and insights of its participants. Therefore, recruiting representative, perceptive, and motivated individuals is paramount for the successful conduct of a FG study.
Conduct and role of the moderator	The nature of discussion and interaction within a focus group session can vary widely. It may involve structured discussions or employ brainstorming techniques such as affinity grouping, teamwork methods, or even role-playing scenarios. Various methods can be employed for data capture during the session, including additional observers taking notes, audio or video recording, among others. The moderator's primary responsibility is to guide the discussion without imparting personal opinions or biases.
Analysis of data	It includes transcribing the recorded sessions if necessary, and then analyzing the data using qualitative analysis methods (e.g., [14, 71])

useful approximation, we can use the lists of secondary studies surveyed in tertiary studies (if they include the primary studies' full list). Personally, two catalogs that I find useful are:

- Kamei et al. [48] paper reports a tertiary study on the use of grey literature in SE, in which the search strategy is open to all secondary studies in SE. The authors identified 446 secondary studies published in reputable venues spanning from 2011 to 2018 (for a comprehensive list, see [49]).
- Budgen et al. [17] paper compiles 49 secondary studies featuring practical recommendations published through 2015. It offers concise summaries of each SR, highlighting their topics and key findings. This resource is valuable for practitioners seeking evidence from secondary studies on specific subjects and for educators looking to introduce EBSE using relevant SR examples.

2.2.2. Emphasis on Qualitative Research Methods

The primary objectives of the research in this thesis are exploratory in nature. The main focus has been the examination of attitudes and opinions within the SE community regarding EBSE. Thus, the research methodology has primarily centered on qualitative methods. In particular, I have employed the following methods:

- **Focus groups.** Focus groups are meticulously organized meetings aimed at gathering the viewpoints of participants regarding a specific topic or subject of interest [60]. Focus groups typically comprise 3 to 12 participants and are led by a moderator who maintains focus during discussions. Participant selection is based on their specific attributes

Table 2.2: Some characteristics of using Case Studies in SE [93, 94]

Selection of a Case	- The deliberate and intentional selection of the case and the units of analysis (e.g., projects, individuals, groups) is essential. - The case could be a specific software project, an organization's software development process, or a unique software-related problem. - This selection should align with the research objectives and questions.
Data Collection	- Various sources can be used, including interviews, surveys, documents, source code, logs, and observations. - Multiple data sources are often used to provide a comprehensive view of the case. - Data collection techniques cover various levels of engagement. First Degree (or Direct Methods): Researchers directly engage with subjects in real time to collect data using methods like interviews. Second Degree (or Indirect Methods): Researchers collect raw data without direct interaction with subjects, such as making observations through video recording. Third Degree (or Independent Analysis of Work Artifacts): Researchers use pre-existing and often compiled data sources, such as analyzing documents like failure reports.
Triangulation	Refers to the use of multiple viewpoints to develop a comprehensive understanding of phenomena. There are several forms such as: Using multiple data sources (Data triangulation), utilizing more than one observer (Observer triangulation), Combining diverse data collection methods (Methodological triangulation), and Incorporating alternative theories (Theory triangulation).
Ethical Concerns	Ethical considerations play a pivotal role in designing a case study in SE, where sensitive, confidential information is often involved. Clear protocols must be established upfront to address key ethical factors, including the following. Ensuring that all parties involved are informed and provide voluntary consent to participate in the study (Informed Consent). Safeguarding sensitive information, especially within organizations (Confidentiality) and Responsible management and dissemination of sensitive findings or data.
Data Analysis	Analysis methods may include qualitative techniques like thematic analysis or content analysis, as well as quantitative methods like statistical analysis, depending on the research goals.

relevant to the session topic. This group dynamic fosters the exchange and elaboration of ideas among participants, enhancing the depth and quality of the information obtained. Table 2.1 presents some other important characteristics of Focus Groups in SE.

- **Case studies.** Runeson et al. define case studies in the following terms [94] *"Case study in SE is an empirical inquiry that draws on multiple sources of evidence to investigate one instance (or a small number of instances) of a contemporary SE phenomenon within its real-life context, especially when the boundary between phenomenon and context cannot be clearly specified."* Table 2.2 describes other important characteristics of Case Studies in SE.

- **Action Research.** Action research aims to contribute to both the practical concerns of practitioners in an immediate problematic situation and to the goals of researchers through ethical joint collaboration [90]. It involves the application of a process that includes diagnosing, planning, intervening, evaluating, and reflecting, allowing for iterations to achieve results incrementally [31]. Table 2.3 describes other important characteristics

Table 2.3: Several key attributes of Action Research [31, 90]

Problem-Centered	The primary focus of action research is addressing specific problems or challenges within a particular context. It aims to bring about practical solutions and positive change in real-world situations.
Context-Specific	It is contextually grounded, meaning it is conducted within a specific environment or setting. The findings and interventions are tailored to the unique characteristics and needs of the particular situation.
Participatory	It involves active participation of the researcher within the context of the study. Researchers work collaboratively with stakeholders, such as practitioners or community members, to address issues and implement changes.
Collaborative Inquiry	It emphasizes collaboration and shared inquiry. It often involves a collective effort among researchers, practitioners, and other stakeholders to generate knowledge and enhance understanding.
Reflective Practice	Reflection is a key element of action research. Researchers continually reflect on their actions, observations, and experiences to gain insights and refine their approaches. This reflective practice is essential for the learning and improvement process.
Flexible Design	It is known for its flexible and adaptive design. Researchers can adjust their strategies based on ongoing observations and feedback, allowing for responsiveness to changing circumstances.
Qualitative and Quantitative Data	It typically employs a combination of qualitative and quantitative data collection methods. This mixed-methods approach provides a comprehensive understanding of the research problem.

of Action Research in SE.

- **Participant observation.** Is a qualitative research method commonly used in various fields, including SE, to gain an in-depth understanding of human behavior and social phenomena within a specific context. It involves the observer remaining in a social situation for the purpose of scientific research [103]. It has the *"unique strength of describing complex aspects of cognition, social interaction, and culture over time"* [58]. Table 2.4 highlights some other important aspects of this research method.

The main issues faced when using qualitative methods were the following:

1. Minimizing bias resulting from the personal viewpoints, preferences, and expectations of researchers. To address this challenge, with my tutors we defined several activities. I maintained a personal diary in which I documented issues and decisions. Transparency was pursued by discussing intentions and research objectives with other researchers and participants. Additionally, I sought guidance from recognized sources on each research method or technique used, particularly adhering to guidelines for reporting qualitative research.

2. Need for ethical treatment of human subjects. This involves taking care of several aspects. For example, treating participants and their opinions with appropriate respect, informing them and being transparent about the research objectives and methods used, informing them when collecting their data (e.g., indicating that a discussion will be recorded in audio), requesting their consent, and use mechanisms to anonymize sensitive or confidential data. Two interesting resources on this topic have recently been published: The ACM Publications Policy on Research Involving Human Participants and

Table 2.4: Several key attributes of Participant Observation [4]

Research Objective	Good for studying social issues or behaviors that remain unclear or insufficiently understood. Particularly relevant for capturing the direct perspectives of the individuals involved. The definition of the research objective is tied to the selection of the setting in which the investigation will take place. This setting could encompass various contexts within SE, such as a specific project team, a software company, or an open-source community.
Role of the Researcher	The researcher becomes an active member of the group being studied and may take on various roles, such as a developer, tester, project manager, or even an observer.
Data Collection	Participant observation do not stand as a data collection method on its own; instead, it represent the role an ethnographer assumes to facilitate the data collection process. Data collection techniques can be categorized into three areas: Observation (i.e., researchers perceiving and recording activities and interactions of individuals in the field setting), Interviewing, and Archival Research (i.e., analyzing materials that have been preserved for research). Triangulation (a technique explained above) is also used in this method.
Reliability, validity, and bias	Three critical aspects to consider are the following. - Reliability hinges on the consistency and accuracy of data collection methods and the ability to replicate observations consistently. - Validity is a crucial factor in determining the authenticity and truthfulness of research findings. It can be assessed through multiple means, including the involvement of multiple observers to cross-verify findings. - Bias Mitigation: Observer bias is a potential challenge in observational research, but it can be mitigated through specific approaches. Observational research often benefits from its natural and emergent nature, which reduces the risk of preconceived biases. Combining observational research with other research techniques can help triangulate findings and minimize the impact of any observer bias that may exist.

Subjects [3] and a Checklist for reporting the process for obtaining informed consent, achieving confidentiality and anonymity [4].

3. Organizing comprehensive, reliable data collection. This process entails the utilization of one or more data collection techniques based on the nature of the study. In this thesis, I employed self-administered questionnaires that included open and closed questions (S1-S2, S3, S4-S5), conducted focus group sessions (Study 3), recorded meetings with participants and the research team, and gathered artifacts and communications produced during the studies.

4. Data analysis. This involves organizing, synthesizing, and interpreting the data to answer the research questions. Several methods were used, although the most significant was thematic analysis. *Thematic analysis works by systematically examining qualitative data to identify recurring themes or patterns within the content. Researchers start by collecting or organizing data, then they code segments of the data, group related codes into themes, refine and name those themes, and provide supporting evidence from the data. The final step involves reporting the identified themes and offering insights into the data's meaning.* [14].

[3] https://www.acm.org/publications/policies/research-involving-human-participants-and-subjects
[4] https://www.e-informatyka.pl/index.php/einformatica/volumes/volume-2022/issue-1/article-9/

Throughout this research, various tools were employed to facilitate the analysis. Google Spreadsheets supported several data analyses (S1-S2), but for more intricate qualitative assessments, I turned to specialized software. In particular, saturateapp.com was used for the analysis of some data from S1-S2 and for analyzing the focus group transcripts (S3) and employed Atlas.ti[5] for managing the extensive dataset in the participant-observation study (S4-S5).

5. Reporting the analysis and its results in an appropriate manner (i.e., providing sufficient information for the readers to know what was done at a level suitable for reproducibility, including the actions taken to avoid personal biases) and reporting in a way that supports the interpretation of results (e.g., including quotes, summaries or snippets of related discussions).

6. Identifying the limitations of qualitative studies. This process entails a critical examination of the research, with the aim of identifying and transparently reporting any factors that could limit the results or their interpretations. It is particularly relevant to reflect on the role of researchers and their backgrounds, as these factors can exert significant influence on various aspects, most notably the analysis and interpretation of the data.

Some lessons learned that I highlight from the use of qualitative methods are:

- Working with human participants and companies requires flexibility. They will not have the same objectives as the researchers. Although it is necessary to define how to work with other stakeholders, in practice it will often be necessary to adapt to their requirements and restrictions.

- Whenever possible, plan for follow-up evaluations. In our research, it was useful to ask participants who had little prior experience with EBSE whether they subsequently found any of the concepts to be useful. It was also worth sending them a summary of the results for validation or clarification (this technique is called member checking [11]).

- In SE there is a lack of well-known standards for assessing the quality of qualitative studies. In this thesis, the most recognized guidelines for each research method or technique were used (e.g., [59, 60, 93, 94]). When there were no guides or references in SE, guides from other disciplines were considered. Of the latter I highlight the following resources: Thematic Analysis [14], member checking as a tool for improving trustworthiness [11], reporting qualitative research [77], qualitative research quality [109]. The last three proved highly valuable, particularly in enhancing the quality of our reports. However, the current page limits of journals and conferences in SE pose a challenge to reporting qualitative studies with the rigor that these guidelines recommend.

- I would note that qualitative research entails a steep learning curve. Assisting to The International Advanced School on Empirical Software Engineering (IASESE) in 2019, during which Carolyn Seaman delivered a session titled *"Observation as a Data Collection Technique for SE Research,"* significantly contributed to improving my understanding of qualitative research.

- Based on my experience, the tools that support qualitative analysis prove to be valuable. However, it is important to keep in mind that these tools are meant to aid the analysis and not substitute for it. They also come with varying learning curves and, in most cases, involve license costs. In our particular situation, the ability to collaborate

[5] https://atlasti.com/

among researchers was highly advantageous, as the versions of Google Spreadsheets and Saturateapp.com used were web-based and supported multiple users.

- Our engineering school lacks an ethics committee. Despite this, my colleagues and I made explicit efforts to adhere to best practices recommended in the literature regarding ethical issues. I acknowledge the importance of handling these concerns with care and responsibility. In discussions with other teachers at our school, I have found that they share similar opinions. I hope my research can help emphasize the importance of these matters and even contribute to the establishment of an ethics committee.

- Our experience highlights the value of employing focus groups in SE research. However, it is important to acknowledge certain peculiarities in their application in our field. In contrast to marketing focus groups, where participants typically represent different interest groups and are often unfamiliar with each other, SE researchers often have less control over participants. In many SE focus groups, it is preferred that participants either know each other or belong to specific groups, such as our focus group in a government agency. Another distinction lies in the focus of marketing groups on immediate reactions and opinions, considering questions like preferences for the ending of a film or the sales promotion video most likely to generate interest in a new product. Unlike marketing focus groups, SE focus groups may have an interest in their impact on participants, necessitating follow-up activities. Initial papers on focus groups in SE tended to be more exploratory, with little reflection on these differences [59, 60]. However, a more recent study includes one FG evaluating participants' perceptions of new method applications [36]. One of the focus groups was part of a longer observational study on the use of planning guidelines for simulation experiments. Despite the differences with their application in the marketing field, my research confirms that focus groups can be valuable in SE. In particular, integrating focus groups into larger studies offers certain advantages. For instance, in our study with the government agency applying FG not only provided a deeper understanding of participants' initial opinions but also established connections and mutual understanding. This, in turn, facilitated the effective use of a follow-up questionnaire some months later.

- In software engineering, assigning labels to qualitative studies can be challenging because many methods lack agreed-upon definitions, and sometimes multiple definitions with varying scopes exist. Furthermore, different researchers may define these labels differently. In my research, I also faced this challenge leading to the following observations:

  - Participant observation seems to be one of the most recently adopted methods (two notable studies are [58, 105]), lacking specific guidelines and not being included in ACM SIGSOFT Empirical Standards for Software Engineering [89]. Given the absence of clear definitions in SE and limited discussions on this method, one might question whether participant observation is a form of case study. Personally, I believe it is not, with the most significant distinction being that in participant observation, the researcher is also subject of study. This unique aspect alters the research design compared to a typical case study. For instance, participant observation involves collecting data that allows for a deep reflection on the role and participation of the researcher, such as maintaining a research diary documenting decisions, and also including in the report the researcher's position on the studied topic.
  - Recently, Wohlin proposed a new definition of a case study [117], stipulating that the researcher(s) should not take an active role in the case under investigation. This

restriction is introduced to clearly differentiate case studies from action research. In the former, the researcher has no active involvement in the phenomenon studied beyond observation and analysis. This characteristic can serve as a distinguishing feature between case studies and both action research and participant observation.

- To differentiate participant observation from action research, I understand the following: Participant observation primarily aims to gain an in-depth understanding of the phenomenon being studied, while action research focuses on creating positive change and improvement within a specific community or organization. In the former, maintaining a neutral stance regarding the project's outcome is necessary, while in the latter, a positive impact is anticipated and actively pursued.

- Finally, Lenberg et al. [63] suggest that SE qualitative research could benefit from utilizing a broad set of empirical methods, placing a stronger emphasis on reflexivity, and adhering to qualitative guidelines and quality criteria. Although this study was published after much of my research had already been conducted and was being reported, I have certainly adhered to these principles and also recommend them. Specifically, I recognize that reflexivity is an underemphasized aspect in SE qualitative research. Reflexivity involves the process by which researchers critically examine their own biases, assumptions, values, and experiences that may influence the research process and findings. In the studies included in this thesis, explicitly efforts were made to incorporate reflexivity. Specifically, I endeavored in all stages to be transparent about our intentions and personal backgrounds regarding the objects of study. My colleagues and I have openly discussed the limitations associated with our roles as researchers and have maintained an ethical research process, including considering the relationship between participants and researchers. Finally, I have kept a chronological and personal research diary and explicitly considered the ethical aspects of all interactions with participants.

2.2.3. Managing a Research Team

The collaboration with stakeholders involved three distinct EBSE course modules for undergraduate students, a workshop with government agency staff, and an EBSE project conducted in partnership with an industry collaborator. Each required the active participation of a research team as well as me as the main researcher. I was responsible for organizing the team members, both to do their required tasks in each project and to support the research effort (e.g., data collection, and debriefing meetings). Although the most significant effort was made in studies S4-S5 all studies required organizing the research in collaboration with a research team.

In general, the issue of management is relevant for any project that involves large-scale empirical research beyond the scope of a single researcher. However, in qualitative research, there are some specific issues.

First, the following opportunities offered by collaborating with other researchers are identified:

- Qualitative studies often necessitate on-the-fly decision-making since they are frequently exploratory, and unforeseen events or opportunities for enhancing the research may arise during their execution. Having multiple researchers enables us to deliberate on these decisions before implementing them.

- There are several aspects of qualitative research methods that remain ill-defined within our community and the involvement of multiple researchers allows for a more objective assessment of the study's limitations throughout the research process.

- Qualitative research entails interpreting results, which can be influenced by various factors, including the contextual elements, education, and experience of the researcher, as well as their stance regarding the investigated topic. Collaborative teamwork reduces reliance on the interpretation of data by a single individual. This not only helps improve or rectify the interpretation of data but also ensures that the results are reported accurately, facilitating comprehension by other researchers and practitioners.

Secondly, I also recognize some potential challenges:

- Motivating other researchers to participate in the research can often be a daunting task, especially when involving researchers who also have industry commitments. It proved immensely helpful to clearly define and agree upon their specific roles and responsibilities in advance as a condition for their participation.
- Managing the assigning and scheduling of tasks among multiple participants was another significant challenge. As previously mentioned, flexibility was essential, and research plans had to be adapted to accommodate the availability of other researchers and participants.
- While other researchers may be interested in collaboration, especially when their contributions are acknowledged, becoming a co-author can introduce conflicts between maintaining objectivity and the desire to publish. In such cases, I recommend making concerted efforts to uphold research rigor and encourage reflection on each researcher's position.
- The validation of the research, its findings, and the reporting process by other researchers can help mitigate researcher bias. However, in other fields, there is some controversy surrounding how other researchers should be integrated into the validation of qualitative synthesis. While some suggest involving different researchers in data coding and using measures like intercoder reliability [78], others, such as Braun and Clark, the proponents of the thematic analysis method, argue that reliability is an inappropriate criterion for assessing qualitative work [15,26]. They argue that interpretation is invariably influenced by the researcher's epistemological approach, their position, and their experience. Additionally, they argue that the use of quantitative measures can pose challenges; for instance, it could lead to a more extroverted researcher or one with a hierarchical relationship to convincing others of their arguments. Some take a more categorical stance, such as Stenbacka [110], who asserts that 'reliability has no relevance in qualitative research, as it is impossible to distinguish between the researcher and the method.' Personally, I believe it is essential to introduce this discussion within the SE community, in which qualitative research predominantly relies on a positivist epistemological approach that often oversimplifies both the issues under investigation and the role of the researcher.

Chapter 3

Summary of Research

This chapter presents a summary of the studies included in this thesis, outlining their objectives, methods employed, key findings, and reflections.

3.1. Overview

As mentioned above, the purpose of this thesis is to contribute to the understanding and potential improvement of EBSE adoption. It has been suggested by several authors that EBSE could help bridge the gap between industry and academia [29, 62]. Extensive research exists that aims to enhance EBSE methods and techniques (see for example [5, 37, 52, 57]), alongside efforts to find mechanisms for improving the transfer of evidence to industry, such as rapid reviews and evidence briefings [21, 23].

However, there seems to be a gap in the literature, as no studies directly address the question of whether non-academic adoption of EBSE is worthwhile. In this thesis, this gap is addressed by presenting studies that explore the issue from three different angles: the role of training as a facilitator of EBSE adoption and its effect on trainees' further working practices, the significance of EBSE in government agencies, and the utilization of EBSE by the software industry.

Figure 3.1 provides an overview of this thesis and its constituent studies. The green boxes represent the areas investigated, while the yellow boxes denote the empirical studies conducted. By delving into these topics, I aim to shed light on the practical value and significance of non-academic adoption of EBSE.

Below is a summary of the studies conducted in this thesis.

3.2. EBSE Training

Other disciplines, in which EBP is implemented, have recognized the critical relevance of appropriate training. Several systematic reviews have highlighted the lack of knowledge and skills as one of the most commonly reported barriers to the adoption of EBP [96,104,115,118]. In fields like medicine, EBP training has a higher level of formalization and consensus, with proposals for core skills in EBP (see for example [3]).

Since EBSE's introduction, some studies reporting teaching initiatives have been published (see for example [6, 16, 24, 43, 46, 53, 74]). However, there has been a lack of a comprehensive review that encompasses all the existing work in this area. Furthermore, no one

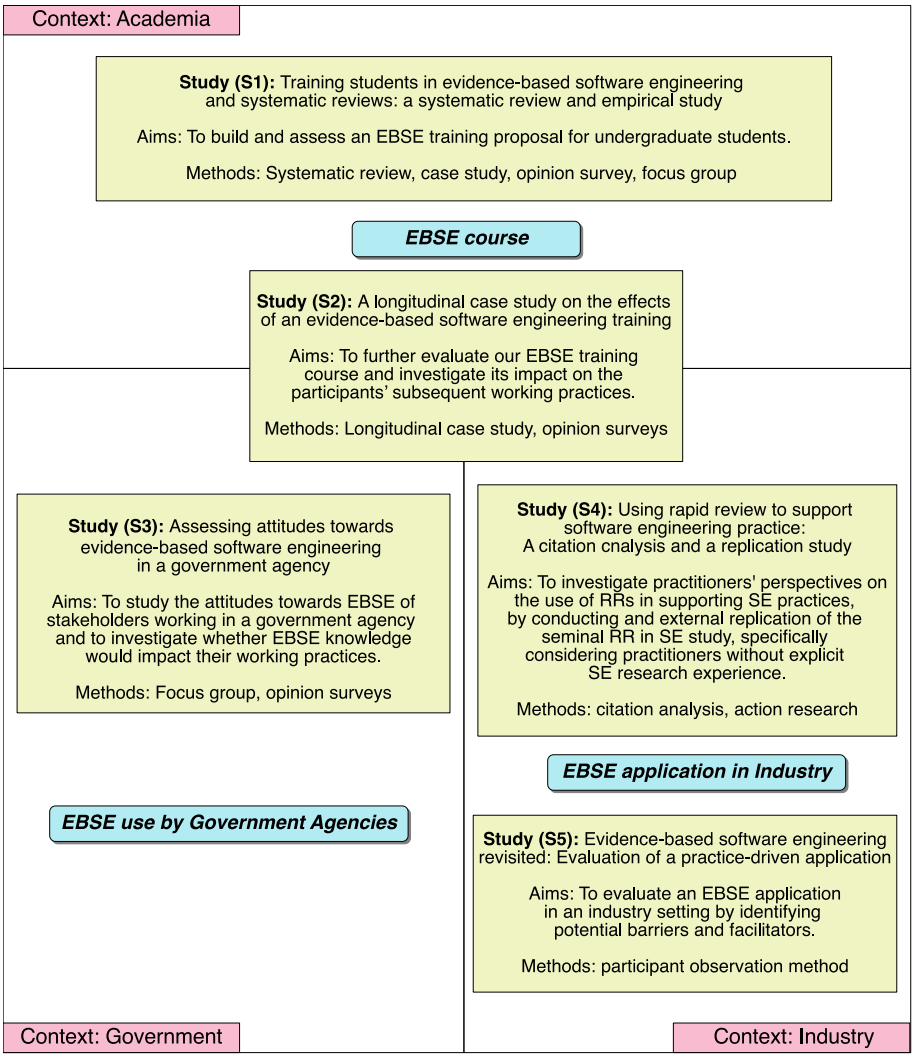


Figure 3.1: Overview of the thesis.

has yet presented a detailed proposal for teaching EBSE, which could serve as a foundation for core skills.

In S1, the initial study on EBSE training (presented in Appendix A), I tried to address these two gaps by (1) reviewing previous publications on EBSE teaching, (2) proposing an undergraduate EBSE course, and (3) conducting an empirical evaluation of the course.

3.2.1. SR on EBSE Training

Firstly, an SR was conducted to examine the previously reported EBSE teaching initiatives thoroughly. The aim was to gain a comprehensive understanding of these initiatives. The research questions addressed through this SR were as follows:

1. Which EBSE teaching initiatives have been reported?
2. In what context (academic program/courses/etc.) is it taught?
3. What is the content taught and what are the methodologies used to teach it?

4. What are the assessment tools used?
5. What are the difficulties found and what are the recommendations provided?
6. What are the benefits for students?

Fourteen reports of EBSE and SRs training experiences were found, with half of the studies focusing on teaching EBSE and the rest examining the EBSE process or attitudes towards it. Although our SR was conducted in 2017 and last updated in December 2019, all the studies found were conducted before 2014. Furthermore, the studies were conducted by universities in seven countries, primarily the UK. The training initiatives involved postgraduate and undergraduate students across various program areas.

The most common educational approach included a brief introduction followed by a practical assignment, although other methods were also used. None of the studies mentioned specific educational theories guiding their approach. All initiatives incorporated a practical assignment, such as participating in a secondary study (SR, limited SR, or mapping study) or performing EBSE steps. The assignments were conducted individually, in groups, or as a whole class, with some studies limiting the scope or providing guidance. However, none of the studies reported using any pedagogical method to support their training method.

Evaluation methods varied and included marking student reports, teacher evaluation of outcomes, and student questionnaires. No specific method appeared to be inherently better.

Common issues reported by multiple studies included the time and effort required, students' ability to conduct SRs or mapping studies, challenges in literature searching, the value of an iterative approach, the benefits of working on team projects, and the importance of focused research questions.

The most common benefits reported were learning literature search and organization skills, gaining knowledge about empirical studies, and developing the ability to assess information on a particular topic.

3.2.2. EBSE and SRs Course for Undergraduate Students

Secondly, my colleagues and I developed an EBSE course, with emphasis on the SR process. I adopted the learning outcome approach as a specific pedagogical method based on codifying the knowledge and skill required of future EBSE users [51, 102].

The course focuses on teaching fundamental EBSE concepts and techniques for practical use. By the end of the course, the goals included students having a comprehensive understanding of essential EBSE concepts, the ability to identify SE issues addressable through evidence, the skills to assess published secondary studies, and active participation in the conduct of SRs.

Learning outcomes (LOs) were used to guide the course design and assessment. In summary, I established over fifty LOs that formed the basis of a syllabus focused on promoting the practical application of EBSE. The course covered topics such as scientific publications, the evidence-based paradigm, SRs in SE, and the process of conducting an SR.

The course followed a structure of alternating theoretical and practical content, with weekly monitoring of students' team assignments focused on conducting a secondary study. To address challenges and recommendations from previous EBSE training studies, I incorporated principles like limiting the workload during certain stages of the EBSE process, providing teacher guidance in topic selection, and offering support to students requiring iterations in their team assignments.

3.2.3. Evaluation of our EBSE and SRs Course

Thirdly, I conducted a case study to deliver and evaluate our EBSE course, considering students' performance and opinions as key factors. The case study's primary goal was to assess whether students achieved the intended learning outcomes of the course, while also gathering insights into any challenges they encountered with the course content and structure. Additionally, the objective was to understand students' perceptions of the method and materials used and their opinions regarding the relevance and value of the acquired knowledge. Hence, the case study was driven by the following research questions:

1. Does our training proposal enable undergraduate students to explain EBSE concepts and contribute to the conduct of an SR?
2. How suitable were the method and materials used according to the students' perception?
3. What difficulties do students observe?
4. What benefits do students observe?

To address these research questions, I collected both quantitative and qualitative data, employing data triangulation to include multiple perspectives. The data encompassed students' opinions gathered through a survey and a focus group, as well as their marks in course tests and teams' SR projects. Additionally, insights from a teachers' debriefing meeting were incorporated to present another viewpoint. The survey and focus group were conducted during the final class of the course, which was organized in stages. Initially, teachers provided detailed feedback to each team on their work, including the grade achieved, comments, and assessment criteria for each LO. Subsequently, students participated in an individual test.

The study results suggested that the EBSE teaching proposal effectively prepares students with more than 3 years of Computer Science/Software Engineering training at university level to participate in the conduct of secondary studies. While the students did not complete a full SR during the course, they demonstrated the acquired skills to perform various activities within the process. Their performance in practical assignments and their opinion on the achievement of the objectives aligned well, reflecting the acquired skills. However, the evaluation of individual tests showed only a minor grasp of EBSE theoretical elements. This outcome is consistent with the course approach, which emphasizes practical training in SR activities rather than theoretical concepts.

The teaching method, featuring a high practical workload and an alternating introduction of theoretical and practical content, proves to be effective. Students faced difficulties with the textbook, likely due to its English language and technical content. Undertaking an SR was challenging for half of the students, and teachers provided support to manage frustration.

In the initial course experience, I identified common issues in previous EBSE training initiatives. Conducting an SR requires significant effort, but students can succeed with an iterative teaching approach and guidance from teachers. Learning the conduct of secondary studies as a team project is valuable, but research questions should align with available research.

Training provided students with a different perspective, though the actual benefits are challenging to assess. Some students see potential professional benefits, while others focus on academic gains. Two former students conducted a limited SR for their capstone project that sought to create a prototype to help patients with the freezing of gait in Parkinson's disease, giving preliminary confirmation of the positive impact of EBSE training.

3.2.4. Impact of our EBSE Training

In the subsequent study S2 (presented in Appendix A), I conducted a further evaluation of our training method. The study aimed to confirm the adequacy of our training and evaluate

its impact on the attitudes and behaviors of the trainees, particularly in their work practices, over different time periods. The investigation was driven by the following research questions:

1. Is our EBSE course adequate to train undergraduate students?
2. Does our EBSE course have any impact on the working practices of the students?

Following the initial course, the EBSE course was delivered two more times and used the same evaluation process, consisting of surveys, focus groups, and teacher assessments. To analyze the training's impact, I conducted two surveys with the students who took our course, one after 7 months after finishing the third course, and a second after 21 months.

In the three courses, a total 44 students took our training course. The course received positive feedback from both the teachers, who assessed the students' teamwork assignments, and the students themselves. At the end of each course and after applying the knowledge gained in practice, the students considered the training to be satisfactory. Notably, all 33 students who reported applying what they learned, acknowledged the training was suitable for their needs and requirements.

The results confirm that EBSE training has enabled more than half of the students to improve their work practices by enhancing their information-gathering and analysis skills and increasing awareness of evidence and research. The findings suggest, for the first time, that training in EBSE enhances practitioners' confidence in the significance of evidence from research and encourages its utilization for informed decision-making. Additionally, the study suggests that the benefits of teaching EBSE are comparable to those observed in other disciplines that teach EBP. Furthermore, the findings indicate that EBSE training positively impacts the performance of novice industry practitioners, not just individuals in traditional decision-making roles such as senior software engineers or project and quality managers.

Key Findings

- The systematic review found no defined formalized EBSE training program.
- The EBSE course was based on current educational theory with 50 learning outcomes and a strong practical component.
- The course was run three times, training a total of 44 undergraduate students.
- Teachers and students judged the course to be acceptable.
- 33 students reported training positively impacted their work practices.
- English and Spanish versions of the course are available [82].

3.3. EBSE Use by a Government Agency

Government agencies and regulatory bodies are key players in the adoption of evidence-based practice, since they support the conduct of SRs, promote good practices, and define quality levels. So far there is no research on EBSE and stakeholders from these types of organizations.

Thus, in study S3 (presented in Appendix B), the aim was to investigate the attitudes of stakeholders working in a government agency (GA) towards EBSE and assess the impact of EBSE knowledge on their working practices. The research questions addressed were the following:

1. Is EBSE a method that should be considered for adoption by GA stakeholders?
2. Does EBSE awareness have any impact on the working practices of GA stakeholders?

To address these questions, I conducted a multi-stage field investigation within AGESIC, a Uruguayan national GA responsible for digital policies. AGESIC's involvement in policy-making and legislation in SE is of great importance. They support government initiatives, define software requirements, and influence purchase conditions. This study aims to contribute to AGESIC's understanding of EBSE and assess its potential usefulness in achieving its goals.

In the first stage, I organized an EBSE awareness lecture in the agency and collected through a focus group participants' perceptions about the value and limitations of EBSE. Sixteen months later, in the second stage, I reached out to the agency and inquired whether the participants had applied the EBSE information presented to them.

Initially, participants expressed that EBSE appeared useful for addressing challenging problems, particularly considering the agency's responsibilities. Perceived barriers to adopting EBSE included the need for institutional support, the lack of government practice reports, insufficient skills or motivation, the cost associated with conducting SRs, and the scarcity of evidence on emerging issues.

In the follow-up survey, although the participants were not conducting SRs themselves, many reported improvements in their information search and evaluation methods to support their work. These findings confirm the potential value of evidence within the context of IT regulatory and government bodies.

Participants indicated using a diverse range of literature, with a particular emphasis on grey literature, including government reports and white papers. This finding suggests the need to tailor the presentation of EBSE concepts and techniques in future similar studies. For instance, incorporating quality assessment checklists specifically designed for grey literature, such as the one proposed by Garousi et al. [38], could be beneficial.

Key Findings

- EBSE was perceived useful for tackling challenging problems and appropriate given agency's role.
- Identified barriers include lack of: institutional support, evidence about government practices, and research skills.
- Participants' information literacy skills improved with knowledge of EBSE.
- Barriers to adopting EBSE align with those observed in other fields.

3.4. EBSE Application in Industry

In the final part of this thesis, I study the application of EBSE in industry, with a particular focus on RRs, which has been promoted as a type of secondary study explicitly created for direct collaboration with practitioners.

Specifically, I led a review team that conducted an EBSE-based project in collaboration with a software company to address industry-specific issues. I worked closely with practitioners to diagnose the problem at hand, employed an RR to gather relevant evidence, and ultimately transferred the findings to the company. This project resulted in two interconnected yet distinct empirical studies and a literature analysis on RRs in SE.

3.4.1. Rationale

Two factors influenced the final part of my research program:

- Given the current trend where the limited number of practice-driven SRs conducted are primarily in the form of RRs, I find it compelling to investigate more about this method. This was the rationale of the next study (S4) which involved: (1) studying previously conducted RRs in SE and (2) conducting an external replication of the first study that proposed the use of RRs in SE [21].
- Only a small number of SRs incorporate non-academic stakeholders and include additional EBSE activities. For instance, out of the 169 SRs identified by Kamei et al. [48] published in 2011, 2014, and 2018, I found only three SRs that were jointly conducted with industry [20,47,107]. While some practice-driven SRs have addressed practical problems, provided evidence to practitioners, and evaluated the benefits of EBSE adoption (e.g., [12, 21, 109]), none of them comprehensively examined the overall application of EBSE or explored the challenges associated with its implementation. Thus, the project was utilized as an opportunity to evaluate the broader EBSE framework, aiming to identify challenges, barriers, and facilitators encountered during its application (S5).

Table 3.1: Comparison of steps of EBSE & Action Research

EBSE Steps	Action Research
1. Converting the need for information into an answerable question.	1. Diagnosis
2. Finding the best evidence with which to answer that question.	2. Planning
3. Appraising evidence validity, impact, and applicability.	3. Intervention
4. Integrating the appraised evidence with expertise and stakeholders' values and circumstances.	
5. Evaluating effectiveness and efficiency in executing previous steps and seeking ways to improve.	4. Evaluation
	5. Reflection

3.4.2. Methodological Complexities

In the context of S4, I employed Action Research, replicating the same research method as in the original study. The action taken was the conduct of an RR to provide evidence to practitioners aimed at addressing their problems. However, from the company's perspective, their involvement was in an EBSE project (i.e., an application of all EBSE steps). EBSE can be defined as a form of action research, where the goal is to adopt evidence-based actions to address a problem (refer to Table 3.1 for comparison of the steps of both). This consideration allows us to utilize the Participant Observation method to evaluate the project as an example of EBSE. This assessment is presented in S5.

It was necessary to reconcile the application of both these methods. As mentioned in Section 2.2.2, Participant Observation seeks to understand the phenomenon being studied, while Action Research focuses on creating positive change. In this case, the aim was to assist the company while also conducting a fair evaluation of the EBSE application. To reconcile these objectives, the project focused on applying EBSE steps as they are defined. This meant explaining in advance to practitioners that our objective, in addition to trying to assist them,

was to evaluate EBSE and that the suggestions for process improvements would be grounded on evidence from research collected through an RR. I was also clarified several times to the participants that both their positive and negative feedback was useful. Furthermore, a detailed record of the activities undertaken and any deviations (such as requests for clarifications or examples for the recommended actions) was maintained. I personally maintained a research journal, and my colleagues and I also used reflexivity on multiple occasions, including meticulously analyzing the studies' limitations, among other aspects.

The two components of the S4 study (i.e., the systematic review of published SE RRs and the replication of the original study proposing RRs for SE) and the S5 study involving the evaluation of EBSE in an industry application are introduced in greater detail below.

3.4.3. A systematic Review

of SE Research Using RRs

In order to assess take-up of RRs in SE, I lead a systematic review of RR research based on citation analysis of two of Cartaxo's papers [21, 22]. The research questions were:

1. What is the extent of take-up of RRs in the SE domain?
2. What was the scope of these studies?
3. What are the methodological characteristics of the reported RRs?
4. Which studies contributed to assessing the value of RRs and what have they found?

We found 23 papers reporting RRs in SE, although one of them analyzed the process used by two previously published RRs [91]. These studies involved researchers from 13 different countries, and more than half of them were published in the first 11 months of 2023.

Of these, 13 studies focused on RRs aimed at acquiring knowledge within specific fields, while nine studies conducted RRs and supplemented or compared their findings with stakeholders' opinions. The majority of RRs primarily relied on white literature sources.

In many instances, adequately understanding the RR process or its outcomes proved challenging due to insufficiently reported information. Only two studies provided comprehensive details crucial for reproducibility, such as the search date, primary study list, and synthesis methods employed. In addition, only five explicitly mentioned utilizing a protocol to guide the process.

Six studies validated the value of RRs by corroborating their findings (or derived artifacts) through collaboration with stakeholders external to the review team. Stakeholders across all studies exhibited a positive attitude towards the results. With the exception of one study involving undergraduates, the remaining studies enlisted practitioners with at least a degree-level educational background.

In three studies, researchers collaborated with companies throughout the RR process, engaging technology experts specializing in topics relevant to the RR questions. This research confirmed that RRs can support collaborations between industry and academia, but, like the RR conducted by [21], the industry collaboration was in the context of advanced R&D projects and involved practitioners with experience in SE research.

3.4.4. Industry Practitioners' Perceptions of RR Evidence

The mapping study confirmed that there was little research aimed at increasing the understanding of RRs in SE. Thus, I considered it important to replicate the original study of RRs in SE [21] and examine their utility in supporting SE practice within a software company. The specific focus was to assess the significance of the evidence derived from an RR for practitioners lacking familiarity with SE research, which represents a novel aspect compared

to prior research. In study S4 (presented in Appendix C), the investigation was guided by the following research question:

1. What are the perceptions and attitudes about using a rapid review to support software engineering practice in a software company?
2. Are there any problems using RR information when collaborator have no SE research experience?

To investigate this, like Cartaxo [21], I adopted action research, a method that combines research with proactive interventions to drive change. I led a small review team that conducted an RR for an agile software development team seeking to enhance their knowledge management practices. Employing action research, the RR conduct, collaboration with practitioners, and their feedback were analyzed. Comparing the results with previous RR evaluations in SE, I collected data on the organization, participant roles, actions taken, and outcomes of the action research process. These data allowed my colleagues and I to assess the reception of the RR recommendations by the practitioners.

Participants generally viewed RR results positively, considering them more reliable than other sources and effective for addressing identified problems. Despite our requesters lacking academic backgrounds and SE research experience, the findings align with Cartaxo's study where participants were from an applied research institute with master's degrees [21]. This underscores that RR's value extends beyond research-background practitioners, emphasizing its usefulness in bridging academia and industry collaboration.

Collaborating in the RR process also improved practitioners' problem understanding and knowledge. Additionally, they expressed appreciation for timely results. Subsequently, some months later, they reported implementing some of the recommendations.

Following Cartaxo et al.'s recommendation, result discussions were incorporated into dissemination activities. In addition to an evidence briefing, a workshop was conducted where attendees engaged in discussing the evidence. This approach effectively disseminated our findings. Both studies suggest that a single-page evidence briefing is insufficient for the practical utilization of results.

The main challenges faced by the review team were the lack of guidelines and examples of using an RR in industry and the difficulty in finding adequate evidence. Presently, the former challenge appears less problematic due to the growing number of RR publications. However, the latter issue suggests that the absence of explicit recommendations and the necessity for qualitative aggregation might jeopardize the rigor of the RR process in SE, particularly when only one researcher is typically involved.

Finally, I found some aspects that deserve reflection about the RR process.

- Although none of the RRs that I surveyed included *quality assessment* of the primary studies (refer to Section 3.4.3), my experience indicates that it is not advisable to rule out this activity before assessing whether this information is required.

- While RR offer a valuable tool for minimizing effort and timescales, I recommend researchers *utilize secondary studies if relevant ones are available*. A high-quality and pertinent SR or mapping study may resolve requesters' issues directly or identify relevant primary studies.

- *Reporting an RR.* Published RRs reports fail to provide necessary information for a thorough understanding of the RR process or its outcomes (refer to Section 3.4.3). However, without *comprehensive reporting of the RR process*, the value of the results to other researchers or practitioners will be limited, regardless of their significance.

Key Findings

- More RRs are being reported in SE, with some conducted collaboratively with practitioners knowledgeable in SE research.
- Reporting practice for RRs is inadequate, reducing the scientific value of their results.
- The replication study addressed a limitation of the original study of RR in SE by collaborating with participants lacking SE research experience.
- Practitioners without SE research experience valued and utilized RR results.
- Researchers and practitioners considered that RR provided more reliable content than other sources.

3.4.5. Evaluation of a Practice-driven Application of EBSE

Despite the widespread adoption of SRs for academic purposes, the broader framework of EBSE has not been thoroughly evaluated. In study S5 (presented in Appendix C), I evaluated the application of EBSE in an industry setting, which included collaborating with practitioners to diagnose the problem, collect evidence through an RR, and transfer the results to the company.

The following research question guided the investigation:

1. What issues, barriers, and facilitators arise when using EBSE in an industry setting?

To answer it, and as a means of assessing the value of the EBSE framework, I intended to identify potential barriers and facilitators to EBSE use, reflecting the viewpoints of both the EBSE team members and the company staff members. My evaluation considered the five steps of the EBSE process in terms of how the steps were performed and any difficulties that were encountered. Specifically, I conducted a thorough examination EBSE application using participant observation and qualitative data analysis of meeting recordings, correspondence, and personal notes. Furthermore, the findings were compared with early concerns raised by the researchers who initially proposed EBSE two decades ago [33, 56].

Practitioners used some of the evidence-based recommendations obtained. However, several barriers to the practice-driven use of EBSE were encountered .

- *Difficulty in obtaining relevant evidence.* I struggled to find evidence that directly addressed the company's specific problems. Existing evidence was either too general or non-existent, requiring us to formulate a broad research question.
- *Complexity of applying evidence.* Applying the evidence I found was challenging because it consisted primarily of high-level recommendations rather than actionable guidelines.
- *Lack of guidelines and examples of using EBSE in industry.* This lack of guidance hindered our EBSE process, and I had to adapt my approach based on the project's circumstances. For example, I adjusted our selection criteria based on discussions with requesters after they validated some of the evidence from the studies obtained in the search stage.

These issues highlight the need for more guidance in the EBSE framework, particularly in addressing broad questions, effectively reporting SE evidence to practitioners, developing practical and context-appropriate recommendations, and evaluating EBSE activities.

Factors that contributed to the success of my project included:

- *Close collaboration and addressing industry partner's needs.*
- *Comprehensive dissemination strategies.* In my case, holding a knowledge transfer workshop in which participants carried out hands-on activities to discuss the recommendations and their implementation was very valuable.
- *The company's commitment to process improvement* facilitated their openness to reflection and acceptance of recommendations.
- *The use of an RR provided valuable results within the resources constraints without compromising scientific rigor.*
- *External expert validation played a crucial role in improving the study's quality* during the RR process and results analysis and reporting.

Despite the recognition of barriers to EBSE for over 20 years, my study confirmed that many of these barriers persist. The lack or inadequacy of evidence remains a significant challenge in adopting EBSE, as identified in previous research [17, 82, 109]. Collaboration between academia and industry also remains a challenge for the community, indicating that EBSE has not fully achieved its goal of bridging the gap between industry and academia.

On a positive note, my study reports the effective utilization of EBSE for enhancing industry-academia collaboration, even when practitioners lack of SE research experience. Nonetheless, applying EBSE presents various hurdles, requiring a diverse skill set that includes academic knowledge and professional experience.

The results suggest that, for effective adoption of EBSE, it is likely to be necessary to use relatively high-level questions, due to limitations of the available research, and we should provide improved guidelines for EBSE steps 2 to 4. Additionally, defining procedures in advance to monitor and evaluate the process changes specified in Step 4 is crucial to effectively completing step 5.

Key Findings

- Limitations of SE evidence led to using broad questions and tailoring results to specific recommendations.
- Implementing EBSE process demanded a combination of professional and research skills.
- Key enablers included close collaboration, effective dissemination, and practitioners' commitment to process improvement.

Chapter 4

Discussion

This section provides a summary of the main findings across all studies, reflects on the applicability and value of evidence-based practice applied to SE, and analyzes the potential role that EBSE can play in relation to the ongoing global changes.

4.1. Key Findings on EBSE Adoption Across Studies

Going beyond the precise aims of each study, they can be considered as interventions wherein participants were introduced to the principles of EBSE through different approaches. To illustrate, in the course, students were immersed in both the theoretical concepts and practical techniques of EBSE (S1 & S2). Likewise, in the study involving the government agency several IT area members were introduced to EBSE concepts through a lecture (S3). Finally, during the EBSE-based project researchers collaborated with practitioners to conduct RR and participate in knowledge dissemination activities (S4 & S5). Considering this viewpoint, below I comparatively interpret the results seeking to identify patterns in the benefits of knowing/using EBSE, perceived/experienced barriers to its adoption, and factors that support its use.

4.1.1. Benefits of Knowing or Adopting EBSE

The results of my research suggest that gaining an understanding of EBSE allows participants to improve information literacy skills (S2 and S3). This can be attributed to several reasons, but it could be probable that after becoming familiar with EBSE, participants gain an awareness that it is both possible and often advisable to consider the quality of the information underlying decision-making in SE, as well as the level of rigor applied in its generation.

Both industry and government practitioners considered the evidence from the SRs to be more reliable than the sources of information they usually used (S3 and S4). In addition, industry practitioners felt that the RR evidence was problem-oriented and that there would be benefits to its application (S4).

Finally, participating in an EBSE-based project seems to have improved the problem understanding of the staff of the industry partner and allowed them to learn new concepts (S4).

4.1.2. Barriers to EBSE Adoption

One of the most challenging barriers faced in the project with the industry partner (S5) was the lack or inadequacy of evidence (S5), which was also perceived as a barrier by the members of the government agency (S3).

Another related issue is the current incompleteness of the reports of SE RRs (S4). The lack of information about the RR process makes it difficult to assess the evidence generated by RRs and under what circumstances it can be applied.

The high cost (in time and effort) of conducting SRs was identified as a barrier by the teachers of the EBSE course (S1) and was perceived as a barrier for the members of the government agency (S3). Furthermore, having short response times was valued as important by the industry partner (S4).

Other difficulties experienced when working with the industry partner (S5) were the complexity of applying the evidence, the lack of guidelines and examples of use of EBSE, and practical problems of working with an industry partner.

Finally, the government agency staff perceived as obstacles the lack of skills and motivation to use EBSE as well as the possible lack of institutional support (S3).

4.1.3. Facilitators to EBSE Adoption

First, my research about the effects of the EBSE training suggests that training significantly facilitates the adoption of EBSE practices (S2). Undoubtedly, possessing sufficient training represents the optimal approach for overcoming the skill gap barrier. The training proposal has predominantly targeted undergraduate students[1], so I would like to investigate training approaches for industry and government practitioners in the future[2].

Second, several factors contributed to the successful achievement of the EBSE-based project (S4, S5). These factors included: (1) Establishing close collaboration with the industry partner and considering their needs throughout the project. (2) Considering an appropriate disseminating of the evidence to facilitate its practical application, a non-trivial task. (3) Utilizing an RR appears particularly appropriate when collaborating with an industry partner due to its flexibility and the capacity to generate results in low-resource settings (in the study of this thesis, limited effort and use by non-experts) while maintaining scientific rigor. (4) The commitment of the industry partner to process improvement, and (5) the involvement of external researchers who validated the progress and rigor, were also noteworthy.

Finally, the brief introduction of EBSE in a government agency (S3) suggests that employing knowledge brokers might aid in introducing EBSE to audiences unfamiliar with scientific literature usage. Furthermore, observing the agency members' substantial reliance on grey material leads to believe that they could benefit from guidance or recommendations on effectively utilizing this type of material (matters that have been previously considered within EBSE).

4.2. Practicality and Worthiness of EBSE

All the studies on EBSE adoption of this thesis show the same consistent pattern of main barriers to evidence-based practice as identified in secondary studies across various disciplines, including medicine. These barriers encompass a lack of relevant research in the literature,

[1] Some postgraduate students participated in the EBSE courses, but their experience was not empirically evaluated.

[2] The intervention with AGESIC could be considered as an introduction to EBSE rather than formal training.

insufficient research skills among stakeholders, and limited time/resources available to stakeholders. Despite the apparent significance of these barriers, there remains a question: why are researchers convinced that evidence-based practice is effective? This conviction appears to stem from the success of Evidence-Based Medicine (EBM).

In the view of my colleagues and me, EBM is deemed successful primarily because it is highly effective in drug testing, owing to the following factors:

- *Evaluation Based on Comparative Studies:* The efficacy of EBM is grounded in comparative studies, with drug trials typically adopting well-understood methodologies such as randomized controlled experiments, often incorporating double (and preferably triple) blinding. These trials are meticulously controlled formal field experiments with established protocols.
- *Passive Recipients of Treatment:* The subjects in these studies are typically passive recipients of the treatment.
- *Passive Role of Doctors:* While doctors are required to diagnose the cause of a patient's illness, they usually act as passive recipients of the results from comparative studies, guiding them on one treatment over another. In the majority of cases, little to no additional training is necessary for them to administer a new drug treatment.
- *Medical Ethics:* Medical ethics dictate that doctors must "do no harm", creating an inherent incentive for proven treatments to be adopted. Additionally, doctors often need to make a formal ethical statement before being licensed to practice.
- *Regulation as a Key Enabler:* The major contributing factor to the success of EBM in drug testing appears to be stringent regulation. Pharmaceutical companies are required to conduct extensive trials before a treatment is approved. Moreover, in many countries, government agencies produce mandatory treatment guidelines, further ensuring adherence to standardized and proven practices.

In the field of software engineering, several challenges shape the research landscape:

- *Conducting Field Studies is Challenging:* Conducting formal field experiments is virtually impossible, and achieving blinding is often infeasible in software engineering contexts. Moreover, comparing two well-known and well-understood methods is the exception rather than the norm. Thus, SE faces hurdles in conducting field studies, with limited industry motivation or established practices for thoroughly evaluating new methods or technologies before implementation. Consequently, major technological changes can lead to unforeseen side effects. For instance, the adoption of object-oriented programming unintentionally resulted in increasing code duplication, which is identified as one of the major sources of defects [97]. Another example is the oversight or deliberate neglect of knowledge management practices in agile software development methodologies. Such oversights can lead to deviations from the company's strategy, wastage of resources, and the acquisition of irrelevant knowledge [76]. The use of artificial intelligence as the foundation for self-driving cars is also under scrutiny, the concerns include at least 25 deaths and issues such as several traffic jams [27].
- *Active Involvement of Subjects:* Subjects are rarely passive recipients of improved methodologies, especially since human-based problem-solving methods are prevalent in SE. Even if the solution is simply a new tool, it often involves a process of product or process migration, along with the need for personnel training. When applying EBSE this becomes even more complicated. The need to use broad questions arises with

various high-level recommendations. Many industry practitioners, apart from large software engineering companies with extensive R&D departments, lack the knowledge to choose the most suitable approaches for their circumstances. Furthermore, they may lack the experience required to implement necessary changes to their processes. Practitioners need to understand the problems SRs aim to address, considering symptoms and context. They also need to know the solutions that address these problems, the context in which these solutions are effective, and how to implement them successfully, which may involve training, adopting new tools, and incorporating new QA practices. Given these complexities, the direct utilization of SRs results by practitioners remains the exception.

- *Ethical Statements Not Required:* Unlike some other fields, software engineers are not required to make any ethical statements before employment.
- *Lack of Regulation:* There is a lack of formal regulation for software products and software production practices.

Considering these circumstances, is EBSE worthwhile? Several ongoing societal changes suggest EBSE could become more important. These global shifts are linked to the widespread dissemination of unrealistic information, the rise of generative artificial intelligence, and the consequential transformations. In the following section, an overview of these challenges is provided, along with a discussion of the potential role of EBSE and empirical SE research.

4.3. Current Global Challenges and the Role of EBSE

The Information Age is the historical period that began in the mid-20th century in which societies are profoundly influenced by technological advancements, particularly in information and communication technologies. Castells, a prominent researcher in the fields of sociology, communication, and technology, underscores the pivotal role of information in this era, emphasizing its significance not merely as data but as a crucial resource [25]. When information is processed and transformed into knowledge, it becomes a driving force shaping economic, social, and cultural structures.

Previously, information and knowledge were scarce; now, they are abundant. However, their abundance does not shield them from manipulation and distortion. Information pollution refers to *"false, misleading, and manipulated online and offline content, created, produced, and disseminated intentionally or unintentionally, with the potential to cause societal or physical harm"* [114]. Manifestations of information pollution include, among others [68,114]: disinformation (false information deliberately created to harm a person, social group, organization, or country), misinformation (false information not created with the intention of causing harm), malinformation (information based on real facts but manipulated to inflict harm on a person, organization, or country), fake news (false information spread under the guise of being authentic news, usually disseminated through news outlets or the internet with the intention to gain political or financial advantage), and clickbait (the deliberate use of misleading headlines to encourage visitors to click on a particular webpage).

Information pollution presents a significant challenge for societies across various levels [114]. On a broader scale, access to quality information plays a crucial role in cultivating public trust, sustaining transparency of governments, and enhancing social cohesion. Particularly noteworthy are its potentially detrimental effects on professional decision-makers whose performance relies on the quality of the information they receive [75].

Moreover, information pollution spans all types of information. Two illustrative examples of this are fake product endorsements and paper mills. Firstly, fake product reviews, such as

those that can be found on Amazon, can significantly impact perceptions of product quality, making their study one of the primary agenda items in social and digital media marketing research [98]. At the other end of the spectrum, scientific literature also faces contamination through fake manuscripts orchestrated by paper mills— *"the process by which manufactured manuscripts are submitted to a journal for a fee on behalf of researchers with the purpose of providing an easy publication for them, or to offer authorship for sale"* [2]. An unpublished analysis shared with *Nature*, estimates that between 1.5 and 2 % of all scientific articles published in 2022 exhibit many similarities to content generated by paper mills [116].

In this general context, the following societal changes suggest that EBSE could become more important:

Need for Information Literacy Skills. The software engineering community, including academia, industry, government, social actors, and professional communities, is also susceptible to the impacts of information pollution. A recent instance of this is the article released by the consulting giant McKinsey on developer productivity [40]. The article, aimed at comprehending software productivity and suggesting a measurement approach, sparked significant controversy, primarily due to the numerous unsupported statements it contained [73]. The findings of this thesis indicate that undergoing training in EBSE or possessing a foundational understanding of its principles enhances the information literacy skills of practitioners, irrespective of whether they are employed in industry or government. This suggests that they are potentially better equipped to consume information thoughtfully, reducing susceptibility to the impacts of information pollution. This is further emphasized by the observation that EBSE training also appears to promote the utilization of evidence from research to support professional practice. Several authors advocate empowering individuals as a primary strategy to mitigate the effects of information pollution [61]. It seems that the knowledge and training in EBSE may contribute to this empowerment within the SE community.

Generative AI Regulation. Lately, there have been several concerns about the impacts of artificial intelligence (AI). Over the past year, generative AI has brought up some serious questions in policies, especially regarding the copyright of the data they use for training (see, for example, [99, 100]). It is crucial for government organizations to consider how to regulate AI. A recent move in this direction was the 2023 AI Safety Summit[3], where 28 countries united to start discussing the safety and regulation of AI. However, some argue that too much attention is paid to big tech companies and not enough to academia [67]. This should prompt us to reflect on what we might be doing incorrectly. It could be that we are not sufficiently dedicated to investigating issues within professional practice, or perhaps we lack the skills to effectively communicate our findings beyond academic spheres or advocate for a more empirically grounded approach to practice. In order to ensure it is fit for purpose, we need better ways to provide evidence of good practice when developing AI systems and we need processes for evaluating the performance of deployed systems. So government bodies need to understand how to use evidence to make decisions about certification. SE researchers need to consider what sort of evidence is needed, how it can be collected, and how to ensure that evidence is fair and unbiased.

Generative AI & Nature of Evidence. AI systems themselves also raise issues about the nature of evidence. In the presence of systems capable of generating fake news and manufactured research, ensuring the authenticity of submitted evidence becomes a critical concern. Consequently, there is a pressing need to further investigate methods for assessing the credibility of all published reports and Internet materials. The focus should not only be on evaluating the quality of primary studies but also on establishing the trustworthiness of all information sources supporting SE practices. The process of achieving this goal is yet to be thoroughly explored. While guidelines for assessing the quality of grey literature (e.g., [38])

[3] https://www.aisafetysummit.gov.uk/

provide a reasonable starting point, tackling the challenges posed by deep fakes is more complex than identifying invalid arguments or unfounded assertions.

Regulation of software products. As mentioned earlier, the issues linked to deployed AI systems are raising significant concerns, leading to an escalating demand for regulations. SE researchers would need to consider the issues involved in the formal evaluation of software products such as protocols for field trials, independent auditing of the product development and maintenance process as well as the detection of fake evidence. Thus, regulation of software products might increase the importance of EBSE and empirical software engineering in general.

Chapter 5

Conclusions

In this chapter, I bring this thesis to a conclusion by summarizing the primary contributions and elucidating their significance for the software engineering community. Lastly, I present my view on potential future research directions within the scope of the topics of this thesis.

5.1. Concluding Remarks

The initial conclusion drawn from this thesis is that the non-academic application of EBSE is complex. This complexity, coupled with the ambiguity surrounding its value, has the potential to foster significant skepticism among researchers and practitioners alike. It is conceivable that these factors have contributed to the scarcity of practice-driven uses of EBSE reported over the two decades following its inception. My research addresses both of these aspects in distinct manners: by presenting certain findings regarding the value and benefits of EBSE utilization, and by making an effort to comprehend its complexity. The latter includes an exploration of barriers encountered during its application, as well as an investigation into the factors that could potentially facilitate its successful adoption. Specifically, I studied the value and complexity of EBSE by (1) investigating the effects of an EBSE training on the work practices of former students, (2) evaluating the attitudes towards EBSE of members of a government agency and the effects of their working practices, and (3) conducting and evaluating a practice-driven application of EBSE to address a software company problem and studying the practitioners' feedback.

The second noteworthy conclusion is that EBSE appears to hold significant value for the SE community. My own research confirms the use of evidence is essential for SE researchers, all stages of this thesis were based on the results of an SR or a RR. Furthermore, acquiring knowledge of EBSE, whether through a brief introduction or a more comprehensive training, has the potential to enhance participants' work practices. This is achieved by fostering an awareness of research and evidence, as well as refining skills related to information gathering and information literacy. The findings of this thesis suggest that, both within government and industry contexts, the insights obtained from SRs are considered more reliable compared to other forms of information utilized for decision-making. Furthermore, EBSE awareness and application appear to foster a positive disposition towards evidence from research, even among individuals who were previously unfamiliar with it.

Effectively utilizing EBSE (i.e., applying all of its steps) requires the ability to translate academic recommendations into practical process changes and proficiency in devising and utilizing effective mechanisms to disseminate knowledge. Consequently, EBSE appears to be

more appropriate as an approach for researchers to enhance collaboration with practitioners, rather than a freely accessible mechanism for all stakeholders as initially expected [33]. The primary obstacle to EBSE adoption seems to be the limitations of evidence within the SE domain. This reinforces the viewpoint of many researchers who emphasize the necessity for research more aligned with the industry's needs.

Akin to the lack of skills being a major barrier to adopting EBSE, training stands out as one of its primary facilitators. In this thesis, previous teaching initiatives related to EBSEs and SRs were compiled, culminating in the formulation of a set of learning outcomes tailored for training both undergraduate and graduate students. The evaluation of the proposal, as well as the assessment of the course built around it, indicates its suitability. Furthermore, its comprehensive nature enables replication and adaptation in different contexts. The educational materials have garnered interest from teachers at different universities, who are keen to incorporate them into their teaching. My colleagues and I are actively receptive to this interest, aiming to provide assistance and incorporate other researchers' feedback. I believe that the proposal presented in this thesis has the potential to serve as a foundational framework for fostering future consensus on EBSE training, akin to the style of EBP training core skills defined in other disciplines (e.g., [3]).

Despite being introduced in SE in 2018, RRs publication has experienced a substantial increase, particularly evident in 2023 compared to previous years. Notably, many of these RRs involve collaborations with industry partners, marking a difference from the predominantly academically motivated nature of SRs. However, my exploration of RRs in SE revealed numerous risks and challenges. Reports are often incomplete, making it difficult to assess their outcomes adequately. Furthermore, the method itself remains relatively understudied, with various aspects remaining unexplored, such as effective strategies for dissemination and knowledge transfer of its results.

Finally, over the course of this research, I have endeavored to explore a closer connection between EBSE and diverse stakeholder groups, primarily those outside of the academic community. This represents a novelty within EBSE research, coupled with the emphasis on assessing the value of EBSE itself, rather than solely focusing on SRs. My findings underscore the potential value that EBSE could bring to regulatory and governmental bodies, as well as to the software industry. However, further exploration in this direction is imperative to gain a deeper comprehension of their specific needs and to identify modes of collaboration with academia that could yield more productive outcomes.

5.2. Future Agenda

Throughout the work presented in this thesis, I have pinpointed various issues that could serve as potential directions for future research. The following list outlines and elaborates on the key or most promising of these lines for future work.

Enhancing the EBSE training. While the EBSE and SRs course for university students has been deemed adequate, there is room for improvement. Specifically, the course requires a stronger emphasis on EBSE and less focus on the SR process. This shift would enhance both the comprehension of topics such as evidence utilization and knowledge translation, along with fostering a more positive perception among students regarding the value of evidence-based practice. To facilitate this improvement, several approaches could be considered: (1) Incorporating a practice-driven EBSE application report (e.g., [50]) into the reading materials. (2) Considering both of the final two EBSE steps, encompassing discussions on contextual factors that influence knowledge utilization (e.g., company size, staff experience, application types, etc.), as well as reflections on the efficacy of the EBSE process and what this means in practice. (3) Introducing an assignment that challenges students to apply evi-

dence for addressing a professional practice problem. For instance, students could be tasked with examining the application of evidence when dealing with a scenario like establishing a new company and deciding between test-driven development and conventional testing approaches. To facilitate this, EBSE training could adopt the approach of teaching through evidence-based scenarios co-created with practitioners (as recommended by Manns and Darrah [66]).

Research and develop EBSE training for industry and government. Up to this point, I have not encountered any reports of EBSE training extended to companies or other non-academic stakeholders. Considering the advantages identified, it becomes compelling to consider implementing interventions aimed at increasing awareness and imparting EBSE training. In doing so, special attention must be paid to the requirements and circumstances of the trainees, necessitating the customization and creation of user-friendly resources. Additionally, in line with my research findings, it would prove highly beneficial to place a distinct emphasis on EBSE and the utilization of evidence from research (as opposed to its generation). This emphasis should include guidelines or recommendations for effectively assessing grey material (e.g., blog or chatGPT outcomes). Also, as Jørgensen suggests [45][1], it seems advisable to teach and encourage practitioners to make the processes for creating local evidence[2] explicit and somewhat rigorous, thus enhancing the quality and availability of this type of evidence.

Using theories and models to support the investigation of EBSE adoption. Gaining deeper insights into enhancing the adoption of EBSE requires a comprehensive understanding of the decision-making dynamics within the SE domain, as well as the distinct roles various types of information play in the decisions of practitioners. A potential avenue for exploration involves conducting surveys amongst software company managers and other practitioners, delving into queries about the stakeholders responsible for SE methodology and tool decisions, as well as the processes underlying such choices. Notably, it has been proposed that shedding light on and dissecting the decision-making process can significantly bolster the uptake of EBP [10]. Moreover, integrating theoretical frameworks to better understand the decision-making procedures used in SE (e.g., [19]) could significantly contribute to this investigative endeavor. Lastly, using theories or models engineered to facilitate a more nuanced comprehension of the adoption of evidence-based practices (for example, Rogers' diffusion of innovations theory [92]) might provide valuable insights in steering this direction.

Study and promote experiences of practice-driven use of EBSE. Finally, given the scarcity of reports on EBSE applications oriented to practice, it seems advisable to further investigate the existing reports and consult their authors seeking to identify challenges faced and lessons learned. In addition, material such as guides or more detailed recommendations could be generated to apply all the EBSE steps. In our research group, we also intend to continue using EBSE to improve the collaborations with companies and government agencies in Uruguay. In this way, we will also have more opportunities to better study the non-academic use of EBSE.

[1] keynote slides in https://web-backend.simula.no/sites/default/files/publications/files/fse-keynote-2014.ppt.pdf

[2] Local evidence refers to "evidence that is available from the specific setting(s) where a decision or action will be taken." [64] This type of evidence is typically generated by companies, organizations, or agencies and is often produced informally or without a systematic approach. Although local evidence may be more directly relevant than studies conducted elsewhere, it can also be less reliable due to significant limitations in its creation process [64].

References

[1] Agency for Healthcare Research and Quality. Evidence-based Practice Centers (EPC) Program Overview. https://www.ahrq.gov/research/findings/evidence-based-reports/overview/index.html. Accessed: 2023-12-12.

[2] Alam, S., Patel, J., Chap, C., Robbie, S., and Moylan, E. Potential paper mills. https://publicationethics.org/publishers-perspective-paper-mills. Accessed: 2024-05-29.

[3] Albarqouni, L., Hoffmann, T., Straus, S., Olsen, N. R., Young, T., Ilic, D., Shaneyfelt, T., Haynes, R. B., Guyatt, G., and Glasziou, P. Core Competencies in Evidence-Based Practice for Health Professionals. *JAMA Network Open 1*, 2 (2018).

[4] Angrosino, M. *Doing Ethnographic and Observational Research*. SAGE Publications, 2007.

[5] Avgeriou, P., Verbeek, M., Bibi, S., Chatzigeorgiou, A., and Ampatzoglou, A. Identifying, categorizing and mitigating threats to validity in software engineering secondary studies. *Information and Software Technology 106* (2018), 201–230.

[6] Baldassarre, M., Boffoli, N., Caivano, D., and Visaggio, G. A hands-on approach for teaching systematic review. In *International Conference on Product-Focused Software Process Improvement* (Monte Porzio Catone, Italy, 2008).

[7] Basili, V., Briand, L., Bianculli, D., Nejati, S., Pastore, F., and Sabetzadeh, M. Software engineering research and industry: A symbiotic relationship to foster impact. *IEEE Software 35*, 5 (2018), 44–49.

[8] Basili, V. R. A personal perspective on the evolution of empirical software engineering. In *Perspectives on the Future of Software Engineering: Essays in Honor of Dieter Rombach*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 255–273.

[9] Beecham, S., Bowes, D., and Stol, K.-J. Introduction to the ease 2016 special section: Evidence-based software engineering: Past, present, and future. *Information and Software Technology 89* (2017), 14–18.

[10] Benfield, A., and Krueger, R. B. Making decision-making visible—teaching the process of evaluating interventions. *International Journal of Environmental Research and Public Health 18*, 7 (2021).

[11] Birt, L., Scott, S., Cavers, D., Campbell, C., and Walter, F. Member checking: A tool to enhance trustworthiness or merely a nod to validation? *Qualitative Health Research 26*, 13 (2016), 1802–1811.

[12] Bjarnason, E., Åberg, P., and bin Ali, N. Software selection in large-scale software engineering: A model and criteria based on interactive rapid reviews. *Empirical Software Engineering* (2023).

[13] Bossavit, L. *The Leprechauns of software engineering.* Laurent Bossavit; Print Edition G15 (14 Octubre 2013), 2013.

[14] Braun, V., and Clarke, V. Using thematic analysis in psychology. *Qualitative Research in Psychology 3*, 2 (2006), 77–101.

[15] Braun, V., and Clarke, V. *Thematic analysis: a practical guide.* SAGE Publications, London, England, 2022.

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

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

[18] Budgen, D., Kitchenham, B., and Brereton, P. EBSE Web Site. https://ebse.webspace.durham.ac.uk/. Accessed: 2024-05-29.

[19] Burge, J. E., Carroll, J. M., McCall, R., and Mistrik, I. Decision-Making in Software Engineering. In *Rationale-Based Software Engineering.* Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 67–76.

[20] Callo Arias, T. B., van der Spek, P., and Avgeriou, P. A practice-driven systematic review of dependency analysis solutions. *Empirical Software Engineering 16*, 5 (2011), 544–586.

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

[22] Cartaxo, B., Pinto, G., and Soares, S. Rapid reviews in software engineering. In *Contemporary Empirical Methods in Software Engineering.* Springer International Publishing, Cham, 2020, pp. 357–384.

[23] Cartaxo, B., Pinto, G., Vieira, E., and Soares, S. Evidence briefings: Towards a medium to transfer knowledge from systematic reviews to practitioners. In *International Symposium on Empirical Software Engineering and Measurement (ESEM)* (Ciudad Real, Spain, 2016).

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

[25] Castells, M. *The Information Age: Economy, Society and Culture*, 2nd ed. ed. The Rise of the Network Society. Blackwell, 2000.

[26] Clarke, V., and Braun, V. *Successful qualitative research: A practical guide for beginners.* SAGE Publications, London, England, 2013.

[27] Cummings, M. L. M. What self-driving cars tell us about ai risks. https://spectrum.ieee.org/self-driving-cars-2662494269. IEEE Spectrum, Accessed: 2024-05-29.

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

[29] Devanbu, P., Zimmermann, T., and Bird, C. Belief & evidence in empirical software engineering. In *International Conference on Software Engineering (ICSE)* (Austin, TX, USA, 2016), pp. 108–119.

[30] Dijkstra, E. W. Letters to the editor: go to statement considered harmful. *Communications of the ACM 11*, 3 (1968), 147–148.

[31] dos Santos, P. S. M., and Travassos, G. H. Action research can swing the balance in experimental software engineering. In *Advances in Computers*, M. V. Zelkowitz, Ed., vol. 83. Elsevier, 2011, pp. 205–276.

[32] Dybå, T., and Dingsøyr, T. Strength of evidence in systematic reviews in software engineering. In *ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)* (Kaiserslautern, Germany, 2008).

[33] Dybå, T., Kitchenham, B., and Jørgensen, M. Evidence-based software engineering for practitioners. *IEEE Software 22*, 1 (2005), 58–65.

[34] Felderer, M., and Travassos, G. H. The evolution of empirical methods in software engineering. In *Contemporary Empirical Methods in Software Engineering*. Springer International Publishing, 2020, pp. 1–24.

[35] Ferlie, E. *Conclusion: From Evidence to Actionable Knowledge?* Oxford Scholarship Online, 05 2005, pp. 182–197.

[36] França, B. B. N., Ribeiro, T., Santos, P. M. D., and Travassos, G. Using focus group in software engineering: lessons learned on characterizing software technologies in academia and industry. In *Ibero-American Conference on Software Engineering (CIbSE)* (Lima, Perú, 2015), pp. 1–15.

[37] Garousi, V., and Felderer, M. 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 (EASE)* (Karlskrona, Sweden, 2017).

[38] Garousi, V., Felderer, M., and Mäntylä, M. V. Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. *Information and Software Technology 106* (2019), 101–121.

[39] Glass, R. The software-research crisis. *IEEE Software 11*, 6 (1994), 42–47.

[40] Gnanasambandam, C., Harrysson, M., Hussin, A., Keovichit, J., and Srivastava, S. Yes, you can measure software developer productivity. https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/yes-you-can-measure-software-developer-productivity. Accessed: 2024-05-29.

[41] Hamel, C., Michaud, A., Thuku, M., Skidmore, B., Stevens, A., Nussbaumer-Streit, B., and Garritty, C. Defining rapid reviews: a systematic scoping review and thematic analysis of definitions and defining characteristics of rapid reviews. *Journal of Clinical Epidemiology 129* (2021), 74–85.

[42] Hassler, E., Carver, J., Kraft, N., and Hale, D. 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 (EASE)* (London, UK, 2014).

[43] Janzen, D. S., and Ryoo, J. Seeds of evidence: Integrating evidence-based software engineering. In *Conference on Software Engineering Education and Training* (Charleston, South Carolina, USA, 2008).

[44] Jørgensen, M. A review of studies on expert estimation of software development effort. *Journal of Systems and Software* (2004).

[45] Jørgensen, M. Ten years with evidence-based software engineering. what is it? has it had any impact? what's next? In *ACM SIGSOFT International Symposium on Foundations of Software Engineering* (Hong Kong, China, 2014).

[46] Jørgensen, M., Dybå, T., and Kitchenham, B. Teaching evidence-based software engineering to university students. In *International Software Metrics Symposium* (Como, Italy, 2005).

[47] Kalenda, M., Hyna, P., and Rossi, B. Scaling agile in large organizations: Practices, challenges, and success factors. *Journal of Software: Evolution and Process 30*, 10 (2018).

[48] Kamei, F., Wiese, I., Lima, C., Polato, I., Nepomuceno, V., Ferreira, W., Ribeiro, M., Pena, C., Cartaxo, B., Pinto, G., and Soares, S. Grey Literature in Software Engineering: A critical review. *Information and Software Technology 138* (2021).

[49] Kamei, F., Wiese, I., Neto, C., Polato, I., Nepomuceno, V., Ferreira, W., Ribeiro, M., Pena, C., Cartaxo, B., Pinto, G., and Soares, S. Replication package of the study "Grey literature in software engineering: A critical review"published in Information and Software Technology (IST). https://doi.org/10.5281/zenodo.6780520, June 2022. Accessed: 2024-05-29.

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

[51] Kennedy, D., Hyl, Á., Ryan, N., Hyland, Á., and Ryan, N. Writing and using learning outcomes: A practical guide. In *Bologna Handbook*. European University Association, 2007, ch. C 3.4-1, pp. 1–30.

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

[53] Kitchenham, B., Brereton, P., and Budgen, D. The educational value of mapping studies of software engineering literature. In *International Conference on Software Engineering (ICSE)* (Cape Town, South Africa, 2010).

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

[55] Kitchenham, B., and Charters, S. Guidelines for performing systematic literature reviews in software engineering. Tech. Rep. EBSE-2007-01, Keele University, 2007.

[56] Kitchenham, B., Dybå, T., and Jørgensen, M. Evidence-based software engineering. In *International Conference on Software Engineering (ICSE)* (Scotland, UK, 2004).

[57] Kitchenham, B., Madeyski, L., and Budgen, D. SEGRESS: Software Engineering Guidelines for REporting Secondary Studies. *IEEE Transactions on Software Engineering 49*, 3 (2023), 1273–1298.

[58] Ko, A. J. A three-year participant observation of software startup software evolution. In *IEEE/ACM International Conference on Software Engineering: Software Engineering in Practice Track (ICSE-SEIP)* (Buenos Aires, Argentina, 2017), pp. 3–12.

[59] Kontio, J., Bragge, J., and Lehtola, L. The focus group method as an empirical tool in software engineering. In *Guide to Advanced Empirical Software Engineering*. Springer London, London, 2008, ch. 4, pp. 93–116.

[60] Kontio, J., Lehtola, L., Bragge, J., and Box, P. O. Using the Focus Group Method in Software Engineering: Obtaining Practitioner and User Experiences. In *International Symposium on Empirical Software Engineering* (Redondo Beach, California, 2004), pp. 271–280.

[61] Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., and Zittrain, J. L. The science of fake news. *Science 359*, 6380 (2018), 1094–1096.

[62] Le Goues, C., Jaspan, C., Ozkaya, I., Shaw, M., and Stolee, K. T. Bridging the Gap: From Research to Practical Advice. *IEEE Software 35*, 5 (2018), 50–57.

[63] Lenberg, P., Feldt, R., Gren, L., Wallgren Tengberg, L. G., Tidefors, I., and Graziotin, D. Qualitative software engineering research: Reflections and guidelines. *Journal of Software: Evolution and Process* (2023).

[64] Lewin, S., Oxman, A. D., Lavis, J. N., Fretheim, A., Marti, S. G., and Munabi-Babigumira, S. Support tools for evidence-informed policymaking in health 11: Finding and using evidence about local conditions. *Health Research Policy and Systems 7*, 1 (Dec 2009), S11.

[65] Lewowski, T., and Madeyski, L. Code smells detection using artificial intelligence techniques: A business-driven systematic review. In *Developments in Information & Knowledge Management for Business Applications : Volume 3*. Springer International Publishing, 2022, pp. 285–319.

[66] Manns, P., and Darrah, J. A structured process to develop scenarios for use in evaluation of an evidence-based approach in clinical decision making. *Advances in Medical Education and Practice 3* (2012), 113–119.

[67] Marcus, G. Controlling AI. *Communications of the ACM 66*, 10 (sep 2023), 6–7.

[68] Meel, P., and Vishwakarma, D. K. Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. *Expert Systems with Applications 153* (2020).

[69] Méndez Fernández, D., and Passoth, J. H. Empirical software engineering: From discipline to interdiscipline. *Journal of Systems and Software 148* (2019), 170–179.

[70] Michal, A. L., Zhong, Y., and Shah, P. When and why do people act on flawed science? effects of anecdotes and prior beliefs on evidence-based decision-making. *Cogn Res Princ Implic 6*, 1 (Apr. 2021), 28.

[71] Miles, M. B., Huberman, A. M., Saldana, J. *Qualitative Data Analysis: A Methods Sourcebook*, 3 ed. SAGE Publications, Ltd, 2014.

[72] Nagappan, M., Robbes, R., Kamei, Y., Tanter, E., McIntosh, S., Mockus, A., and Hassan, A. E. An empirical study of goto in c code from github repositories. In *Joint Meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on the Foundations of Software Engineering (ESEC/FSE)* (Bergamo Italy, 2015).

[73] North, D. Mckinsey developer productivity review. https://dannorth.net/mckinsey-review/. Accessed: 2024-05-29.

[74] Oates, B., and Capper, G. Using systematic reviews and evidence-based software engineering with masters students. In *International Conference on Evaluation and Assessment in Software Engineering (EASE)* (Swindon, UK, 2009).

[75] Orman, L. Fighting information pollution with decision support systems. *Journal of management information systems 1*, 2 (1984), 64–71.

[76] Ouriques, R., Wnuk, K., Berntsson Svensson, R., and Gorschek, T. Thinking strategically about knowledge management in agile software development. In *International Conference on Product-Focused Software Process Improvement (PROFES)* (Wolfsburg, Germany, 2018), pp. 389–395.

[77] O'Brien, B. C., Harris, I. B., Beckman, T. J., Reed, D. A., and Cook, D. A. Standards for reporting qualitative research: a synthesis of recommendations. *Academic Medicine 89*, 9 (2014), 1245–1251.

[78] O'Connor, C., and Joffe, H. Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods 19* (2020).

[79] Pizard, S. Ingeniería de software basada en evidencia: Dos casos de aplicación en Uruguay. https://www.youtube.com/watch?v=58F1M9O-bH0. Accessed: 2024-05-29.

[80] Pizard, S., Acerenza, F., Casella, V., Moreno, S., García, R., Lezama, J., and Vallespir, D. Conceptos de Ingeniería de Software Basada en Evidencias (version 2). https://hdl.handle.net/20.500.12008/25557. Tech. Rep., Universidad de la República, School of Engineering, 2019.

[81] Pizard, S., Acerenza, F., Otegui, X., Moreno, S., Vallespir, D., and Kitchenham, B. Training students in evidence-based software engineering and systematic reviews: a systematic review and empirical study. *Empirical Software Engineering 26*, 3 (2021), 50.

[82] Pizard, S., Acerenza, F., Vallespir, D., and Kitchenham, B. Materials for an Evidence-based software engineering & Systematic reviews course. https://doi.org/10.5281/zenodo.7464649, Dec. 2022.

[83] Pizard, S., Acerenza, F., Vallespir, D., and Kitchenham, B. Assessing attitudes towards evidence-based software engineering in a government agency. *Information and Software Technology 154* (2023).

[84] Pizard, S., and Herbert, J. ¿Más valor generado en el contexto ágil? Show me the evidences! https://www.youtube.com/watch?v=U2YfxpcQCEI. Accessed: 2024-05-29.

[85] Pizard, S., Lezama, J., García, R., Vallespir, D., and Kitchenham, B. Using rapid review to support software engineering practice: a citation analysis and a replication study. https://hdl.handle.net/20.500.12008/44579, 2024.

[86] Pizard, S., Vallespir, D., and Kitchenham, B. A longitudinal case study on the effects of an evidence-based software engineering training. In *IEEE/ACM International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET)* (Pittsburgh, PA, USA, 2022), pp. 1–13.

[87] Pizard, S., Vallespir, D., and Kitchenham, B. Evidence-based software engineering revisited: evaluation of a practice-driven application. https://hdl.handle.net/20.500.12008/44572, 2024.

[88] Potts, C. Software-engineering research revisited. *IEEE Software 10*, 5 (1993), 19–28.

[89] Ralph, P., Baltes, S., Bianculli, D., Dittrich, Y., Felderer, M., Feldt, R., Filieri, A., Furia, C. A., Graziotin, D., He, P., Hoda, R., Juristo, N., Kitchenham, B. A., Robbes, R., Méndez, D., Molléri, J. S., Spinellis, D., Staron, M., Stol, K., Tamburri, D. A., Torchiano, M., Treude, C., Turhan, B., and Vegas, S. ACM SIGSOFT Empirical Standards for Software Engineering Research. https://github.com/acmsigsoft/EmpiricalStandards. Accessed: 2024-05-29.

[90] Rapoport, R. N. Three dilemmas in action research: With special reference to the tavistock experience. *Human Relations 23*, 6 (1970), 499–513.

[91] Rico, S., Ali, N. B., Engström, E., and Höst, M. Experiences from conducting rapid reviews in collaboration with practitioners — Two industrial cases. *Information and Software Technology 167* (2024), 107364.

[92] Rogers, E. M. *Diffusion of innovations*, 5th ed. Free Press, New York, NY, 08 2003.

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

[94] Runeson, P., Host, M., Rainer, A., and Regnell, B. *Case Study Research in Software Engineering: Guidelines and Examples*, 1st ed. Wiley Publishing, 2012.

[95] Rycroft-Malone, J., Seers, K., Titchen, A., Harvey, G., Kitson, A., and McCormack, B. What counts as evidence in evidence-based practice? *J Adv Nurs 47*, 1 (July 2004), 81–90.

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

[97] Saini, N., Singh, S., and Suman. Code clones: Detection and management. *Procedia Computer Science 132* (2018), 718–727. International Conference on Computational Intelligence and Data Science (ICCIDS).

[98] Salminen, J., Kandpal, C., Kamel, A. M., Jung, S.-g., and Jansen, B. J. Creating and detecting fake reviews of online products. *Journal of Retailing and Consumer Services 64* (2022).

[99] Samuelson, P. Legal Challenges to Generative AI, Part I. *Communications of the ACM 66*, 7 (2023), 20–23.

[100] Samuelson, P. Legal Challenges to Generative AI, Part II. *Communications of the ACM 66*, 11 (2023), 16–19.

[101] Santos, R. E., and Silva, F. Q. d. Motivation to perform systematic reviews and their impact on software engineering practice. In *ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)* (Baltimore, MD, USA, 2013), pp. 292–295.

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

[103] Schwartz, M. S., and Schwartz, C. G. Problems in Participant Observation. *American Journal of Sociology 60*, 4 (1955), 343–353.

[104] Scurlock-Evans, L., and Upton, D. 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 (2015), 369–399.

[105] Sedano, T., Ralph, P., and Péraire, C. Software development waste. In *IEEE/ACM International Conference on Software Engineering (ICSE)* (Buenos Aires, Argentina, 2017), pp. 130–140.

[106] Shull, F., Mendoncça, M. G., Basili, V., Carver, J., Maldonado, J. C., Fabbri, S., Travassos, G. H., and Ferreira, M. C. Knowledge-sharing issues in experimental software engineering. *Empirical Software Engineering 9* (2004), 111–137.

[107] Sierra, J. M., Vizcaíno, A., Genero, M., and Piattini, M. A systematic mapping study about socio-technical congruence. *Information and Software Technology 94* (2018), 111–129.

[108] Sommerville, I. *Software Engineering*, 9 ed. Addison-Wesley, Harlow, England, 2010.

[109] Song, Q., Borg, M., Engström, E., Ardö, H., and Rico, S. Exploring ml testing in practice: Lessons learned from an interactive rapid review with axis communications. In *International Conference on AI Engineering* (Pittsburgh, PA, USA, 2022).

[110] Stenbacka, C. Qualitative research requires quality concepts of its own. *Management decision 39*, 7 (2001), 551–556.

[111] Stol, K.-J., and Fitzgerald, B. The ABC of Software Engineering Research. *ACM Trans. Softw. Eng. Methodol. 27*, 3 (2018).

[112] Theisen, C., Dunaiski, M., Williams, L., and Visser, W. Writing good software engineering research papers: Revisited. In *IEEE/ACM International Conference on Software Engineering Companion (ICSE)* (Buenos Aires, Argentina, 2017).

[113] Theisen, C., Dunaiski, M., Williams, L., and Visser, W. Software engineering research at the international conference on software engineering in 2016. *SIGSOFT Softw. Eng. Notes 42*, 4 (2018), 1–7.

[114] United Nations Development Programme. Mapping and Analysis of Efforts to Counter Information Pollution in Europo and Central Asia Region. Tech. rep., November 2022.

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

[116] Van Noorden, R. How big is science's fake-paper problem? https://www.nature.com/articles/d41586-023-03464-x. Nature News, Nov 2023, Accessed: 2024-05-29.

[117] Wohlin, C. Case study research in software engineering—it is a case, and it is a study, but is it a case study? *Information and Software Technology 133* (2021).

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

Appendix A

Evidence-based Software Engineering Training

Empirical Software Engineering (2021) 26: 50
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[1] · Fernando Acerenza[1] · Ximena Otegui[2] · Silvana Moreno[1] · Diego Vallespir[1] · Barbara Kitchenham[3]

Accepted: 19 February 2021 / Published online: 31 March 2021

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

Communicated by: Scott Fleming

✉ Sebastián Pizard
spizard@fing.edu.uy

Extended author information available on the last page of the article.

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.

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; 2019b; 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 Sections 7 and 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 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.

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”))

Table 2 Results of selection process

	First Stage						Second Stage
	Papers Agreed		Papers	Papers	Kappa	Papers	Papers
Search	Include	Exclude	Disagreed	Total		Selected	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

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 1) 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 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 1.

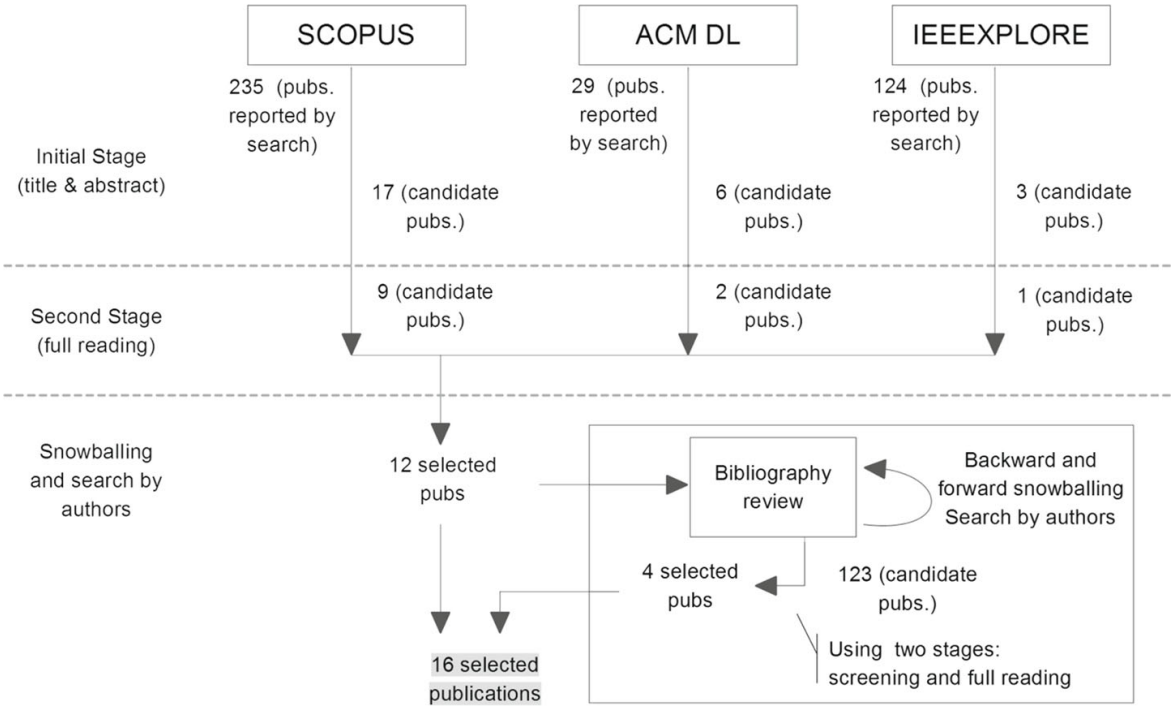


Fig. 1 Surveyed literature flowchart

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 1. 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 1.

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 Fig. 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.

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

Table 3 General characteristics of studies on EBSE teaching initiatives

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

Table 3 (continued)

Id	Paper	Summary of aims of the study	Main motivation	Type of study	Overall quality (% of relevant questions)
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 4 Context of 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)

(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.
– 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:

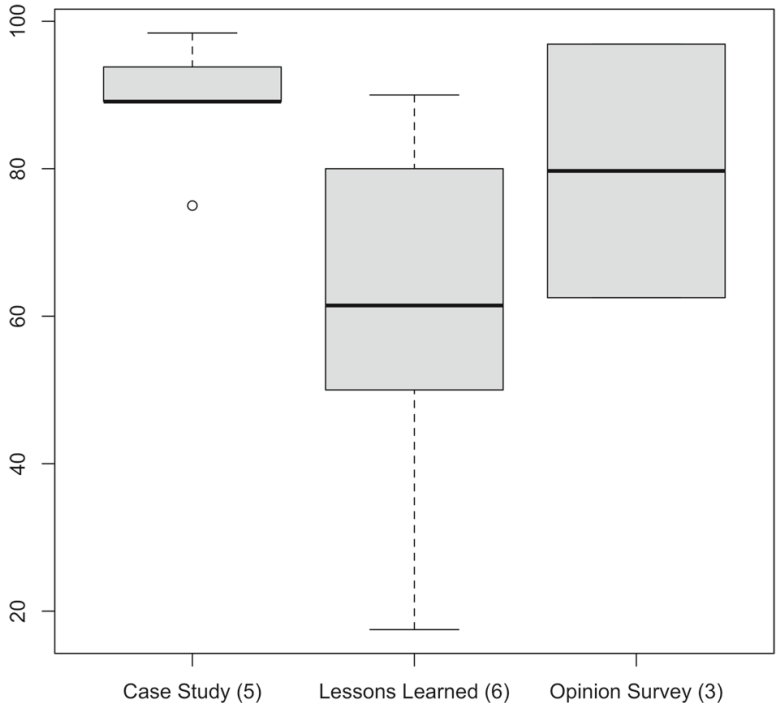


Fig. 2 Quality score for types of study (number of studies in parenthesis)

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).

Table 5 Content and Methodology of EBSE teaching initiatives

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 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		

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.

Table 7 Common issues and recommendations

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 8 Common benefits

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	

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).

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 2 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 2 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 2. 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.

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 Section 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.

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

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
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

Table 9 (continued)

Id.	Learning outcome	Bloom's level
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		
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

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.

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

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

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 3. 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 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 Fig. 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.

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).

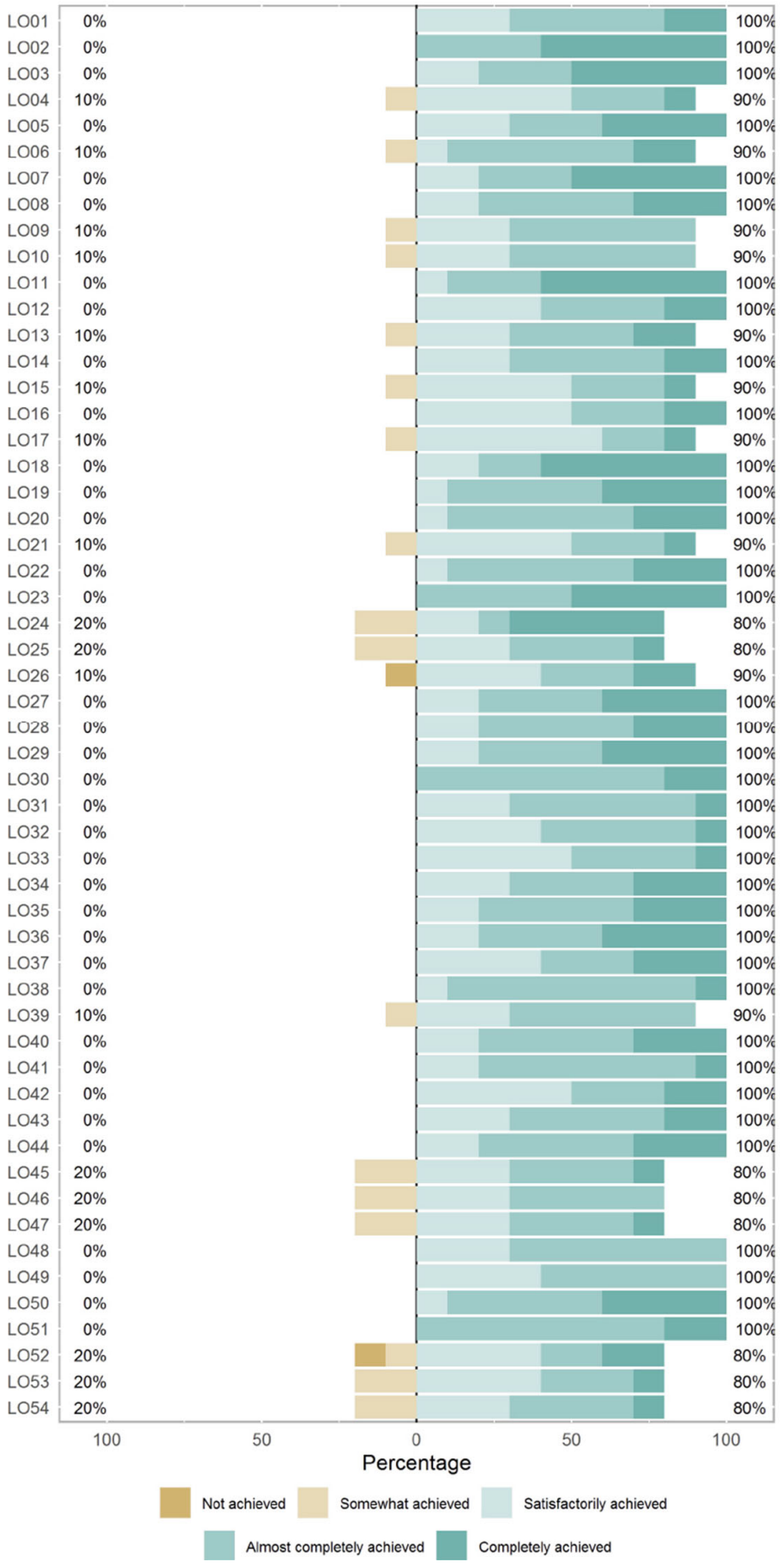


Fig. 3 LO achievement according to student opinion survey

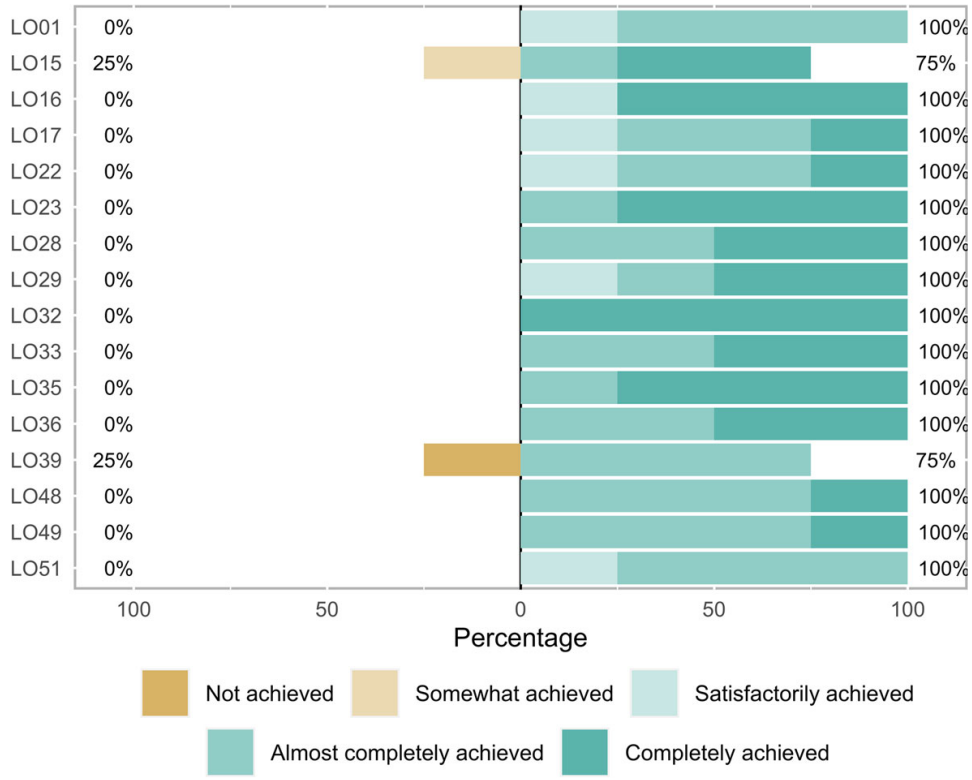


Fig. 4 LO achievement according to learning assessments from team assignment

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 4. 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.

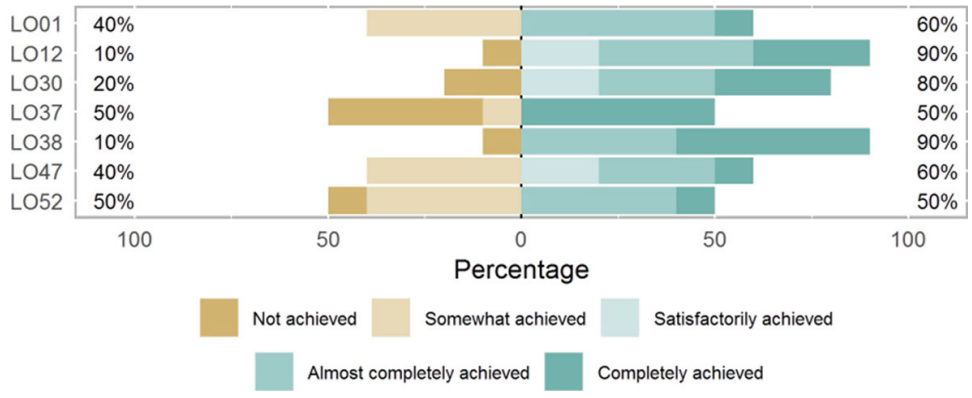


Fig. 5 LO achievement according to learning assessments from individual tests

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 4. 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.

7.3.1 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 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.

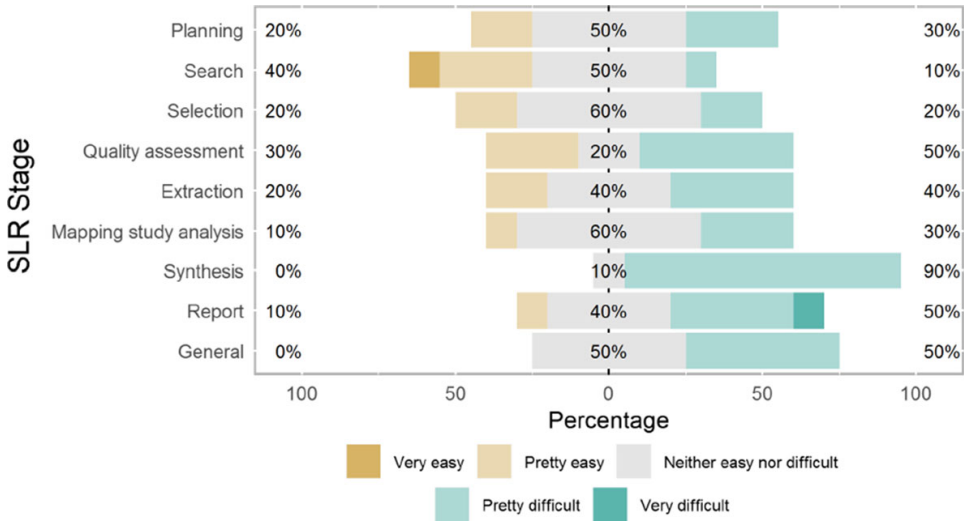


Fig. 6 Opinion survey on the difficulty of conducting an SLR

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).

7.3.2 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 Fig. 4) and their opinion on the achievement of objectives (see Fig. 3) are quite similar and give an account of the acquired skills. However, the evaluation of the individual tests (see Fig. 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 Section 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

Section 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 7.3.2); 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 Section 2.7.2 and in the authors' participation in the case study presented in Section 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.

Appendix 1: 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.

1.1 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.

1.2 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.

1.3 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

Table 11 Adapted search strings

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"))

- 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

Table 12 Papers by search

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

  - 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.

1.4 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 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

Table 13 Initial agreement in the categorization of studies

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

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 ($p < 0.05$).

Appendix 2: 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, educational 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 over-simplification 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.

Appendix 3: 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.

3.1 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)

3.2 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.

3.3 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.

3.4 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.

3.5 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.

Appendix 4: Case Study Additional Data

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

4.1 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.

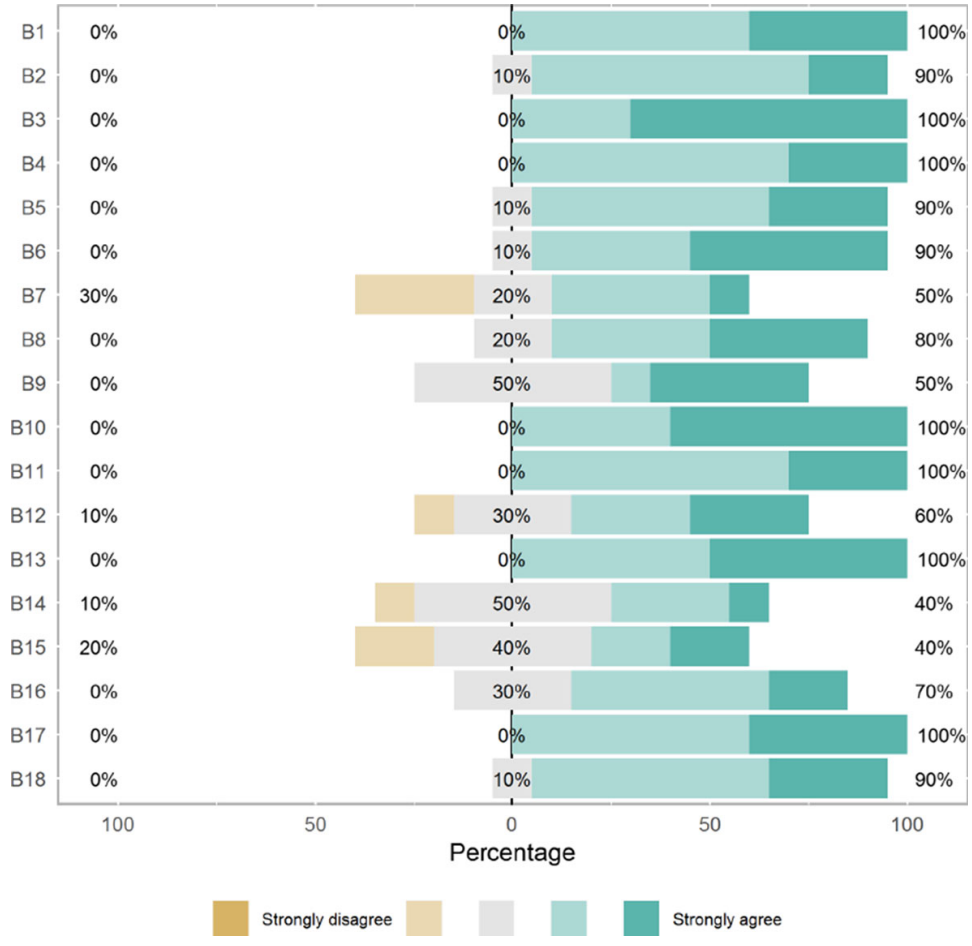


Fig. 7 Opinion survey on general aspects of the course and teamwork

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

- 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

4.2 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.

4.3 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.

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

Author 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.

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.

Declarations

Conflict of Interests The authors declare that they have no conflict of interest.

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 Educ 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 Netw 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. Wiley, Hoboken

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 Softw 35(5):44–49

Beall J (2012) Criteria for determining predatory open-access publishers by Jeffrey Beall. 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. 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. Clinic Psychol Sci Pract 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. Amer Assoc Higher Educ Bullet 51

Brereton P (2011) A study of computing undergraduates undertaking a systematic literature review. IEEE Trans Educ 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 Educ 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-Informat Softw Eng J 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 Softw Eng 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 Softw Eng 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. Inform Softw Technol 53(9):899–913

DeFranco JF, Laplante PA (2017) A content analysis process for qualitative software engineering research. Innov Syst Softw Eng 13(2-3):129–141

Dybå T, Dingsøyr T (2008) Empirical studies of agile software development: A systematic review. Inf Softw Technol 50(9-10):833–859

Elo S, Kyngäs H (2008) The qualitative content analysis process. J Adv Nurs 62(1):107–115

Emerson RM, Fretz RI, Shaw LL (2001) Participant observation and fieldnotes. SAGE Publications Ltd, chap 24 352–368

Feinstein AR, Cicchetti DV (1990) High agreement but low kappa: I. the problems of two paradoxes. J Clin Epidemiol 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. Educ Assess Evaluat Accountabilit 28(3):205–223

Janzen D, Ryoo J (2009) Engaging the net generation with evidence-based software engineering through a community-driven web database. J Syst Softw 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, pp 143–150. June 26-30, 2006. Revised papers Springer Berlin Heidelberg, chap Educational objectives for empirical methods

Kasoju A, Petersen K, Mäntylä M (2013) Analyzing an automotive testing process with evidence-based software engineering. Inf Softw Technol 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 1–30

Kitchenham B, Brereton P (2013) A systematic review of systematic review process research in software engineering. Inf Softw Technol 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, Boca Raton

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, chap 4, pp 93–116

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

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

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

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

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. Empir Softw Eng 23(3):1594–1663

Rice W (2006) Moodle E-learning course development. Packt Publishing, Birmingham

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

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

Sadeghi-Bazargani H, Tabrizi JS, Azami-Aghdash S (2014) Barriers to evidence-based medicine: A systematic review. J Eval Clin Pract 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 Evid-Based Nurs 16(3):176–185

Schoepp K (2019) The state of course learning outcomes at leading universities. Stud High Educ 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. J Evid-Inform Soc 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. Educ Sci 6(4):37:1–37:12

Stol KJ, Fitzgerald B (2018) The ABC of software engineering research. ACM Trans Softw Eng Methodol 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. Br J Occup Ther 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. Assess Evaluat Higher Educ 38(5):567–579

Viera AJ, Garrett JM (2005) Understanding interobserver agreement: The kappa statistic. Fam Med 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. Br J Gen Pract 62(600):511–521

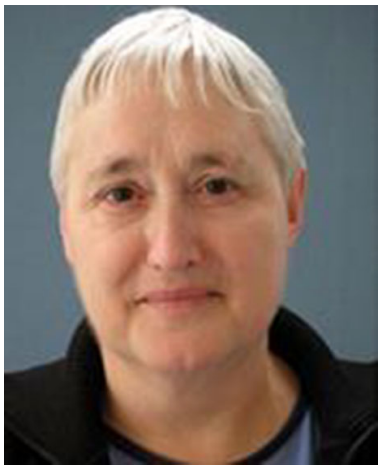
Sebastián Pizard is teaching assistant in the Computer Science Institute at Universidad de la República (Udelar), working in the Software Engineering Group since 2008. He holds a B.Sc. degree in Computer Engineering from Udelar and a master degree in Computer Science from Program for the Basic Sciences Development of Uruguay (PEDECIBA). He has 10 years of experience in the software industry, taking different technical and managerial roles in medium and large software development and maintenance projects for local government agencies. He is a PhD student at PEDECIBA and his main research interest is the use of evidence-based practice in software engineering.

Fernando Acerenza holds a B.Sc. degree in Computer Engineer and a masters degree in Software Engineering from Universidad de la República (Udelar), Uruguay. He works in the Udelar Software Engineering Group as teaching assistant at the Engineering School. His current research interest is evidence-based software engineering. Besides, he has relevant experience in the software industry, working for 14 years, where he is currently in charge of the Technology and Innovation department in a local fintech company.

Ximena Otegui is a PhD student in Education at the Faculty of Humanities and Educational Sciences of the Universidad de la República (Udelar), Uruguay. She holds a master degree in University Teaching (Udelar). Her current research interests are engineering teaching and active learning in engineering education. She is director of the Teaching Department of the Engineering School (Udelar). Her works involve providing pedagogical support to professors and teachers, developing pedagogical-training courses, and tutoring teaching teams interested in improving their teaching and assessing methods.

Silvana Moreno holds a B.Sc. degree in Computer Engineering from Universidad de la República (Udelar) and a master degree in Computer Science from Program for the Basic Sciences Development of Uruguay (PEDECIBA). She has worked in the Udelar Software Engineering Group since 2012, where she is currently teaching assistant. She is a PhD student at PEDECIBA and her current research interest is the software design practice.

Diego Vallespir holds a PhD. in Computer Science from the Universidad de la República (Udelar). He is associate professor at the Computer Science Institute of the Engineering School (Udelar), where he has worked since 2001. He is Director of the Udelar Software Engineering Group. He also works as Coordinator of the Uruguay Software Engineering Program (IS.uy). His current research topics are software development processes and quality management with an empirical approach and, particularly, working with the Personal Software Process and the Team Software Process.



Barbara Kitchenham is Professor of Quantitative Software Engineering at Keele University in the UK. She has worked in software engineering for over 40 years both in industry and academia. She has published over 150 software engineering journal and conference papers. Her main research interests are the application of evidence-based practice to software engineering, the use of statistical methods in software engineering including meta-analysis and robust methods, and open and reproducible science. In 2019, she was awarded the IEEE Technical Committee Distinguished Women in Science & Engineering(WISE) Leadership award.

Affiliations

Sebastián Pizard[1] · Fernando Acerenza[1] · Ximena Otegui[2] · Silvana Moreno[1] · Diego Vallespir[1] · Barbara Kitchenham[3]

Fernando Acerenza
acerenza@fing.edu.uy

Ximena Otegui
xotegui@fing.edu.uy

Silvana Moreno
smoreno@fing.edu.uy

Diego Vallespir
dvallesp@fing.edu.uy

Barbara Kitchenham
b.a.kitchenham@keele.ac.uk

1 Engineering School, Universidad de la República, Montevideo, Uruguay

2 Teaching Department, Engineering School, Universidad de la República, Montevideo, Uruguay

3 School of Computing and Mathematics, Keele University, Keele, Staffordshire, ST5 5BG, UK

2022 IEEE/ACM 44th International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET)

2022 IEEE/ACM 44th International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET) | 978-1-6654-9592-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICSE-SEET55299.2022.9794238

A longitudinal case study on the effects of an evidence-based software engineering training

Sebastián Pizard
School of Engineering, Universidad de la República
Montevideo, Uruguay
spizard@fing.edu.uy

Diego Vallespir
School of Engineering, Universidad de la República
Montevideo, Uruguay
dvallesp@fing.edu.uy

Barbara Kitchenham
School of Computing and Mathematics, Keele University
Keele, UK
b.a.kitchenham@keele.ac.uk

ABSTRACT

Context: Evidence-based software engineering (EBSE) can be an effective resource to bridge the gap between academia and industry by balancing research of practical relevance and academic rigor. To achieve this, it seems necessary to investigate EBSE training and its benefits for the practice. *Objective:* We sought both to develop an EBSE training course for university students and to investigate what effects it has on the attitudes and behaviors of the trainees. *Method:* We conducted a longitudinal case study to study our EBSE course and its effects. For this, we collect data at the end of each EBSE course (2017, 2018, and 2019), and in two follow-up surveys (one after 7 months of finishing the last course, and a second after 21 months). *Results:* Our EBSE courses seem to have taught students adequately and consistently. Half of the respondents to the surveys report making use of the new skills from the course. The most-reported effects in both surveys indicated that EBSE concepts increase awareness of the value of research and evidence and EBSE methods improve information gathering skills. *Conclusions:* As suggested by research in other areas, training appears to play a key role in the adoption of evidence-based practice. Our results indicate that our training method provides an introduction to EBSE suitable for undergraduates. However, we believe it is necessary to continue investigating EBSE training and its impact on software engineering practice.

KEYWORDS

Evidence-based software engineering, evidence-based practice, training, longitudinal case study

ACM Reference Format:
Sebastián Pizard, Diego Vallespir, and Barbara Kitchenham. 2022. A longitudinal case study on the effects of an evidence-based software engineering training. In *44nd International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET '22), May 21–29, 2022, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3510456.3514150

ICSE-SEET '22, May 21–29, 2022, Pittsburgh, PA, USA

ACM ISBN 978-1-6654-9592-9/22/05...$15.00
https://doi.org/10.1145/3510456.3514150

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 [35]. EBSE has been widely adopted by researchers. For example, Budgen et al. report having found 178 systematic reviews (SRs) published only in the most highly ranked software engineering journals between 2010 and 2015 [9]. In contrast, little is known about the adoption of EBSE by other stakeholders (e.g., government or industry practitioners). Although lack of EBSE adoption is known to be a problem [13, 14, 18, 22, 34], to our knowledge, there are only three studies that report the application of EBSE in non-academic settings [11, 30, 42].

Other disciplines where the adoption of evidence-based practice (EBP) is being studied and promoted, have identified the critical relevance of appropriate training. In particular, several systematic reviews placed the lack of knowledge and skills as one of the most commonly reported barriers to adoption [55, 56, 61, 64].

Motivated by these findings, we have undertaken a research program intended to contribute to improving EBSE adoption by developing and evaluating an EBSE training initiative appropriate for delivery in a university environment and assessing its possible effects on professional practices.

The initial stage of our research [48], involved undertaking a systematic review (SR) aimed at assessing previous EBSE training initiatives, developing an EBSE course using the learning outcome approach aimed at codifying the knowledge and skill required of future EBSE users, and the delivery and evaluation the EBSE course based on students' performance and initial opinions after the course.

The research reported in this paper provides (1) further evaluation of our training method, and (2) an assessment of the impact that the training had on the attitudes and behaviors of students, particularly those related to their working practices. After the first couse, we delivered the EBSE course two more times and used the same initial evaluation process as we did for the first course which was based on surveys, focus groups, and teacher assessments undertaken as part of the training module. Subsequently, to analyze the impact of training, we conducted two surveys of the course participants (one after 7 months of finishing the last course, and a second after 21 months).

The remainder of this paper is structured as follows. In Section 2 we briefly summarize work related to EBSE training and the effects of EBSE use in industry. Section 3 presents our proposal for teaching EBSE and SRs and the context in which it was delivered. In Section 4, we specify our research objectives and questions, and describe the research strategy that was used to address them. Section 5

Table 1: Reported benefits of academic EBSE trainings

Benefit	Reported as results in
Acquire or improve research skills	[17, 33]
Become aware of the value of aggregating evidence	[5]
Learn how to search the literature and organize results	[33]
Learn how to assess the relevance, validity or quality of the information on a topic	[17]
Practice the use of digital libraries	[17]

Catal [17] did not indicate how he obtained these results.

presents the data collection and analysis processes. The results are presented and discussed in Section 6. Finally, we present our concluding remarks in Section 7.

2 RELATED WORK

Our study seeks to evaluate the impact of an EBSE training module on the subsequent work practices of the trainees. So in this section, we provide an overview of EBSE training research, and also of uses (or proposals for use) of EBSE aimed at improving collaboration between industry and academia.

2.1 EBSE training

Unlike other disciplines in which evidence-based practice training is widely studied (see, for example, [23, 36, 37, 40]), there are not many studies of EBSE training. In a recent SR [48], we found only 14 investigations of teaching EBSE (reported in 16 articles: [5, 7, 8, 15–17, 25, 26, 29, 33, 38, 45, 49, 50, 52, 60]). The studies reported EBSE training courses with postgraduates and undergraduate students that took place before 2014 and were carried out by universities in seven countries (i.e., Brazil, Canada, Italy, Norway, Turkey, UK & USA). The main purpose of half the studies was related to the teaching of EBSE, while the rest attempted to study either the EBSE process or attitudes towards EBSE.

All studies included a practical assignment. Typically, it involved participating in the conduct of a secondary study, i.e., an SLR, a limited SLR, or a mapping study. Only three studies focussed on teaching EBSE rather than on the conduct of SRs [29, 49, 50]. None of the studies reported any subsequent evaluation of possible impacts of the training on the professional practice of those trained.

However, three studies report benefits obtained by graduate students from taking EBSE courses [5, 17, 33]. Table 1 summarizes their results. Two of them reported having used questionnaires that the students answered after finishing their practical work [5, 33]. The reported benefits were acquiring research skills, being aware of the value of the evidence, and learning to search and organize information. However, Catal does not indicate how he obtained his results [17], so, despite being included in Table 1, we do not use them in the discussion of our own results in Section 6.3.

2.2 Effects of EBSE in practice

Although EBSE explicitly includes within its steps the transfer of knowledge to professional practice, until now more emphasis has been placed on conducting secondary studies than on transferring its results to practice. To our knowledge, there are only three EBSE application reports in non-academic contexts.

First, Kasoju et al. tried to analyze and improve an automotive testing process using among other things EBSE (which included conducting a systematic review) [30]. The improvement proposals received positive comments from practitioners, although their implementation is not reported.

Second, Cartaxo et al. reported the conduct of a rapid review (a limited systematic review) in search of practical recommendations that would help an agile development organization with customer collaboration issues [11]. Although the focus was the evaluation of rapid reviews in software engineering, it is also an application of EBSE. Some of the study recommendations were implemented with good results. The results of the rapid review seemed more reliable to the practitioners than the information they usually used, which increased their confidence in their decisions.

Finally, Lewowski & Madeyski reported the conduct of an SR commissioned by a software development company [42], which sought to learn about advances in predicting code smells using AI to improve its platform for automated code review. The SR was an initial stage in a joint research & development project between practitioners and researchers. The use of the results by the company is not discussed in the paper.

Two other initiatives proposed approaches to transfer knowledge from scientific evidence to SE practice using EBSE as a key practice. First, Cartaxo et al. proposed a model that uses rapid reviews and evidence briefings (one-page reports of evidence) [12]. Second, Badampudi et al. proposed and evaluated a framework for knowledge transfer based on: identification of knowledge (which can be done through secondary studies), transfer to the medium (e.g., using evidence briefings), and contextualization of evidence (for this the authors propose the use of Bayesian synthesis) [3, 4].

Finally, another two recent publications have explicitly suggested or proposed that EBSE may be a valuable resource to bridging the gap between industry and academia.

Devandu et al. investigated the beliefs of Microsoft programmers and how these were related to current empirical evidence [19]. To achieve this, they conducted a survey of 564 Microsoft workers and a case study. They reported that practitioners placed more importance on personal experience and the experience of colleagues than on empirical evidence. Therefore, the authors suggest that, given its successful application in medicine, knowledge of EBP might help practitioners to place more trust on verified evidence.

In an opinion article, Le Goues et al. argued that it is possible to improve the use of research in SE, and to obtain better practical benefits, by better organizing and synthesizing its results [39]. In summary, they suggest the use of EBSE associated with (1) achieving consensus on a formal framework of levels of evidence (i.e., outcomes from secondary and primary studies) and mechanisms to determine the level of confidence in the research results (taking into account types of methods, execution of studies and strength of evidence), (2) clearly identifying the methods and results of studies, (3) encouraging the publication of secondary studies and (4) educating software engineers on how to use the proposed framework.

To summarize, although there are several proposals to use EBSE to better connect academia with industry, we do not know of any

that propose EBSE training as the main strategy nor have we found any studies of the impact of EBSE training on professional practice.

3 EBSE TRAINING MODULE FOR UNIVERSITY STUDENTS

In our university, Universidad de la República, we have a Computer Science (CS) curriculum (a five-year degree comparable to the IEEE/ACM's proposal for the CS undergraduate curriculum [27]). The program comprises 450 credits with certain minimums by areas, e.g. 70 credits in Mathematics. One credit corresponds 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.

In 2017, we introduced a non-mandatory EBSE and SR course. To take our EBSE course, students must have passed the mandatory course on SE. So, students would take this course during the fourth or fifth year of their degree (270 credits approx). In addition, the EBSE course gives students 7 credits upon approval.

We aimed to teach EBSE fundamental concepts and techniques for practical use. After the course, students should understand fundamental EBSE concepts, identify SE issues that may be addressed by using evidence, assess published secondary studies on SE, and participate in the conduct of SR. A limitation of our course was that it emphasized conducting SRs more than EBSE [48].

We used learning outcomes (LOs) to guide both the design of the course and the method we use to assess it. LOs specify what students are expected to know, understand, or be able to demonstrate after the course [32]. Due to this paper's focus, we do not discuss the LOs in detail, instead, we describe the course, and the results, in terms of topics (which group several LOs). The details of the LOs used for the course can be found in our previous study [48]. However, we present two LOs as an example (they should be read as knowledge or skills that the student will achieve by the end of the course).

- Describe the protocol sections of an SLR.
- Participate in the selection stage of primary articles for an SLR with multiple reviewers

In summary, we specified more than fifty LOs that defining a syllabus that aims to promote the practical application of EBSE, this includes:

- Basic aspects of scientific publications (e.g., how to distinguish between scientific papers from other types of publications) and introduction to research in SE (e.g., what research methods are commonly used in SE and for what purpose). These topics are necessary to understand EBSE and it is necessary to include them because many students do not know them when taking our course.
- Introduction to the evidence-based paradigm and characteristics of SRs in SE. Based on chapters 1 to 3 of the reference book on EBSE and SRs [34].
- The process of an SR and characteristics of each of its stages. Namely, SR planning, search for primary studies, study selection, study quality assessment, data extraction from primary studies, analysis in mapping studies, qualitative syntheses, reporting of an SR, and knowldege translation and diffusion). Based on chapters 4-10, 12 and 14 of the EBSE book [34].

The course was organized as an alternating introduction of theoretical and practical content and a weekly follow-up of the students' progress in their team assignment (i.e., the conduct of a secondary study). In addition, based on the difficulties and recommendations reported in previous EBSE training studies [5, 7, 15, 16, 29, 33, 38, 45, 49], we chose the following principles to develop the course:

- The students' team assignment workload would be limited in some stages of the EBSE process.
- Students would be assisted by teachers to choose their review topic.
- Students who needed to perform iterations in some stages of their team assignment would be supported.

The course was 14-weeks long and had one non-compulsory on-site class a week that is 3.5-hours long. Table 2 shows the topics of the classes and the team assignments scheduled for each week. During weeks 11-13 students work on completing their team assignments.

As a practical team assignment, students were required to define and conduct guided activities for an SR. Each team chooses the topic of their SR and the research questions, according to their own interests. The teachers guided them in their selection so that the work, both in scope and complexity, could be tackled in the time available. The teams were made up of two or three students who worked together throughout the course and delivered the final report of their SR in week 14. The weekly class covered the theoretical content of a stage of the SR, which the students then applied to their own SR before the next class.

Most classes had two main parts: a lecture by a teacher and a reserved time for teamwork and questions to teachers. The lectures usually took less than an hour during which the teacher explained the main concepts of the SR stage that were studied that week. Students were asked to pre-read the material, which was made up of chapters from the EBSE book [34]. Subsequently, the teacher presented the weekly assignment task, which consisted of completing the activities of the SR stage previously discussed.

Students could ask questions or present problems they had about the assignment for the current or previous weeks. The two teachers answered questions and were present in the classroom for the entire class. Students were expected to spend four hours weekly outside of the classroom to complete the required assignment. To discuss their problems during the course, students could use a Moodle platform site, where teachers published material and answered questions. The basic format of the course was unchanged through all three years. In 2018, the course content was extended to introduce rapid reviews and evidence briefings, but otherwise unchanged.

Sebastián Pizard was not only responsible for the design and evaluation of the course, but he was also lead researcher of the course evaluation and one of the two teachers on each of the three courses. During 2017 and 2018 he was accompanied by Fernando Acerenza and in 2019 by Cecilia Apa. All three are active researchers with completed master's degrees and eight years (or more) of experience of conducting lectures and tutorials.

4 RESEARCH GOALS AND METHOD

The overall objective of our research program is to contribute to the understanding of training as a facilitator for the adoption of EBSE.

Table 2: Course timetable, including class topics and student assignments.

Week	Topics	Team assignments
1	Basic aspects of scientific publications, Introduction to SE research, Evidence-based paradigm, SRs in SE	Classify the sections of a scientific paper.
2	Planning an SR	Establish the purpose and the need for the SR to be performed by each team. Propose and validate the research questions.
3	Searching for primary studies	Define the search strategy for the SR. Create and validate the search string.
4	Study selection	Define inclusion and exclusion criteria for the SR. Define and conduct the selection process, obtaining 20-30 primary studies per student.
5	Study quality assessment	Define and conduct the quality assessment procedure for the selected primary studies.
6	Data extraction from the studies	Define an extraction form and use it to extract data from the primary studies.
7	Mapping study analysis	Classify primary studies according to commonly used schemes and schemes relevant to the research questions.
8	Introduction to data synthesis, Qualitative synthesis	Use qualitative synthesis to answer the research questions, considering the limitations of the review process.
9	Reporting a systematic review	Report the results and the whole process.
10	Knowledge translation and diffusion	-
14		Deadline for team assignment

In this article, we extend the evaluation of our training module reported in our previous study [48] to include data from two more years, and we report the impact of the training on the participants' subsequent working practices. Specifically, our research questions are:

RQ1: Is our EBSE course adequate to train undergraduate students?
RQ2: Does our EBSE course have any impact on the working practices of the students?

We consider that course adequacy (RQ1) is directly related to the successful achievement of the predefined LOs by the students. These LOs (and the teaching methodology used) are intended for students to understand EBSE and apply it by conducting the steps of the SLR process. The achievement of these LOs (as explained below) was assessed through teacher evaluations of student's assignments and the perceptions of the students themselves.

Similarly, we consider as impacts of the EBSE course (RQ2) any change in attitudes and behaviors of students caused by what they learned. The possible impacts were investigated surveying the students, with which the results are associated with their perceptions. Furthermore, we cannot isolate our course from the rest of the curriculum, with which it could be thought that students would exercise some EBSE practices in other courses. From our knowledge of the curriculum, we understand that there are no other courses that include our syllabus and that this should not be a threat to our work.

We treat our study as a *longitudinal case study*, which, according to Yin, seeks to study *'how certain conditions and their underlying processes change over time'* [63]. Fucci et al. argue that longitudinal studies are useful to study a particular event and its impact, for example, the introduction of a practice in a company [21]. This research method enables studying dynamics of change in complex and changing contexts such as those related to software engineering [44], and, therefore, to investigate the effects of learning EBSE.

The collection of data in different waves (i.e., separate time periods) allows better understanding of the evolution of the impact of training. In particular, we can study both their initial impressions of EBSE and any subsequent impact on their professional practice.

The data collection waves carried out are summarized in Figure 1. Since it is a longitudinal study, certain aspects of our research design are worth mentioning [20]. The amount of time between the waves was chosen for convenience and based only waiting a reasonably long period between the end of the course and circulating our surveys, because we wanted to allow enough time for participants to encounter opportunities to use their training. We sought the participation of all the students of the three courses. Data collection was always anonymous (the tools used allowed checking a single response per subject). Anonymous collection has certain advantages [2]: (1) participants respond more honestly if they trust that they cannot be identified in any way, (2) anonymous data reduces confirmation bias, and (3) anonymous data collection helps to comply with ethical and legal requirements. The most notable disadvantage, in our case, is that we cannot perform intra-individual comparisons across time. We consider that the design used was adequate to meet the research objectives, which are exploratory, while also encouraging participation. The unit of analysis used for most analysis was data from individual students. However, some course assessments were based on the assessment of team projects.

To improve reliability (e.g., to avoid misinterpretation of the data), we used a rigorous pre-planned analysis process. We also used two approaches proposed by Runeson et al. to improve reliability [53]: systematic tracking of all data, and peer debriefing (i.e., participation of at least two researchers). In addition, to enhance reporting clarity and completeness, we used the O'Brien et al. checklist for reporting qualitative research [46].

	Course evaluation surveys & focus group Teachers' course assessments Nov 2017, Nov 2018 & Nov 2019	First follow-up survey Jun 2020	Second follow-up survey Aug 2021
Sample	44 students (100%)	37 students (84%)	29 students (66%)
Adequacy (RQ1)	Evaluation of training.	Evaluation of training after use.	
Effects (RQ2)	Study of perceived benefits	Study of new skills usage experiences and changes motivated by the training	Study of usefulness of new skills and frequency of use.

Figure 1: Summary of research waves.

5 PARTICIPANTS, DATA COLLECTION AND DATA ANALYSIS

This section describes the participants, and presents details about data collection and analysis.

The data collection and analysis procedures can be separated into those performed for the course delivery data and those carried out in the follow-up surveys. All of the surveys were designed and conducted following guidelines and recommendations proposed in the literature [31, 59].

5.1 Subjects Selection and Ethical Issues

Each student was the main unit of analysis, however, on several occasions, it was necessary to analyze student teams, e.g. while assessing their practical assignment. Our course is non-mandatory and students were encouraged to take it by a typical course information entry on the university website.

During the first class of the course, the teachers explained to the students that they were researching on EBSE training, and reported the aims and the data collection procedures that they planned to use. They were also told that the information collected was to be managed confidentially (i.e., course assessment reports would not link grades or test scores to individual students or teams), that participation or not in the study would not affect their learning experience or evaluation, and that they could withdraw from the study at any time without leaving the course. For each of the three courses, all students agreed to participate voluntarily and signed an informed consent form.

When asked to answer each of the two follow-up surveys, the students were also reminded about the objectives of the research, the need for accurate and honest answers, and the confidentiality of any reported information.

5.2 Courses data collection and analysis

During each course (2017, 2018, 2019), we collected quantitative and qualitative data. The data include the students' opinions - collected through a survey (will be referred to as SU) and a focus group (FG) - as well as the academic results obtained from the learning assessments (LA). The survey and the focus group were carried out in the last class. The learning assessments include grades achieved in team practical assignments and individual written tests.

The analysis presented in this paper does not include the individual written tests. In these tests, the results obtained were quite satisfactory in all the courses, although not comparable, since their questions sought to evaluate different learning outcomes.

The details of the data analysis process can be found in our previous study [48]. As significant changes we can report:

(1) We counted students' responses to whether they had work activity or not and their weekly workload (all obtained through SU). Work positions were analyzed and classified using descriptive coding and the constant comparative method.
(2) Students' opinions (SU) and teachers' assessments (LA) of the achievement of learning outcomes (initially expressed by using a five-point agreement scale) were grouped and counted as positive or negative.
(3) Students' opinions on the benefits of the training (SU) were grouped according to similarity. Opinions that indicated benefits in their work practices were analyzed using descriptive coding and the constant comparative method.

5.3 First follow-up opinion survey

Seven months after finishing the last course we conducted the first follow-up survey. The main purpose of the survey was to collect the former students' opinions about any impact the EBSE course had on their work and academic practices. It is worth mentioning that at that time we wanted to know any impact of the course, although later in the following survey we asked only for effects on professional practices.

The three authors defined the purpose of the post-course survey and the initial draft of the survey. Subsequently, Pizard created the questionnaire, which was reviewed by Kitchenham and Vallespir. The self-administered questionnaire had six questions. We used a set of closed and open questions to allow participants to explain their answers more completely. The questionnaire was designed in SurveyMonkey and was available from June 23 to July 2, 2020.

The translation of the survey questions (originally in Spanish) is reproduced below, the flow of the survey is presented within square brackets.

F1 What year did you take our EBSE course?
- 2017 / 2018 / 2019

F2 After finishing the course, have you used any of what you have learned in your professional or academic practice? For example, some of the following activities.
- Conducting a secondary study (systematic review, mapping study, or rapid review).

- Using an existing secondary study (systematic review, mapping study, or rapid review).
- Using search engines for scientific articles.
- Undertaking a critical appraisal of reports that compare different technologies.
- Other activities, indicate which.

F3 *[If at least one option selected in F2]* Do you think the course adequately trained you for those activities? Please explain your answer in detail.
- Yes / No / Comments

F4 *[If empty answer to F2]* Do you think that your knowledge about EBSE could be used in the future? For what?

F5 Did knowing about EBSE MOTIVATE you to improve or change any activity in your professional practice related to software development?
- Yes / No

F6 *[If answer to F5 was 'Yes']* In which activities, tasks, problems or situations do you perceive that you have changed after learning about EBSE?

To analyze F1, F2, F3, and F5 we counted the number of responses in each category. In F3, F4, and F6 the textual responses were collated and we used descriptive coding and constant comparative method to classify the responses.

Initially, Pizard performed all of the analysis described above. Then, Vallespir verified the result of the quantitative responses, reviewed the coding of the F1, F2, and F3, and did a separate analysis of F6. Subsequently, in two meetings they both reached an agreement on the coding of all the questions. Finally, Pizard made the grouping of the final codes and description of the results.

As an example of the qualitative analysis carried out, Figure 2 presents the coding of some F6 responses, including the codes identified by each researcher and the result of the agreement meeting.

Id	Student response	Researcher 1	Researcher 2	Final agreement
F6R11	[In] looking for bibliography of reliable sources, and it improved my judgment for choosing papers.	search information from reliable sources evaluate information	Activities with changes: -- Information search -- Reliable sources -- Evaluate information	- Search for information -- From reliable sources - Analyze and evaluate information
F6R18	For example, before I used to assess certain methodologies in software engineering based on my own perception or comments from other people, now I try to read more studies to have a more educated opinion, for example, in which scenarios it's really useful to apply TDD.	value of evidence	Value of evidence Activities with changes: -- Study reading -- Evaluate technologies	- Awareness of value of evidence

Figure 2: Fragment of coding for question F6 ('Changes motivated by knowing EBSE').

5.4 Second follow-up opinion survey

Finally, twenty-one months after finishing the last course we conducted a second follow-up survey. With this survey, we investigated whether the students had used their training in professional practice (rather than an academic context) identifying which skills had been used and how often specific skills were mentioned.

Again, we designed a self-administered questionnaire with a set of closed and open questions. The questionnaire was designed in Google Forms and was available from August 27 to September 3, 2021. The translation of the questions (originally in Spanish) and the flow of the survey are shown below.

S1 Are you working or have you worked in the software industry?
- Yes / No

S2 Have you found useful in your professional (NOT academic) practice what you learned in the EBSE and SRs courses?
- Yes / No

S3 *[If answer to S2 was 'Yes']* Indicate what news skills from the course you used or found useful. If you can, add examples.

S4 *[If answer to S2 was 'Yes']* On a scale from 0 (not at all) to 10 (all the time), how would you rate the frequency in which you use or have used things learned in the course in your professional practice?
- 0 / .. / 10

We counted the answers to each category of the closed questions. The textual responses to the open questions were collated and we used descriptive coding and the constant comparative method to classify the responses.

Initially, Pizard performed all of the analysis described above. Subsequently, Vallespir reviewed the results. In a meeting, differences were discussed until an agreement was reached. Finally, Pizard did the final aggregation of codes and the description of the results.

6 RESULTS

Figure 3 shows demographic information describing our students obtained from the end of course survey. In particular, Figure 3 reports the percentage of male and female students, their age range, and details of any paid employment. Three students reported having more than one position (e.g., team leader & software developer), which we assume means they take on different roles in different projects. The vast majority of the students had worked in industry, and most of the positions that they held were technical rather than managerial. Our university is public and free of charge, and there is a lot of demand for IT workers. This is why most of our undergraduate students have part-time or full-time jobs in their last two years of studies.

6.1 RQ1: Is our EBSE course adequate to train undergraduate students?

The evaluation results for each course and for the first follow-up survey are summarized below.

Courses evaluations. The three courses were taught one per year, from 2017 to 2019. Students' attendance was 10, 18, and 16 respectively. All students passed the module.

To assess the adequacy of our course we investigated (1) the levels of achievement of the learning outcomes, both perceived by the students and those assessed by the teachers, and (2) the

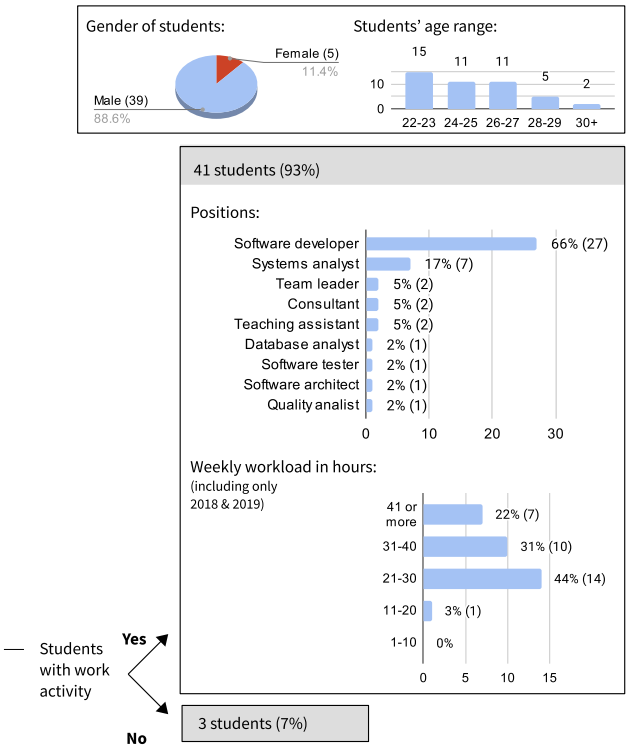


Figure 3: Students demographics

students' opinions of the course, based on the survey and the final focus group.

First, we studied students' opinions and teachers' assessments of the achievement of learning outcomes, which were initially rated using a five-point assessment scale. Assessments allocated to scale points 3, 4 and 5 indicated that successful understanding and use of a topic was achieved[1]. The subject success rate for a topic was calculated as the percentage of subjects who judged their own personal achievement to be successful for that topic. The team success rate was calculated as the percentage of team projects that were judged as successful for a specific topic by the course teachers. The left section of Table 3 shows the student-assessed success rates and the right section shows the teacher-assessed success rates for the team project. Topics that were not evaluated in the team assignments were left blank.

Looking at the self-assessment results, the learning achievement levels of all the topics seem quite good, since success levels are 65% or above in all years. Some topics seemed consistently difficult to learn. Study selection, qualitative synthesis, and knowledge translation and diffusion were the three topics that some students in each year found relatively difficult. With respect Knowledge translation, a limitation of the course was that it did not include any practical examples. For example, one student pointed out *'I consider that I couldn't adequately address these objectives [those related to Knowledge translation] because, since there is no practical instances [in the course], and only remain theoretical, I couldn't understand it well'*. For the other two topics, we did not find specific causes that explain why success was low nor why succes rates varied over the years, and we assume it is a natural result of dealing with different courses and different cohorts of students.

[1]The five-point agreement scale used was: 1. Not achieved at all | 2. Very little achieved | 3. Satisfactorily achieved | 4. Almost completely achieved | 5. Completely achieved

Table 3: Student self-assessments of their learning achievements and teachers assessments team-based learning achievements.

Topic	Student self-assessed success rate			Teacher-assessed project success rate		
	2019	2018	2017	2019	2018	2017
Basic aspects of scientific publications	96%	91%	97%	-	-	-
Evidence-based paradigm	90%	96%	95%	-	-	-
SRLs in SE	100%	100%	100%	-	-	-
Planning	97%	100%	95%	94%	81%	92%
Search	93%	94%	98%	100%	100%	100%
Selection	84%	86%	92%	92%	93%	100%
Quality assessment	88%	97%	100%	83%	64%	100%
Extraction	98%	100%	100%	100%	100%	100%
Mapping study analysis	94%	93%	97%	83%	86%	75%
Synthesis	93%	89%	94%	100%	71%	100%
Report	100%	100%	100%	100%	86%	100%
Knowledge translation	65%	80%	80%	-	-	-
SR process	100%	100%	100%	100%	100%	100%
SE Research	95%	94%	-	-	-	-

The number of observations corresponds to the following detail. In 2017, 10 students & 4 teams. In 2018, 18 students & 7 teams. In 2019, 16 students & 6 teams.

In the teamwork assessment, both planning and mapping study analysis caused problems for a minority of teams in all three years.

Thus, overall, learning achievement levels were mostly satisfactory, and problem topics were relatively consistent.

Students from all three courses expressed general satisfaction with the training. In all the focus groups held at the end of each course, the students expressed a positive opinion about the weekly work dynamics, i.e., the introduction of theoretical concepts and practical work in teams. They also valued positively the role of a practical assignment within the general course marking process.

Post-use evaluation. In the first follow-up survey, seven months after finishing the last course, we asked the students if they had used lessons learned from the course and if the training received had been adequate for that use (see first survey, question F3).

All students who reported using EBSE (89% of the total) indicated that the course trained them adequately. Five students added comments on the adequacy of the course. Four students confirmed in their comments that the course was complete. Although two added that it was complete considering it was a short course and another that it seemed complete given their superficial use of EBSE. Finally, another student indicated that the workshop approach and teamwork were aspects that greatly helped learning.

6.2 RQ2: Does our EBSE course have any effects on the working practices of the students?

The effects of EBSE training on students' attitudes and behaviors were studied in three instances: in the final survey of each course and in the two follow-up surveys. The results obtained are presented following the same sequence.

Perceived benefits of the course. The students' opinion about the benefits of training[2] is grouped into three perspectives. Figure 4 summarize their opinions. It also includes the analysis of the responses that indicated benefits in the work practices, since they are the results of interest to answer our research question.

First, 45% of the students had, what we considered is, a pragmatic view and felt that the course helped them understand EBSE or how to conduct an SR. A student summarized the training benefits as follows: *'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'.*

Second, 30% of the students reported that after taking the course they have a tool to carry out more reliable research in software engineering. Two students thought that they did not see it as applicable in the industry, one of them explained: *'Although we learned an interesting method for research in software engineering, for now I do not see it as very applicable in my working life'.*

Finally, 25% believe that the course prepared them for better professional practice. In this regard, a student said: *'On the one hand, [the course allowed me] to become more aware of all the baseless decisions that are made in our field. On the other hand, [I achieved] a greater knowledge about scientific studies and the different search engines'.* The most perceived benefits were: improving information searching skills, having a new tool to support decision-making, and being aware of the value of the evidence.

First follow-up survey. The survey was answered by 37 students, out of a total of 44 who took one of the courses (i.e., a response rate of 84%). Figure 5 summarizes the results related to the impact of the EBSE course on the former students' behavior. To investigate this impact we used three students' views: experiences of using concepts learned in the course, their perspectives of using concepts in the future, and changes the course motivated to their working practices.

EBSE usage experiences. The majority of students (89%) indicated having used concepts learned in the EBSE and SRs course (see the upper part of Figure 5).

Six students added answers in the free text option 'Others'. One of them did not answer the question so it was discarded. Of the other five responses, all referred to the use of what was learned in the course (e.g., reading SRs or execution of some stages) in academia, that is, in their capstone project (4) or in other courses (1). We accumulated and highlighted in red in Figure 5 any of these responses that semantically correspond to a category not previously selected by the student.

The most frequently performed activity was using search engines to retrieve scientific articles, 73% of those surveyed indicated having done this after completing the course. Using a secondary study and critically evaluating reports that compare technologies were the next two most performed activities (38% and 35% of respondents respectively). Lastly, conducting a secondary study was the least frequently performed activity, although it was done by 6 (16%) of the students surveyed.

Prospects of using EBSE concepts Four students (11%) indicated that they had not used concepts they learned in the course. Three of them responded about prospects for using ideas from the course in the future. One student expressed that a scientific approach is very valuable in decision-making but concludes that it is difficult to apply outside academic environments. The other two students thought that their EBSE knowledge could be useful in the future: to investigate industry-related topics, as a tool in the capstone project, and to analyze scientific papers).

Changes motivated by knowing EBSE. Almost half of the respondents (49%) indicated that knowing EBSE motivated them to change their working practices. The activities in which the students noticed changes are shown in the lower part of Figure 5.

For 30% of the surveyed students, knowing EBSE motivated them to change the way they seek information. Some of them reported using EBSE to search for information for new or unusual problems, on topics with little information available or on topics about new technologies. Others said that the new knowledge allowed them to search more comprehensively or to use more reliable sources. One of them recognized changes in: *'looking for bibliography of reliable sources, and it improved my judgment for choosing papers'.*

Students also identified information assessment and the value of the evidence as activities that were positively influenced by the course (22% and 14% respectively). Both of these skills are related and indicate enhancement in the students' abilities to evaluate the information they collect or receive. One student said: *'For example, before I used to assess certain methodologies in software engineering based on my own perception or comments from other people, now I try to read more studies to have a more educated opinion, for example, in which scenarios it's really useful to apply TDD'.*

Some students also indicated improvements in their technical writing skills (14%), problem-solving skills (5%), and communication skills (5%). To illustrate the latter, some students reported being able to better communicate or substantiate their ideas. These improvements may not solely be due to the knowledge of EBSE, but may also be due to the methodology and activities of the course.

Finally, one student reported a better understanding of the scientific research process, and another to have carried out a secondary study in their academic activity.

Second follow-up survey. The second survey was answered by 29 students (i.e., a response rate of 66%). All indicated that they were working or had worked in the software industry (S1).

A 55% of respondents found what they learned in the EBSE course useful in their (non-academic) professional practice (S2). Figure 6 shows skills from the course that respondents indicated were useful or used (S3), and the frequency of use (S4). In S3, many students did not indicate skills learned in the course that they use (what was requested) but the activities improved by the training. Since both correspond to the effects of the training, our analysis of that response includes both useful new skills and activities enhanced by

[2]The question asked to the students was: *'[SU-C1] Indicate, according to your criteria, what are the main contributions of the course to your education'.*

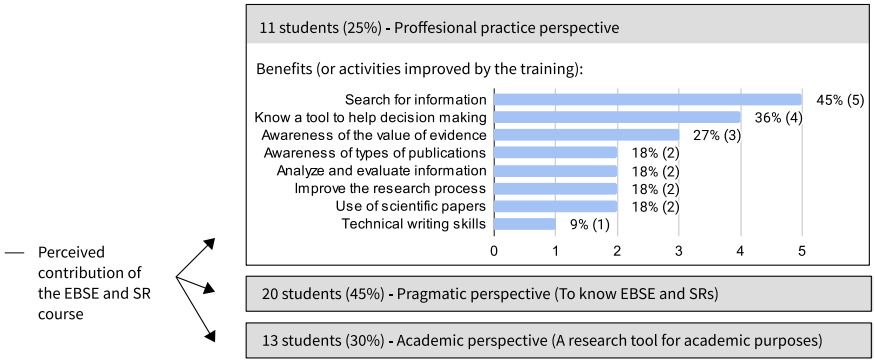


Figure 4: Perceived benefits of the course (Nov 2017, Nov 2018 & Nov 2019).

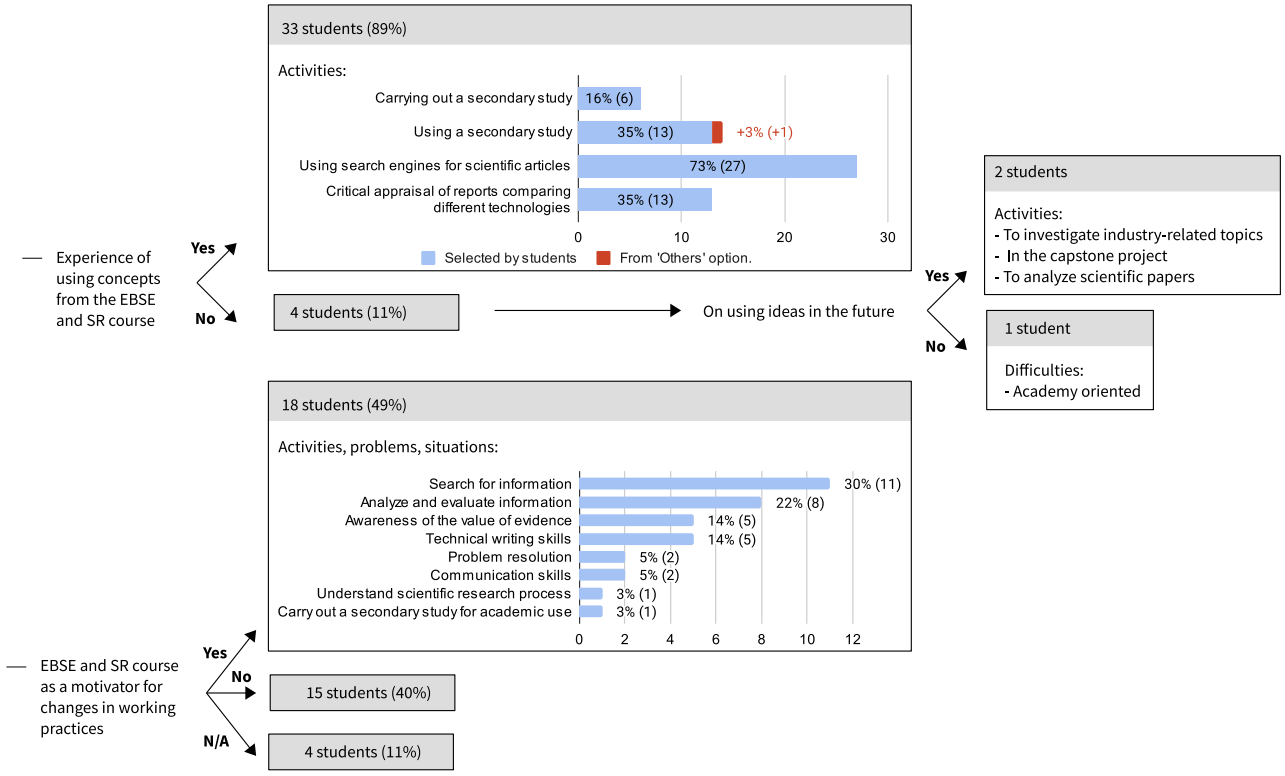


Figure 5: Effects of the course on students' working practices (June 2020).

the training. One of the answers to S3 did not answer the question so it was discarded.

Four out of ten of those surveyed said they have improved the way they search for information after taking the course. On this a student commented that now they can: *'recognize useful information, from reliable sources and carry out more precise searches'*.

A fifth reported using scientific literature after taking the course. One former student said: *'I learned to use scientific articles as a source of information and evidence on a day-to-day basis (and how to search for them correctly) [...] It also helped me to look for reviews at times I wanted to know quickly the state of the art on a specific topic'*. Related to this, two students also reported understanding the scientific research process was a useful new skill.

As in the previous survey, some of the former students reported improvements in their abilities to assess and value the information, both in the evidence assessment and in the analysis and evaluation of information. Two students also indicated that they were more aware of the different types of publications. One of them claimed: *'I also believe that the training from the course allows me to identify the*

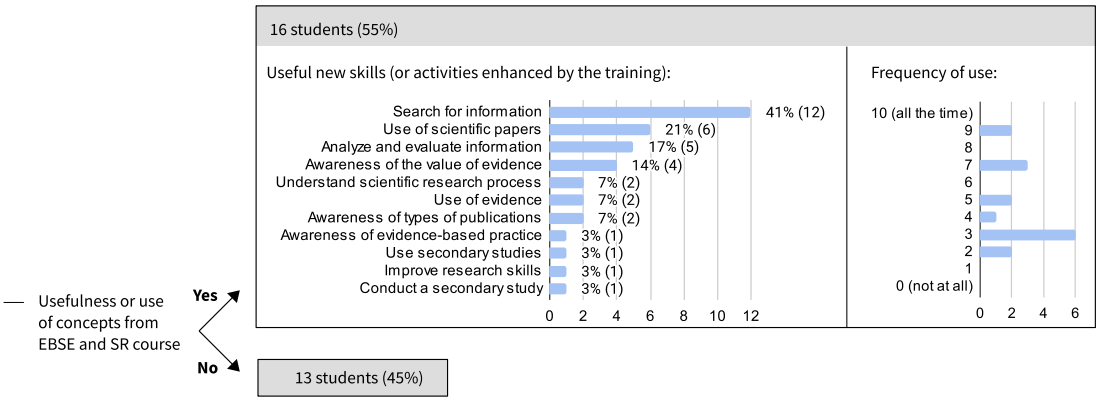


Figure 6: Effects of the course on students' working practices (August 2021)

different types of papers, whether they are primary studies, reviews, etc. and this is essential when looking for information'.

Some students also reported using evidence. In our course, we do not teach concepts and techniques to address in detail the use of evidence and students may not know if they are actually using evidence. This is a non-trivial problem, for example, the term evidence is already controversial in other areas [54].

Other impacts of the EBSE training were reported to be: awareness of the evidence-based practice, using secondary studies, improving research skills, and conducting secondary studies.

6.3 Discussion of findings

Several aspects of our results deserve reflection and analysis.

Course adequacy. Our training was perceived as adequate by the teachers when assessing students' teamwork assignments and by the students themselves at the end of each course and after they tried to use what they learned in practice. In the latter case, all students who reported using what they were taught found the training adequate.

In health care and medicine, there is a greater formalization of the teaching of evidence-based practice (EBP), for example, there are proposals for EBP core skills (see for example [1]) and recognized tests for the minimum skills required [51]. These disciplines have stringent accreditation requirements because of their direct impact on human welfare. However, having such tools in software engineering could help us to better and more consistently assess the adequacy and the effects of EBSE training. Our training proposal is based on the concept of learning outcomes (see [48] for the full description) that could be used as a starting point for discussions on this issue.

Effects of EBSE training. Although few students reported having read or participated in the conduction of secondary studies, their responses confirm that the training allowed them to learn and apply the principles of evidence-based practice. The most reported effects in both surveys were: improvements in information search and analysis skills, and awareness of the value of the evidence. In the second post-course survey, the use of scientific papers was also frequently mentioned. The students reported different frequencies of use of the EBSE-related skills, but overall usage appears to be more than sporadic.

In brief, our EBSE training provided the students with an awareness of research and scientific evidence and improved their information gathering and information literacy skills. All of these effects are consistent with previous research assessing the impact of EBP training. For example, several nursing studies report that EBP training improves students' search skills and information literacy skills [23], and also improves students' research skills and their confidence in research [47].

The effects of our training are consistent with the initial opinions of the students about the benefits of the course. Also, they are consistent with the previous studies on the benefits of learning EBSE (i.e., improvements in research skills, awareness of the value of adding evidence, and improvement in the search and organization of information [5, 33]). This could indicate: (1) a relationship between initial attitudes to EBSE and subsequent use of EBSE-related skills, and (2) that opportunities for individuals to reflect on the benefits of a training course could help them to identify ways to use EBSE related-skills at a later date. The first issue is under study in other areas of EBP. For example, it has been shown that knowledge of EBP and practitioners' attitudes towards EBP influence its adoption [58] and that attitudes towards EBP are considered key for its adoption [62]. The second issue requires further research, for example, by exploring reflective instances in groups and investigating whether early adopters can influence other students.

Our results seem to indicate, we believe for the first time, that EBSE training makes practitioners more confident in the value of scientific evidence and fosters its use for support decision-making. To our knowledge, this is also the first study that provides support for the potential of the proposals (i.e., [19, 39]) to use EBSE to bring practitioners closer to scientific evidence and collaborate in closing the gap between academia and industry.

Future of EBSE education. Currently, curricula guides for undergraduate students in CS and SE do not consider evidence-based practice [27, 28]. Certainly, more studies are needed to justify the inclusion of EBSE, for example, in crowded curriculums common in Europe and the USA.

However, in other disciplines, EBP teaching is more strongly promoted given the benefits of its use by practitioners. For example, evidence-based practice has been included as a core component of the curriculum of undergraduate, postgraduate, and continuing education of health programs throughout the world, and many accreditation boards also expect all clinicians to have competent training in EBP [1].

It is also necessary to point out that, from our understanding, each discipline has its own tailored content and approaches to teaching evidence-based practice. This seems to be also the case in our field, although EBSE appears to be similarly teachable to SE and CS students. This study serves as a sample of this since in our university there is a unique curriculum (in particular, CS focused).

Given the continuous renewal of our discipline, it may be necessary to reimagine what skills tomorrow's software engineers will require and explore ways to bring their practice closer to scientific evidence. To our knowledge, there is a lack of studies on the benefits of adopting EBSE in non-academic environments. Even so, we believe it is important to begin to discuss the inclusion of EBSE teaching in curricula guides, since, as our results show, its teaching can contribute to its adoption and to the use of scientific evidence in professional practice. As Mary Shaw, in the first prospects on the teaching of SE, claimed: *'The greatest danger to software engineering curriculum designers is lack of imagination'* [57].

6.4 Limitations of this study.

Although the results reported in this study have addressed the fact that our previous evaluation of the training module was based on a small sample, many of the limitations discussed in our previous paper [48] apply equally to this study. In particular, the limitations related to the good subject problem among voluntary participants, using incentives to encourage participation at the end of each course, and the course emphasis on SRs rather than EBSE apply also to this study.

Question F6 of the first follow-up survey caused some confusion. The question is a continuation of question F5 and it sought to investigate which activities of their working practices the students had improved from learning EBSE. Two responses out of a total of 18 included references to academic (and non-professional) activities. Many other answers did not indicate whether they correspond to professional or academic activities. Nonetheless, we did not discard data during analysis, because it confirms that EBSE training has general benefits to undergraduate students. In the second follow-up survey we specifically investigated benefits related to the students' industry-related working practices.

Surveys rely on respondents' reflection and reporting on their attitudes and behaviors, and this can bias the results in different ways [41]. For example, people tend to remember events that are most important to them. Although we tried to mitigate its effects by asking participants several different questions about the same issue and requesting examples to support their previous answers, our results must be considered with this limitation in mind.

7 CONCLUDING REMARKS

Our study reports that EBSE training has positive impacts on the professional practices of those trained. In particular, the majority of our EBSE course students reported using their training, not just in an academic context, but also in their jobs in industry. They reported a greater awareness of research and evidence, and improvements to their information search and analysis skills. As a consequence, EBSE training could be useful to train practitioners in concepts and techniques of evidence-based practice but also to foster collaboration between industry and academia.

The major novelties of our study are that it:

- Involves the delivery of three courses of EBSE and SRs to university students based on a predefined set of learning outcomes.
- Is an attempt to evaluate the impact of an EBSE training on the attitudes and behaviors of the trainees (particularly, in their work practices) in different periods.
- Confirms that EBSE training has enabled more than half of students to enhance their work practices, by improving their information gathering and analysis skills, and through better awareness of evidence and research.
- Confirms that the benefits of teaching EBSE are similar to those obtained from teaching EBP in other disciplines.
- Suggests that EBSE training improves the performance of novice industry practitioners not just individuals with conventional decision-making roles such as senior software engineers or project and quality managers.

EBP training in medicine has been shown to be more effective if it is integrated or supported with clinical practice [24, 37], also bringing practitioners even closer to scientific evidence. This could be achieved in EBSE training, for example, by teaching EBSE using scenarios of use of evidence created together with practitioners (as suggested by Manns and Darrah [43]). Knowing the decision-making process in the software industry seems to be the key to orienting EBSE training to the needs of the industry. To address this, it could be useful to survey managers of software companies asking them who makes decisions about SE methods and tools and how such decisions are made. It has been suggested that analyzing and making the decision-making process visible can improve EBP adoption [6]. In addition, using theoretical proposals for studying the decision-making process (e.g., [10]) could also contribute in this direction.

ACKNOWLEDGMENTS

First of all, we thank all the students who participated in our EBSE courses. We thank Fernando Acerenza and Cecilia Apa, of the School of Engineering, Universidad de la República, Uruguay, for their collaboration as teachers in the 2017-2018 and 2019 EBSE courses, respectively. We also thank Ximena Otegui, of the Teaching Department, School of Engineering, Universidad de la República, Uruguay, for her collaboration in creating and improving the LOs of the course. Finally, we thank the reviewers for their helpful comments, which have allowed us to improve our manuscript.

REFERENCES

[1] Loai Albarqouni, Tammy Hoffmann, Sharon Straus, Nina Rydland Olsen, Taryn Young, Dragan Ilic, Terrence Shaneyfelt, R. Brian Haynes, Gordon Guyatt, and Paul Glasziou. 2018. Core Competencies in Evidence-Based Practice for Health Professionals. *JAMA Network Open* 1, 2 (2018). https://doi.org/10.1001/jamanetworkopen.2018.0281

[2] Lillian M. Audette, Marie S. Hammond, and Natalie K. Rochester. 2020. Methodological Issues With Coding Participants in Anonymous Psychological Longitudinal Studies. *Educational and Psychological Measurement* 80, 1 (2020), 163–185. https://doi.org/10.1177/0013164419843576

[3] Deepika Badampudi, Claes Wohlin, and Tony Gorschek. 2019. An Evaluation of Knowledge Translation in Software Engineering. In *International Symposium on Empirical Software Engineering and Measurement*, Vol. 2019-Septe. IEEE, 1–6. https://doi.org/10.1109/ESEM.2019.8870165

[4] Deepika Badampudi, Claes Wohlin, and Tony Gorschek. 2019. Contextualizing research evidence through knowledge translation in software engineering. In *International Conference on Evaluation and Assessment in Software Engineering*. 306–311. https://doi.org/10.1145/3319008.3319358

[5] Maria Baldassarre, Nicola Boffoli, Danilo Caivano, and Giuseppe Visaggio. 2008. A hands-on approach for teaching systematic review. In *International Conference on Product-Focused Software Process Improvement*. 415–426.

[6] Angela Benfield and Robert B. Krueger. 2021. Making decision-making visible—teaching the process of evaluating interventions. *International Journal of Environmental Research and Public Health* 18, 7 (2021). https://doi.org/10.3390/ijerph18073635

[7] Pearl Brereton. 2011. A study of computing undergraduates undertaking a systematic literature review. *IEEE Transactions on Education* 54, 4 (2011), 558–563.

[8] Pearl Brereton, Mark Turner, and Rumjit Kaur. 2009. Pair programming as a teaching tool: a student review of empirical studies. In *Conference on Software Engineering Education and Training*. 240–247.

[9] David Budgen, Pearl Brereton, Sarah Drummond, and Nikki Williams. 2018. Reporting systematic reviews: Some lessons from a tertiary study. *Information and Software Technology* 95, November 2016 (2018), 62–74. https://doi.org/10.1016/j.infsof.2017.10.017

[10] Janet E. Burge, John M. Carroll, Raymond McCall, and Ivan Mistrik. 2008. *Decision-Making in Software Engineering*. Springer Berlin Heidelberg, Berlin, Heidelberg, 67–76. https://doi.org/10.1007/978-3-540-77583-6_5

[11] Bruno Cartaxo, Gustavo Pinto, and Sergio Soares. 2018. The role of rapid reviews in supporting decision-making in software engineering practice. In *International Conference on Evaluation and Assessment in Software Engineering*. Association for Computing Machinery, New York, NY, USA, 24–34.

[12] Bruno Cartaxo, Gustavo Pinto, and Sergio Soares. 2018. Towards a model to transfer knowledge from software engineering research to practice. *Information and Software Technology* 97, January (2018), 80–82. https://doi.org/10.1016/j.infsof.2018.01.001

[13] Bruno Cartaxo, Gustavo Pinto, and Sergio Soares. 2020. *Rapid Reviews in Software Engineering*. Springer International Publishing, Cham, 357–384. https://doi.org/10.1007/978-3-030-32489-6_13

[14] Bruno Cartaxo, Gustavo Pinto, Elton Vieira, and Sergio Soares. 2016. Evidence briefings: Towards a medium to transfer knowledge from systematic reviews to practitioners. In *International Symposium on Empirical Software Engineering and Measurement*. Association for Computing Machinery, New York, NY, USA, 57:1–57:10.

[15] Jeffrey Carver, Edgar Hassler, Elis Hernandes, and Nicholas Kraft. 2013. Identifying barriers to the systematic literature review process. In *International Symposium on Empirical Software Engineering and Measurement*. 203–212.

[16] Daniela Castelluccia and Giuseppe Visaggio. 2013. Teaching evidence-based software engineering: Learning by a collaborative mapping study of open source software. *ACM SIGSOFT Software Engineering Notes* 38, 6 (2013), 1–4.

[17] Cagatay Catal. 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 (2013), 1–2.

[18] Fabio Da Silva, André Santos, Sérgio Soares, A. Frana, Cleviton Monteiro, and Felipe Maciel. 2011. Six years of systematic literature reviews in software engineering: An updated tertiary study. *Information and Software Technology* 53, 9 (2011), 899–913.

[19] Prem Devanbu, Thomas Zimmermann, and Christian Bird. 2016. Belief & evidence in empirical software engineering. In *International Conference on Software Engineering*, Vol. 14-22-May-. 108–119. https://doi.org/10.1145/2884781.2884812

[20] Garrett M Fitzmaurice and Caitlin Ravichandran. 2008. A primer in longitudinal data analysis. *Circulation* 118, 19 (2008), 2005–2010. https://doi.org/10.1161/CIRCULATIONAHA.107.714618

[21] Davide Fucci, Simone Romano, Maria Teresa Baldassarre, Danilo Caivano, Giuseppe Scanniello, Burak Turhan, and Natalia Juristo. 2018. A longitudinal cohort study on the retainment of test-driven development. In *International Symposium on Empirical Software Engineering and Measurement*. https://doi.org/10.1145/3239235.3240502

[22] Edgar Hassler, Jeffrey Carver, Nicholas Kraft, and David Hale. 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*. Association for Computing Machinery, New York, NY, USA, 31:1–31:10.

[23] May Elin T. Horntvedt, Anita Nordsteien, Torbjørg Fermann, and Elisabeth Severinsson. 2018. Strategies for teaching evidence-based practice in nursing education: A thematic literature review. *BMC Medical Education* 18, 1 (2018), 1–11. https://doi.org/10.1186/s12909-018-1278-z

[24] Dragan Ilic and Stephen Maloney. 2014. Methods of teaching medical trainees evidence-based medicine: A systematic review. *Medical Education* 48, 2 (2014), 124–135. https://doi.org/10.1111/medu.12288

[25] David Janzen and Jungwoo Ryoo. 2009. Engaging the net generation with evidence-based software engineering through a community-driven web database. *Journal of Systems and Software* 82, 4 (2009), 563–570.

[26] David S. Janzen and Jungwoo Ryoo. 2008. Seeds of Evidence: Integrating Evidence-Based Software Engineering. In *Conference on Software Engineering Education and Training*. 223–230.

[27] 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.

[28] 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.

[29] Magne Jørgensen, Tore Dybå, and Barbara Kitchenham. 2005. Teaching evidence-based software engineering to university students. In *International Software Metrics Symposium*. 24:1–24:8.

[30] Abhinaya Kasoju, Kai Petersen, and Mika Mäntylä. 2013. Analyzing an automotive testing process with evidence-based software engineering. *Information and Software Technology* 55, 7 (2013), 1237–1259.

[31] Mark Kasunic. 2005. *Designing an effective survey*. Technical Report CMU/SEI-2005-HB-004. Carnegie Mellon University Software Engineering Institute, Pittsburgh, PA 15213-3890.

[32] Declan Kennedy, Áine Hyl, Norma Ryan, Áine Hyland, and Norma Ryan. 2007. *Writing and using learning outcomes: A practical guide*. European University Association, Chapter C 3.4-1, 1–30.

[33] Barbara Kitchenham, Pearl Brereton, and David Budgen. 2010. The educational value of mapping studies of software engineering literature. In *International Conference on Software Engineering*. 589–598.

[34] Barbara Kitchenham, David Budgen, and Pearl Brereton. 2015. *Evidence-based software engineering and systematic reviews*. CRC Press, London.

[35] Barbara Kitchenham, Tore Dybå, and Magne Jørgensen. 2004. Evidence-based software engineering. In *International Conference on Software Engineering*. IEEE Computer Society, USA, 273–281.

[36] Konstantinos Kyriakoulis, Athina Patelarou, Aggelos Laliotis, Andrew C Wan, Michail Matalliotakis, Chrysoula Tsiou, and Evridiki Patelarou. 2016. Educational Evaluation for Health Professions Educational strategies for teaching evidence-based practice to undergraduate health students: systematic review. *J Educ Eval Health Prof* 13 (2016). https://doi.org/10.3352/jeehp.2016.13.34

[37] Camilla Marie Larsen, Anne Seneca Terkelsen, Anne Marie Fiala Carlsen, and Hanne Kaae Kristensen. 2019. Methods for teaching evidence-based practice: A scoping review. *BMC Medical Education* 19, 1 (2019). https://doi.org/10.1186/s12909-019-1681-0

[38] Mathieu Lavallée, Pierre Robillard, and Reza Mirsalari. 2014. Performing systematic literature reviews with novices: An iterative approach. *IEEE Transactions on Education* 57, 3 (2014), 175–181.

[39] Claire Le Goues, Ciera Jaspan, Ipek Ozkaya, Mary Shaw, and Kathryn T. Stolee. 2018. Bridging the Gap: From Research to Practical Advice. *IEEE Software* 35, 5 (2018), 50–57. https://doi.org/10.1109/MS.2018.3571235

[40] Elaine Lehane, Patricia Leahy-Warren, Cliona O'Riordan, Eileen Savage, Jonathan Drennan, Colm O'Tuathaigh, Michael O'Connor, Mark Corrigan, Francis Burke, Martina Hayes, Helen Lynch, Laura Sahm, Elizabeth Heffernan, Elizabeth O'Keeffe, Catherine Blake, Frances Horgan, and Josephine Hegarty. 2019. Evidence-based practice education for healthcare professions: An expert view. *BMJ Evidence-Based Medicine* 24, 3 (2019), 103–108. https://doi.org/10.1136/bmjebm-2018-111019

[41] Timothy C. Lethbridge, Susan Elliott Sim, and Janice Singer. 2005. Studying software engineers: Data collection techniques for software field studies. *Empirical Software Engineering* 10, 3 (2005), 311–341. https://doi.org/10.1007/s10664-005-1290-x

[42] Tomasz Lewowski and Lech Madeyski. 2022. *Code Smells Detection Using Artificial Intelligence Techniques: A Business-Driven Systematic Review*. Springer International Publishing, 285–319. https://doi.org/10.1007/978-3-030-77916-0_12

[43] Patricia Manns and Johanna Darrah. 2012. A structured process to develop scenarios for use in evaluation of an evidence-based approach in clinical decision making. *Advances in Medical Education and Practice* 3 (2012), 113–119. https://doi.org/10.2147/AMEP.S37510

[44] Laurie McLeod, Stephen MacDonell, and Bill Doolin. 2011. Qualitative research on software development: A longitudinal case study methodology. *Empirical Software Engineering* 16, 4 (2011), 430–459. https://doi.org/10.1007/s10664-010-

9153-5

[45] Briony Oates and Graham Capper. 2009. Using systematic reviews and evidence-based software engineering with masters students. In *International Conference on Evaluation and Assessment in Software Engineering*. 79–87.

[46] Bridget C O'Brien, Ilene B Harris, Thomas J Beckman, Darcy A Reed, and David A Cook. 2014. Standards for reporting qualitative research: a synthesis of recommendations. *Academic Medicine* 89, 9 (2014), 1245–1251.

[47] Athina E. Patelarou, Enkeleint A. Mechili, María Ruzafa-Martinez, Jakub Dolezel, Joanna Gotlib, Brigita Skela-Savič, Antonio Jesús Ramos-Morcillo, Stefano Finotto, Darja Jarosova, Marta Smodiš, Daniela Mecugni, Mariusz Panczyk, and Evridiki Patelarou. 2020. Educational interventions for teaching evidence-based practice to undergraduate nursing students: A scoping review. *International Journal of Environmental Research and Public Health* 17, 17 (2020), 1–26. https://doi.org/10.3390/ijerph17176351

[48] Sebastián Pizard, Fernando Acerenza, Ximena Otegui, Silvana Moreno, Diego Vallespir, and Barbara Kitchenham. 2021. Training students in evidence-based software engineering and systematic reviews: a systematic review and empirical study. *Empirical Software Engineering* 26, 3 (2021), 50. https://doi.org/10.1007/s10664-021-09953-9

[49] Austen Rainer and Sarah Beecham. 2008. A follow-up empirical evaluation of evidence based software engineering by undergraduate students. In *International Conference on Evaluation and Assessment in Software Engineering*. 78–87.

[50] Austen Rainer, Tracy Hall, and Nathan Baddoo. 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*. 91–100.

[51] Kathleen D Ramos, Sean Schafer, and Susan M Tracz. 2003. Validation of the Fresno test of competence in evidence based medicine. *BMJ* 326, 7384 (2003), 319–321. https://doi.org/10.1136/bmj.326.7384.319

[52] Talita Ribeiro, Jobson Massollar, and Guilherme Travassos. 2018. Challenges and pitfalls on surveying evidence in the software engineering technical literature: An exploratory study with novices. *Empirical Software Engineering* 23, 3 (2018), 1594–1663.

[53] Per Runeson, Martin Host, Austen Rainer, and Bjorn Regnell. 2012. *Case Study Research in Software Engineering: Guidelines and Examples* (1st ed.). Wiley Publishing.

[54] Jo Rycroft-Malone, Kate Seers, Angie Titchen, Gill Harvey, Alison Kitson, and Brendan McCormack. 2004. What counts as evidence in evidence-based practice? *Journal of advanced nursing* 47, 1 (2004), 81–90.

[55] Homayoun Sadeghi-Bazargani, Jafar Sadegh Tabrizi, and Saber Azami-Aghdash. 2014. Barriers to evidence-based medicine: A systematic review. *Journal of Evaluation in Clinical Practice* 20, 6 (2014), 793–802.

[56] Laura Scurlock-Evans and Dominic Upton. 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 (2015), 369–399.

[57] Mary Shaw. 1987. Education for the Future of Software Engineering. May (1987), 344–357. https://doi.org/10.1007/978-1-4612-4720-3_25

[58] Tim Stroobants, Johan Vanderfaeillie, Caroline Andries, and Frank Van Holen. 2016. Youth care workers' perspectives on and adoption of evidence-based practice. *Children and Youth Services Review* 71 (2016), 299–307. https://doi.org/10.1016/j.childyouth.2016.11.017

[59] Marco Torchiano, Daniel Fernández, Guilherme Travassos, and Rafael De Mello. 2017. Lessons learnt in conducting survey research. In *International Workshop on Conducting Empirical Studies in Industry*. IEEE Press, USA, 33–39.

[60] Mark Turner, Rumjit Kaur, and Pearl Brereton. 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*. 21:1–21:15.

[61] Dominic Upton, Danielle Stephens, Briony Williams, and Laura Scurlock-Evans. 2014. Occupational therapists' attitudes, knowledge, and implementation of evidence-based practice: A systematic review of published research. *British Journal of Occupational Therapy* 77, 1 (2014), 24–38.

[62] Traci L. Wike, Sarah E. Bledsoe, Jennifer I. Manuel, Mathieu Despard, Lydia V. Johnson, Jennifer L. Bellamy, and Candace Killian-Farrell. 2014. Evidence-Based Practice in Social Work: Challenges and Opportunities for Clinicians and Organizations. *Clinical Social Work Journal* 42, 2 (2014), 161–170. https://doi.org/10.1007/s10615-014-0492-3

[63] Robert Yin. 2017. *Case Study Research and Applications: Design and Methods 6ed* (6ed ed.). SAGE Publications.

[64] Sandra Zwolsman, Ellen Te Pas, Lotty Hooft, Margreet Wieringa-De Waard, and Nynke Van Dijk. 2012. Barriers to GPs' use of evidence-based medicine: A systematic review. *British Journal of General Practice* 62, 600 (2012), 511–521.

Appendix B

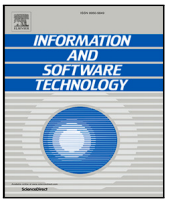
Evidence-based Software Engineering Use by a Government Agency

Information and Software Technology 154 (2023) 107101

Contents lists available at ScienceDirect

Information and Software Technology

journal homepage: www.elsevier.com/locate/infsof



Assessing attitudes towards evidence-based software engineering in a government agency

Sebastián Pizard [a,*], Fernando Acerenza [a], Diego Vallespir [a], Barbara Kitchenham [b]

[a] School of Engineering, Universidad de la República, Montevideo, Uruguay
[b] School of Computing and Mathematics, Keele University, Keele, UK

ARTICLE INFO

Keywords:
Evidence-based software engineering
Evidence-based practice
Field investigation
Government agency
Focus group
Follow-up study

ABSTRACT

Context: Evidence-based practice (EBP) has allowed several disciplines to become more mature by emphasizing the use of evidence from well-designed and well-conducted research in decision-making. Its application in SE, Evidence-based software engineering (EBSE) can help to bridge the gap between academia and industry by bringing together academic rigor and research of practical relevance. To achieve this, it seems necessary to improve its adoption.

Objective: We sought both to study the attitudes towards EBSE of stakeholders working in a government agency (GA) and to assess whether knowledge of EBSE would impact their working practices.

Method: We conducted a multi-stage field investigation in an Uruguayan national GA that is responsible for digital policies. First, we organized an EBSE awareness lecture and we collected and analyzed participants' perceptions of the value and limitations of EBSE. Sixteen months later, in a second stage, we contacted the agency and asked participants whether they had made use of the information about EBSE we presented to them.

Results: Initially, participants reported that EBSE seemed useful for tackling challenging problems and, in particular, considered its use appropriate given the agency's responsibilities. Perceived barriers to EBSE adoption were the need for institutional support, the lack of government practice reports, inadequate skills or motivation, the cost of conducting systematic reviews, and the lack of evidence about emerging issues. In the follow-up survey, although the participants were not undertaking systematic reviews themselves, many reported improvements in how they searched for and evaluated information to support their work.

Conclusion: Our study presents some insights to better understand EBSE adoption. With the exception of GA-specific issues, perceived value and barriers to adoption were consistent with those reported in software engineering and other disciplines. Our follow-up study confirms the potential value of evidence in the context of IT regulatory and government bodies.

1. Introduction

In medicine, the use of evidence-based practice (EBP[1]) has made it possible to address the information needs of practitioners without detracting from the rigor levels of the research community [1]. Kitchenham et al. introduced EBP into software engineering (SE), calling it evidence-based software engineering (EBSE) [2]. EBSE aims to improve decision-making related to software development and maintenance by integrating the best current research evidence with practical experience and human values.

In practice, aggregating scientific results, interpreting them, and using them in professional practice occurs from the interaction of various stakeholders, in different social processes (e.g. including the interpretation and discussion of evidence by professional networks and communities of practice) [3]. The complexity of these processes and the importance of considering different stakeholders were pointed out in the first EBSE paper [2]. In particular, the authors reflected that the adoption of EBSE would require *'extensive collaboration and long-term commitment among individual research groups worldwide, and active support from other stakeholders such as practitioners in industry, certification bodies, etc'*.

In healthcare, Government agencies (GAs) are considered key stakeholders in EBP adoption as they support the conduct of SRs, promote

* Corresponding author.
E-mail address: spizard@fing.edu.uy (S. Pizard).

[1] In this paper, we use EBP as a generic term when our comments relate the use of evidence in any applied discipline. We use EBSE when our comments concern the use of evidence in the context of SE, CS and IT only.

https://doi.org/10.1016/j.infsof.2022.107101
Received 29 March 2022; Received in revised form 26 October 2022; Accepted 28 October 2022
Available online 3 November 2022

good practices and define acceptable levels of quality. For example, the Agency for Healthcare Research and Quality [4] within the United States review relevant literature and prepare evidence reports, and help to translate them into practice and teaching programs. Another example is the UK National Institute for Health Research whose Systematic Review Program supported the conduct or update of more than 1500 SRs during the years 2007–2011 which impacted more than 240 sets of practice guidelines [5].

In our field, Unterkalmsteiner and Gorschek's work confirms that GAs can benefit from SE technologies for requirements specification, requirements management, and project management, among activities, but did not consider how such technologies should be selected [6], which is an issue that EBSE might be expected to address. But, GAs could be not only potential users of EBSE in their own work or when providing recommendations to industry. They could be, also, potential advocates for EBSE if they adopt a policy of supporting recommendations with available empirical evidence. Therefore, it is important to investigate EBSE in the context of GAs.

Investigating attitudes and behavior is a widely used approach to understanding the needs of EBP stakeholders (see, for example, [7–9]). However, the study of non-academic stakeholders has not been extensively explored in EBSE research. We are only aware of three reports of non-academic use of EBSE [10–12], all three with industry companies and only one includes a report of the attitudes of the industry stakeholders [11].

Thus, when the School of Engineering at the Universidad de la República was invited to give a presentation on its research to the Uruguayan E-Government agency (AGESIC), we took the opportunity to introduce EBSE to AGESIC staff and assess their attitudes to the concept.

In the first stage, we conducted a field study in the form of an EBSE awareness lecture to AGESIC members, followed by a focus group session to collect and analyze their perceptions of the value and limitations of EBSE. In the second stage, we followed up our investigation with a survey that sought to determine whether the ideas we presented had any impact on the agency's working practices. The second stage was not initially planned but, when we prepared a report on the focus group for AGESIC, we realized that feedback on staff adoption of EBSE was an important part of assessing the potential value of EBSE.

Our work is complementary to current research on EBSE, and is original in that it (1) is an attempt to assess the potential value of EBSE, and not the value or utility of SLRs, (2) involves participants from an important group of stakeholders so far not investigated in relation to EBSE, (3) includes the comparison of our results with five systematic reviews on EBP adoption in other disciplines, and (4) reports a follow-up study, that confirmed that more than a third of the EBSE lecture participants had begun to seek and to use evidence as part of their working practices.

The remainder of the paper is organized as follows. In Section 2, we review previous research. In Section 3, we present our study context, i.e., AGESIC aims and characteristics. Section 4 describes our research questions and strategy. We present ethical issues, characteristics of study conduct, and data analysis in Section 5. Section 6 reports the results of our study in the context of our research questions. In Section 7, we include a discussion of our results and their significance. We present the main limitations of our study in Section 8 and our recommendations for researchers in Section 9. Section 10 includes concluding remarks and directions for further research.

2. Related work

This section includes related work on EBSE research with non-academic stakeholders which we have supplemented with a brief preliminary review of the evidence-based practice (EBP) adoption research in other disciplines.

2.1. EBSE non-academic stakeholders

Budgen et al. [13] considered EBSE as an innovation and used part of Rogers' diffusion of innovations theory [14] to understand the progress of its adoption. They suggest that in SE there are two separate EBSE adoption cycles. The first one is found in the research community, in which, according to the authors, there is widespread adoption of the systematic review as an important research method and a major support tool for EBSE. The second cycle is focused on the users of the evidence (i.e., industry, government, etc.) and is probably in the 'innovator' stage (i.e., the very initial stage). Budgen's argument suggests that it is important to investigate EBSE adoption by stakeholders other than academics.

To our knowledge, there are only three reports of non-academic use of EBSE, and all of them involve conducting a secondary study for a company. Kasoju et al. [10], used the EBSE process (which included conducting a systematic review), along with a case study and value stream mapping technique, to analyze an automotive test process and propose improvements. Although the proposals were not carried out, they were presented to the practitioners who provided feedback. Unfortunately, the authors do not report details of that last activity.

In 2018, Cartaxo et al. reported another EBSE application, in this case, conducting a rapid review (a limited systematic review) to collect practical recommendations in order to help an agile software development company with customer collaboration issues [11]. Although their study was focused on proposing and evaluating the use of rapid reviews, it is also an application of EBSE. The practitioners found that the results of the rapid review seemed more reliable than the information they usually used and that this increased the team's confidence in the decisions they made. They also believed that the review process allowed them to better understand the problem and to learn new concepts.

Finally, Lewowski & Madeyski reported the conduct of an SR commissioned by a software development company [12]. The company, responsible for a platform for automated code review, sought to learn about research related to the prediction of code smells using artificial intelligence. The SR was a preliminary step in a joint research & development project. Although the authors reported decisions that took the company into account (e.g., not requiring completeness of the search for primary studies), the use of the results or company satisfaction is not discussed.

These initiatives contribute to a better understanding of the use of EBSE. Even so, we believe that it is useful to adopt a broader perspective (i.e., including other potential key stakeholders such as government agencies or professional communities) to study the use of EBSE in depth (e.g., investigating whether they find the approach useful, or analyzing what challenges arise when trying to use evidence).

2.2. Evidence-based practice adoption in other disciplines

Since evidence-based practice has a longer history in other disciplines, it seemed useful to gain some insight into the challenges of its adoption with which to compare our results. For this purpose, we conducted a rapid review searching for *secondary studies* on the adoption or impact of EBP or systematic reviews (details are presented in [15]). We found five studies of interest from several fields (i.e., social work [16], occupational therapy [9], physiotherapy [17], medicine [18], and healthcare policy [19]). We analyzed three aspects of the studies:

1. *Perceived Value*: In general, practitioners and stakeholders hold positive views towards EBP [9,16,17], and many expressed the view that it is important and useful for their work [16] or necessary for the role they play [17].

Table 1
Barriers to EBP adoption.

	Field & Secondary study			
Key barriers to EBP	Social work [16]	Occupational therapy [9]	Physiotherapy [17]	Medicine [18]
Relevance, applicability, availability, quality of research evidence	×	×	×	×
Lack of time/workload pressure	×	×	×	×
Lack of knowledge, training or inadequate skills	×	×	×	×
Lack of resources (e.g., funding, staff, guidelines)	×		×	×
Lack of organizational support	×	×		×
Lack of teamwork or communications problems		×		×

2. *Barriers to Adoption*: Table 1 presents the limitations or barriers to EBP adoption identified in at least two SLRs. The most common barriers were associated with: characteristics of research evidence, lack of time or workload pressures, lack of training, or inadequate skills.
3. *Mechanisms to Improve Adoption*: Possible strategies or facilitators to improve the adoption of EBP reported by at least two SLRs included: improving access to resources (e.g., scientific literature or staff) [9,16,17], improving the dissemination of evidence [9,17,19], promoting collaboration between researchers and practitioners [16,19], and improving the applicability of evidence (i.e., by analyzing its context and applicability, or by increasing the practical/applied value of research) [17,19]. Two studies [17,19] suggested that the preference for acquiring evidence from colleagues implies that an efficient disseminating strategy could be the use of knowledge brokers. Knowledge brokers are individuals with good communication, negotiation, and problem-solving skills, who are familiar with the culture and environment of researchers and practitioners [20].

3. Study context: AGESIC

The Agency for the Development of Electronic Government and Information Society and Knowledge (AGESIC) is an executive unit under the President of Uruguay. It was created in 2006 with the aims of *'(1) seeking to improve citizen services, using the possibilities offered by information and communication technologies, and (2) promoting the development of the information society in Uruguay with emphasis on the inclusion of the digital practice of its inhabitants and the strengthening of the skills of the society in the use of technologies'* [21]. Its growth has been sustained, from 30 employees in 2007 to more than 180 in 2015 [22]. In 2019, the year in which the first part of this study was carried out, AGESIC had 0.2% of the national public budget.[2]

The responsibilities of AGESIC arise directly from the Uruguayan Digital Government Strategy [25], of which the agency is a promoter and incubator of initiatives within the entire government. The strategy is a roadmap for transformation and innovation in government management and is based on the following areas/objectives:

- *Proximity Government.* Facilitating citizens' access to all government information and services (ensuring that 100% of the Central Administration services can be done online from start to end).
- *Open Government.* Improving relations with citizens through transparency, open data, and promotion of citizen participation and innovation.
- *Smart Government.* Exploit existing information and use it to move to a more proactive attitude that anticipates the needs of citizens and prevents problems. This includes, among other issues, the sub-objective of *'Strengthening evidence-based decision-making processes'*, which is aimed at creating and improving data analysis capabilities internal to the government (internal evidence). Thus, if the use of scientific evidence and EBSE were of interest or utility, it would be included in this area.
- *Efficient Government.* Improving cross-organizational management in the government and promoting the adoption of common solutions.
- *Whole-of-Government.* Encouraging technological integration and data interoperability as a basis for the management of systems development and evolution.
- *Reliable Digital Government.* Ensuring the response to risks, threats, and challenges that arises using digital technologies in the government.

In particular, AGESIC carries out these activities (among others):

- Conducts national and comprehensive programs in Uruguay in relation to IT. For example, promoting the intensive use of information and communications technologies in the health sector to improve the quality and continuity of care.
- Elaborates policies, plans, and national strategies in matters of governance, integration, interoperability, government business architecture, human capital, and acquisitions related to information and communication technologies in public agencies.
- Creates technical standards relating to products and services related to information and communication technologies in public agencies.

As a concrete example, AGESIC actively participated in, and made important demands on privacy and information security, in the national software development project to incorporate contact tracing for COVID-19 (Uruguay was one of the first three countries to use it) [26]. AGESIC can also recommend, execute and fund research and development initiatives, but they are not a research funding agency. They also hire suppliers (including the Universidad de la República) when they need consulting services. However, none of the members we spoke with knew whether the agency had commissioned (or used) secondary studies.

An issue of major importance from the viewpoint of our research is that AGESIC carries out policy-making and legislation in SE. It supports government initiatives in the efficient use of information technologies (e.g., helping to define software requirements and purchase conditions—including software quality requirements) and provides recommendations to other organizations and to society on the use of software and technologies (e.g., informing on the risks of new technologies). By helping to establish conditions for state purchases, the agency also defines requirements that must be met by national and foreign state software providers. Therefore, the agency could significantly influence the adoption and dissemination of EBSE in our country. However, the agency would first need to learn about EBSE and evaluate whether it is a practice that they would find useful and appropriate to support their goals. Our study aims to contribute to this purpose.

AGESIC has eight main areas and some decentralized units. Some areas are, for example, Organizational Transformation, Citizen Services, Information Security, and Information Technology. In our study, we

[2] Calculated based on information from [23,24].

Table 2
Characteristics of the AGESIC IT Area and its divisions.

Information technology area	Aims to contribute to the achievement of an optimized management of IT in the public administration, through the investigation of trends and innovative technology products that contribute to the digital transformation of the country.	
Division	Division objective	Division staff
Government architecture	Promote a common vision of IT through a government architecture that facilitates the integration, interoperability, and optimization of technologies and solutions for the implementation of better services for citizens.	8
IT optimization	Research, development, and promotion of methodologies, good practices, and tools that contribute to the creation of value and facilitate the effective management of information and technological assets in government organizations.	5
Emerging technologies	Research, develop and promote the use of innovative and emerging technologies that contribute to the creation of value and facilitate effective management in government organizations.	7
Data management	Develop and strengthen systems, platforms, frameworks, and good practices for the management of information data, in order to favor a public administration based on data and the optimization of public services.	2

only worked with the Information Technology area. Table 2 presents the main characteristics of this area, which has four divisions and 24 staff members (i.e., director, deputy director, managers, and consultants).

4. Research questions and method

The purpose of our research is to study the attitudes of GA stakeholders towards EBSE and assess their motivation to adopt it. In particular, we try to answer the following research questions:

RQ1: Is evidence-based software engineering a method that should be considered for adoption by government agency stakeholders?
RQ2: Does EBSE awareness have any impact on the working practices of government agency stakeholders?

We investigated the research questions in the specific context of AGESIC using a multi-stage research design, with an initial focus group study, followed by a survey sixteen months later. This design allowed us to study in-depth participants' first impressions of EBSE by allowing them to discuss its value and limitations. Later, the follow-up survey allowed us, in a non-intrusive and cost-effective way, to evaluate possible impacts on the work practices of the participants.

As evidence-based practice and its application were unknown to all but one of the agency staff, we preceded our focus group with a brief presentation of EBSE and the use of evidence. Following this, we undertook a focus group study where the participants discussed the value and limitations of EBSE.

The focus group method seemed to be appropriate to achieve our purpose as it can be used to obtain initial feedback on new concepts [27], in this case, EBSE was unknown to almost all the participants. In addition, it is a method that has already been used to study attitudes towards EBP in other fields (e.g., medicine [28], dentistry [29], psychiatry [30], nursing [8]). However, our study is inconsistent with definitions of focus group characteristics used in marketing research [31], in particular, the participants were known to one another, and selection was based on convenience and not to cover a specific demographic. In essence, we conducted a field study utilizing, where appropriate, focus group methods.

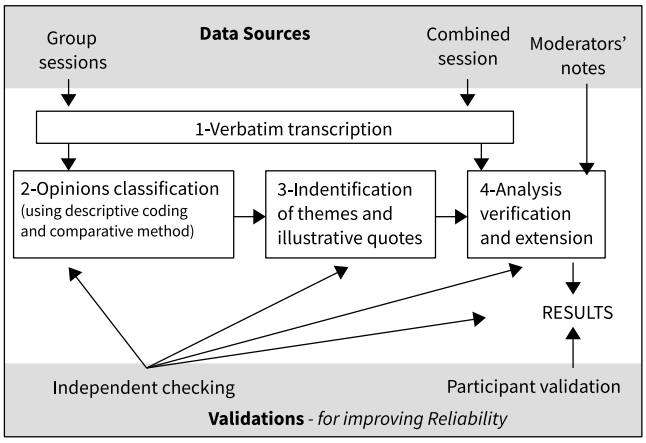


Fig. 1. First stage data analysis process.

The need to validate our report with the participants prompted us to investigate whether knowledge of EBSE had caused them to make any changes to their working practices. So, sixteen months after we conducted the EBSE awareness lecture, we circulated a self-administered questionnaire to participants to ask them whether they had subsequently made any changes to their working practices.

We believe that our initiative of giving an EBSE awareness intervention and using non-intrusive methods to evaluate its impact is the most appropriate (and most ethical) approach to begin investigating whether its adoption is possible and beneficial for non-academic stakeholders. Using other more intrusive methods (e.g., action research) aimed at changing working practices seems to be very premature since the benefits of adopting EBSE in non-academic settings are unknown. Furthermore, it would be fair to say that although the agency's IT division director was keen for his staff to participate in an EBSE talk, a focus group and a survey, any other more invasive research method would have been hard to justify.

We used the Tong et al. checklist for reporting interviews and focus group research [32] and the O'Brien et al. checklist for reporting qualitative research [33]. In particular, the fourth author evaluated the initial report against the checklists and made suggestions that were later addressed by the first author.

In this section, we have reported only the most important study methods. However, full details of the design and conduct of our study are available in our supplementary material [15].

5. Ethical issues, data collection and analysis

In the following subsections, we present ethical considerations and the main characteristics of the data collection and data analysis processes from both the focus group and the follow-up survey.

5.1. Ethical considerations

In consultation with the director of the IT area, we agreed to the following conditions: to inform the participants about the study and request that they sign an informed consent form if they wanted to participate, to ensure the confidentiality of their opinions in any published report, to validate the results with the participants, and to obtain the agency's approval of the report prior to its publication.

Before starting the focus group, the first author explained to the participants that their discussion was going to be used to study EBSE adoption. He described the data collection procedures that would be used, i.e. notes from the moderators and audio recordings. Participants were also told that the information collected would be treated confidentially (i.e., none of their opinions will be linked to their identities in public reports or be made available to anyone other than the researchers and study participants) and that they could withdraw from the study at any time. Individuals who agreed to participate voluntarily signed an informed consent form before starting.

In the follow-up survey, participation was optional but not anonymous. Before answering the questionnaire, participants were reminded that we would preserve their anonymity in the dissemination of the research results.

5.2. Focus group

The focus group was conducted on August 26, 2019 at the agency's facilities. We used the guidelines developed by Kontio et al. [27,34] and recommendations proposed by França et al. [35] to plan the focus group. The most important characteristics of the research design, conduct, and validation are discussed below.

A call for participation via email was made within the Information Technology area of AGESIC. In the preceding week, we sent a short questionnaire to assess the demographics of those who had confirmed their attendance. Full details are in the supplementary material [15].

Before the focus group discussions, the first author of this paper gave a presentation on EBSE. This initial briefing lasted an hour and a half and had the following three main parts (1) a brief introduction to SE research, (2) a brief introduction to EBSE, and (3) a presentation of an SR, published in Spanish (the participants' mother tongue), on a topic of potential interest to the participants. More precisely, the example SR dealt with metrics to evaluate mobile applications for people with visual impairments [36], a topic of interest to the agency since are responsible for developing web accessibility guidelines for government applications [37].

Thirteen agency staff attended the lecture but only eleven attendees agreed to take part in the discussion, but one of them participated only in the combined session.[3] The focus group had two different discussion sessions. First, the participants worked in **group sessions**. For this, two groups were formed (without any segmentation strategy), each one with five participants and a moderator. They were invited to discuss the usefulness of EBSE. The following questions were used to prompt discussions on its benefits and limitations: *'What is EBSE useful for?'* and *'What is EBSE not useful for?'*. They were encouraged not only to say whether EBSE would be useful or not, but also to raise arguments or situations that would support their position. They were given 20 min for the discussion. Second, a **combined session** was carried out, also lasting 20 min, in which each group shared and discussed their reflections with the other group.

We used qualitative thematic analysis to analyze the data (which is an approach recommended for analyzing focus group data [38]) with a realist approach as our theoretical framework. Our analysis process is shown in Fig. 1. It is based on the work of Braun and Clarke [39] and its main features are familiarization with the data, initial coding, and identification of themes together with quotes that illustrate them. Some relevant adaptations that we made are:

- We conducted an in-depth analysis of the group session transcripts. Subsequently, we validated the results with the information from the combined session and moderator notes. This seemed the most appropriate approach since during the combined session the participants only reviewed the opinions they had expressed in the group sessions and the moderators' notes also summarized the group sessions.
- For the classification of opinions we used the techniques of descriptive coding and the constant comparative method [40] and the freeware tool Saturate.[4]
- When reporting the results, as the language used by the participants was Spanish, we worked with an experienced translator to create accurate English versions of the selected quotes.
- The first author carried out the analysis, and, subsequently, the second author did an independent checking of the process and its results. Both authors discussed the comments until they reached an agreement on all issues.

In December 2020, we contacted the agency and sent participants a summary of the study results (with content similar to Section 6.1).[5] Seven participants responded. They agreed with the results summary and did not suggest any changes.

5.3. Follow-up survey

The follow-up survey sought the participants' opinions about any impact that knowing EBSE had on their work practices. Furthermore, to better understand the context in which the participants could make use of the evidence generated by secondary studies, we tried to investigate what type of literature the participants usually utilize. This was an issue that we had not investigated before and that, after analyzing the results of the first stage, we thought might help to interpret our findings.

The survey, designed according to recommendations proposed in the literature [42,43], was conducted simultaneously with the validation of the first stage results.

The survey instrument was a self-administered questionnaire with eleven questions (some of which were open questions, to allow the participants to explain their answers more completely). The question-

[3] Although it was not recorded in audio, the researchers registered that one participant abandoned the meeting before the end of the group session and another entered after the combined session began.

[4] http://www.saturateapp.com/

[5] This is a method of validating results sometimes referred to as member checking [41].

naire was designed in Google Forms, and was available for one month (from December 23, 2020 to January 26, 2021). The translation of the questions (originally in Spanish) and the flow of the survey are shown below.

S1 Have you used any ideas seen in the EBSE lecture in your working practices?*

- Yes
- No *[Skip to question 3 and after to question 6]*

S2 Indicate what you have done

- Conducting a secondary study (systematic review, mapping study, or rapid review).
- Using an existing secondary study (systematic review, mapping study, or rapid review).
- Using search engines for scientific articles.
- Undertaking a critical appraisal of reports that compare different technologies.
- Other activities, indicate which.

S3 Indicate your first and last name. This information will be used to relate your answers to previous results.*

S4 Of the ideas you used, did you find anything particularly difficult?*

S5 Of the ideas you used or their results, did you find anything particularly useful?* *[Skip to question 8]*

S6 What is the main reason why you did not use anything seen in the EBSE lecture?*

- The ideas do not fit the way of working.
- I have had no chance or need.
- I tried to use the ideas but was unsuccessful.
- The lecture was not enough for me to understand how to apply the ideas of evidence-based practice.
- Other reasons, indicate which.

S7 Do you think that your knowledge about EBSE could be used in the future? For what?*

S8 Did knowing about EBSE MOTIVATE you to improve or change any activity in your working practices?*

- Yes *[Skip to question 9]*
- No *[Skip to question 10]*

S9 In which activities, tasks, problems or situations do you perceive that you have changed after learning about EBSE?*

S10 What kind of literature do you use in your practice at the agency? *

S11 Could you indicate titles or links (this is preferable) to two typical examples of the literature you use?*

For quantitative questions, we counted the number of responses in each response category. For qualitative questions, the textual responses were collated and we used descriptive coding and constant comparative method to classify the responses. The responses from S11 (i.e., examples of publications used by participants) were listed and mapped with the results from S10 (i.e., types of publications used). The result of this mapping was also classified into the grey literature layers, as explained later. Similarly to what was done before in the first stage, the first author performed the initial data analysis that was later reviewed and discussed with the second author.

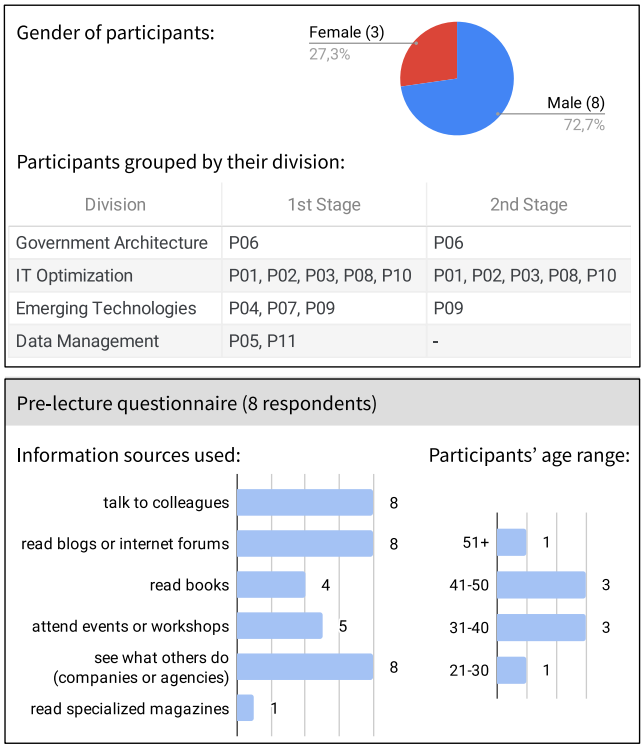


Fig. 2. Participants demographic information.

6. Results

Fig. 2 shows demographic information describing the participants. In particular, Fig. 2 reports the percentage of male and female participants, the agency division they work for, the information sources they report using, and their age range.

None of the agency's participants, except the participant identified as P01, had any ties to our university or potential conflict of interest with our research. P01 worked at the agency and was also an assistant teacher and junior researcher at our university. Despite this, P01 reported no conflict of interest with our research. P01 was the only participant who knew about EBSE beforehand.

Below we answer the research questions using the results of the two stages of our study.

6.1. RQ1: Should EBSE be considered for adoption by GAs

We analyzed the focus group discussions and grouped them into three categories: positive attitudes towards EBSE (i.e., perceptions about the value and benefits of using it) indicated with [+], negative attitudes (i.e., limitations, risks, and challenges about the use of EBSE) indicated with [−] and other non-theme related issues arising from the data analysis indicated with [0]. The results are listed below.

[+] **Innovation, uncertainty or risk.** Participants identified that EBSE could be useful in finding innovative solutions and solving problems associated with large risks and uncertainties. This includes, for example, the following scenarios: the need to innovate when creating a new department, trying to solve a problem after several failed attempts (this item was agreed by both groups), avoiding reinventing known solutions, and tackling a little-known problem with a small budget.

For example, in one group, the participants discussed the usefulness of the approach when dealing with an unknown topic, but also without a budget to hire experts:

P04 [EBSE is useful] in the context of a project when you run out of budget money. I don't know. You can't call a consulting firm that does I-don't-know-what or [do] things like that [which require money]. Maybe for things that we know that we are not going to have the budget for or things like that, I am going to apply this kind of thing.
P05 As a first approximation.
P04 Sure. So as not to go in blind. Could be. Or for example, it can be used to discard or accept paths of action. I don't know. I did a review and 80% of the articles tell me to go this way, I'm going there. If they tell me not to, I'm going the other way.
P10 Or make decisions.
P04 As a feasibility analysis, something of that kind, fast, discarding fast.

— Q01 (from discussion group 2)

[+] **Role of the agency.** Both groups discussed the EBSE usefulness in relation to the role of the agency, since they perceive that it is a concept that promotes standard responses to common problems and disseminates good practices. To illustrate this, in the following fragments the participants expressed the view that EBSE would help make the agency's recommendations more aligned with its discourse, and that its use is in accordance with the role model functions of the agency, respectively.

P05 [...] I think [EBSE] is a method that, as we suggest [to others] to work based on data and evidence, it would be good to support it.
P04 Sure, exactly. Take our own medicine.

— Q02 (from discussion group 2)

P01 ... we are role models, we participate worldwide in certain activities and we are role models of that. So it seems to me that when the publications we make point to our role model, it seems to me that it could also apply, use or do [systematic literature] reviews.

— Q03 (from discussion group 1)

Other possible uses of EBSE that were discussed, although to a lesser extent (according to duration and number of participants involved in the discussion), were:

- [+] **It helps to formalize and give more support to internal research**, especially for the very common reports on the state of the art made in the agency.
- [+] **For feedback for continuous improvement**, for example, by doing a periodic SLR on a practice used in the agency and comparing the results with actual performance.
- [+] **The vision of EBSE as one more tool that was available from now on.**

[−] **Institutional support.** Participants of one group identified the importance of institutional support for the adoption of EBSE. As the following quote shows, the institution would need to give time and priority to carry out this type of practice.

P04 If we give people space, there is not one of us who would be incapable of doing such a review. I think we can all do it. So you have to take the time to do it. Because it is not the same as saying AGESIC suggests a certain line of action but based on all this that we saw before.
P10 In line with priorities, it is going to be in the 7800th place
P04 Ahh of course, but I think this is not something you do in two days.
P05 No, I am not saying no, but I think it is a method that also, as we herald [to others] to work based on data and evidence, it would be good to support it.

— Q04 (from discussion group 2)

[−] **Lack of evidence of governments practices.** In one group, the participants discussed for several minutes the lack of reports of governments practices, an issue that was later agreed on by the other group in the combined session. During the discussion, they indicated that they currently find little evidence on issues applied to governments. Then, as shown in the following excerpt, in the combined session, it was also pointed out that although there are government publications, an important concern is their context (e.g., region or form of work) and therefore their applicability. So it is difficult to find useful publications with contexts similar to that of our country.

R1 Are you sure you can't find any government publications?
P04 We can find them, yes, they are not the majority.
P03 And in addition to government, the issue of the regions of the planet has to do ... with the way of working and ...
P04 We find a lot of publications from European governments for example. The UK has a lot of government publishing practice so we found a lot of things [from them]. But when you are going to apply [recommendations from these papers] ... For example, the Estonian government, which is one of the most avant-garde digital governments, has a lot of publications. But the Estonian context is not so similar to ours, due to a regulatory framework and a lot of [other] things. You can find [papers] but later when you are going to use them, it has its ... [complexities]
P03 And many of the things you find at government level sometimes, especially in Latin America and the area around here, are more like expressions of desire than things that actually [are] ... evidence.

— Q05 (from joint session)

[−] **Team Profile.** The team profile was seen as a barrier to the adoption of EBSE in one group, both the required skills and the motivation or 'research spirit'. With respect to the latter issue, we assume that it was difficult for some participants to understand the difference between the use of secondary study outcomes and the conduct of such studies.

P08 Assuming we had to carry out a review of whatever, I think it is important that the person who does it 'wants to do it' in some way because it is not 'anyone does it', it seems to me that they have to have a research spirit. That person has to do it ... I think that is important. I think that, for example, it would be difficult for me to review this style with this ...
R2 Ah, you say? But how about the use [of SLR]?
P08 No, the use would give it an extraordinary value. But for it to be perfect, for it to be good, for it to be in the best possible way, it seems to me that the commitment of the person who carries it out must have that research spirit.

— Q06 (from discussion group 1)

[−] **Emerging issues.** One group identified that for new issues, they did not expect EBSE to be very useful, based on the perception that they were going to find few pieces of evidence. As the following excerpt shows, participants expressed interest in primary study methods that could be used to obtain quickly evidence on new issues.

P04 [...] with the emerging issues sometimes there is nothing to review, we do it by intuition, so to speak.
R2 Of course, when there is no evidence of the use of anything, one has the theory...
P04 But you see that there are things that, for example, I don't know, it occurs to me, maybe it's a bad example, but in medicine they say 'clinically proven'.
R2 That means there is evidence.
P04 Sure, sure. But they tested it, let's say, by putting cream on the patients... Even that, which is empirical, let's say, because they go and look, has a system of doing, so to speak. It is done with certain rules and other things. In order to do these empirical things, do you have a similar process model?
R1 No, there are techniques that we already mentioned: you can do interviews, surveys, focus groups...

— Q07 (from joint session)

[−] **Cost and opportunity.** Cost and its relationship to opportunity was also a theme discussed by participants in one group, as seen in the following excerpt.

P01 What I also see is an important disadvantage regarding the time that you dedicate [to undertaking an SLR], that's why I think it's important what they said about the amount of time, right? Because sometimes, for example, if you want to look for something and you are, let's say, doing the review, and the review takes two or three years. It seems to me that yes, that the cost surely may not merit what you are doing, right?
P03 Yes, opportunity cost
P08 You finish doing the review and there is another technology, something else ...
P01 Sure ...
P03 But that is part of the planning. You also have ... I think it's like any project where you have to say 'I have these restrictions to see if I get any results'. And these restrictions imply making a plan, taking them into account and determining which method [to use]...

— Q08 (from discussion group 1)

[−] **Usual paths.** Introducing a new technique requires changes in routines. A participant discussed this theme in the following excerpt.

P05 [...] what I see that there are different usual paths: if I want to incorporate a new methodology... If I want to incorporate a new technology, you usually call a consultancy firm from the market.

— Q09 (from discussion group 2)

[−] **Curb breakthrough ideas.** Finally, although one participant saw that EBSE could curb breakthrough ideas, the discussion led to another participant saying just the opposite as shown in the following excerpt.

P03 This system [EBSE] could have the risk, correct me if I'm wrong, that it could stop an idea that could be disruptive and that could come ...?
P01 The systematic review ...
P03 Maybe I'm talking nonsense...
P01 ... I see that it is like a kick for precisely looking for that. If there is no disruptive thing that you have the idea to do, if someone else didn't think of that disruptive thing. It's like a space where you can do the review and say: 'Well, I was based on this review I didn't find anything and I thought to do it this other way'.
P03 Well, it can be too.

— Q10 (from discussion group 1)

[0] **Generate data for future work.** In both groups, participants agreed that they not only found EBSE useful but also found it important to start generating data to use themselves in the future or to serve as input for future reviews. For example, in order to address the lack of government practices publications or how to make adaptations to governments requirements, as one participant explained:

> *P04 We tend to substitute profit for efficiency or social impact: management efficiency or social impact. So how do we translate profit into social impact or management efficiency? And there surely are a lot of patterns that we can draw. [...] that allows us to find clues to say to another government, 'Look, when you see a technology or something in the industry that is supported by profit, try this'.*
>
> — Q11 (from discussion group 2)

[0] **Junior researcher participant.** As mentioned above, P01 is (and was during our study) an assistant teacher at our school and a junior researcher. P01 contribution to the group meeting was then analyzed separately. Some observations about their behavior were the following:

- When another participant indicated that EBSE seemed to be able to curb breakthrough ideas, P01 responded by bringing to the discussion one of the benefits of EBSE which is understanding more about an unknown area (e.g., through a mapping study).
- P01 was the participant in the group who started the discussion about the role of the agency and the need to have more substance in its publications and how EBSE could help with that.
- On the cost of conducting a review, P01 introduced the theme of costs in the group, not only indicating the cost/opportunity ratio (something quite important when doing secondary studies), but also provided another perspective, indicating that EBSE could help in costs, for example, by helping not to reinvent solutions to a problem.

In summary, from our observations, the participant who had some understanding of SE research, seemed to have contributed a deeper perspective to the discussion and helped the group several times to discuss some aspects with a higher level of abstraction and offered different points of view.

6.2. RQ2: Effects of EBSE awareness on the working practices of GA stakeholders

Fig. 3 summarizes the results related to the impact of the EBSE lecture on the behavior of the participants. We used three participant perspectives to study this impact: any experience of using EBSE concepts, their perspectives of using EBSE concepts in the future, and any changes the lecture motivated to their work practices.

EBSE usage experiences. Only one participant (P01, the junior researcher) indicated having used any content from the EBSE lecture [S1]. In particular, P01 reported using existing secondary studies, and search engines for scientific articles [S2]. Regarding the difficulties encountered, P01 expressed:

> *P01 Putting it into practice costs a lot; you have to have training and enough time. It's very difficult first to direct towards what you're looking for and then be able to find in it what really adds value.*
>
> — Q12 (from S4)

During the analysis (March 2021) we contacted P01 by a messaging app to clarify our understanding about the answers. First, P01 confirmed using concepts from the EBSE lecture at the agency and that, more generally, excluded academic activity when answering the survey questions. In particular, P01 reported having read an SR on gamification in SE and having used search engines to learn more about quality and testing in quantum computing (both were, at that time, topics of interest to the agency). P01 also confirmed that lack of knowledge and training, and lack of time were the main difficulties faced. Furthermore, P01 commented that it was quite difficult to find relevant evidence, but strongly suspected that this was due to the lack of training. P01 pointed out:

> *P01 In comparison, I believe my lack of training and time would represent 90% of the causes of the challenges, and the difficulty of finding relevant evidence the other 10%.*
>
> — Q13 (from personal conversation through audio messages)

Prospects of using EBSE concepts. The other six participants indicated that they had not had the opportunity or need to use what they learned [S6]. On whether they would use it in the future [S7], one participant answered negatively, reporting that this was due to having very limited knowledge of EBSE. The rest answered affirmatively, and they indicated that they would apply it to do research on topics of interest, evaluate information, and compare technologies [S9]. For example, P03 wrote:

> *P03 [...] to tackle some new, infrequent topic in which we have little experience and which requires a certain level of rigor.*
>
> — Q14 (from S7)

Changes motivated by knowing EBSE. Four out of seven participants indicated that knowing about EBSE motivated them to improve or change some aspects of their working practices [S8]. In particular, they perceived changes in how they do research, assess information, and compare technologies. To illustrate, P10 answered:

> *P10 [I notice changes] mainly when doing some research on topics of interest. Now we always look for various visions on the topics we are investigating and then have our own [vision].*
>
> — Q15 (from S9)

6.3. Commonly used publication types

In the second stage, we also asked participants to specify what type of publications they commonly use in their working practices. To have more elements of analysis and verification, the questionnaire asked for types and examples of publications in different unrelated questions. Table 3 presents the mapping of the types of publications and examples used by the respondents classified according to the model of three categories for non-commercially published material of Adams et al. [44].

In summary, Adams et al. propose the following. First, they defined 'grey literature' as that not controlled by commercial organizations (e.g., government reports). Second, they called 'grey information' that which is published more informally or not published (e.g., personal notes, emails). Third, they used the term 'grey data' as an intermediate category, which is more formal and public than 'grey information' but less formal than 'grey literature' (e.g., tweets, blogs).

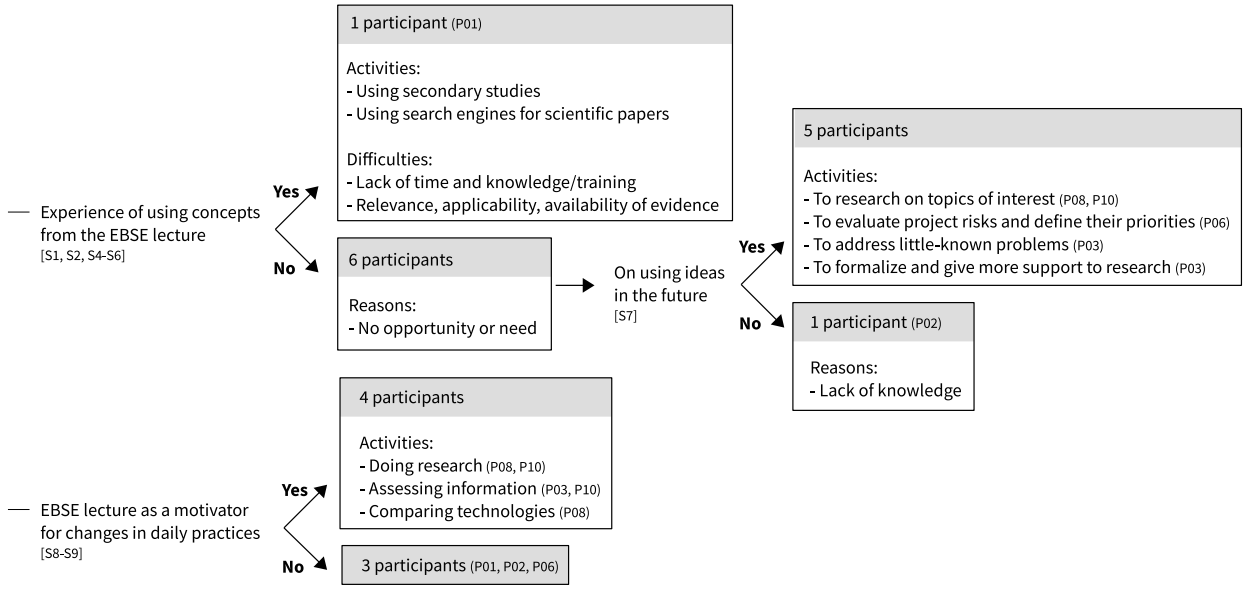


Fig. 3. On the impact of the EBSE awareness lecture on participants' working practices.

Table 3
Types and examples of literature used by participants grouped according to the model of Adams et al. [44].

Type of publications [S10]	Examples [S11]	Material type
Scientific papers (P03)	scielo.org (P09), clei2019.utp.ac.pa (P03)	White literature
Publications from other governments (P01, P03)	www.gov.uk/government/publications/synthetic-data (P01)	Grey literature
Magazines (P09)	www.technologyreview.com (P09)	
Uruguayan Institute of Technical Standards publications (P03)	–	
Research and advisory companies/organizations publications (P01)	www.iaccm.com (P03), www.gartner.com/en (P03), www.ict4v.org (P03)	
Technical documentation of products (P02)	hive.apache.org (P02), www.hitachivantara.com/es-latam/products/data-management-analytics/pentaho-platform.html (P02)	
Technical reports (P06)	publications.opengroup.org (P03)	
Internet forums (P02, P06)	es.stackoverflow.com (P02)	Grey data
Technological innovations publications (P09)	–	
Web pages (P06, P08, P09, P10)	–	

Table 3 shows the nature of the publications consulted by the participants (see questions S10 and S11 of the follow-up survey form in Section 5.3). Regarding these results, two issues are interesting to highlight.

First, on the white literature used by practitioners. Scielo is a regional platform that provides open-access papers, many of them in Spanish. CLEI is a regional computer science conference. We believe that the participants use these sources due to a lack of knowledge of other more well-known academic search engines (e.g., SCOPUS) or because they prefer quick access to literature in Spanish.

Second, it is clear that a large percentage of the responses refer to non-commercially published material. In particular, it seems that many types of literature the participants commonly use corresponds to grey literature.

7. Discussion of findings

Several aspects of our results deserve reflection and analysis.

Perceived benefits. In general, the participants believed that EBSE may be suitable for finding innovative solutions, and solving problems associated with large risks and uncertainties. In addition, they perceived that the use of EBSE would be very appropriate considering the agency's responsibilities. Both cases seem to indicate that they perceive that the evidence generated by SLRs is more reliable than other types of information. This is a similar result to that obtained by Cartaxo et al. when they presented some rapid review results to a software company staff [11].

Lack or inadequacy of evidence. Participants believed that the lack or inadequacy of research evidence, particularly evidence related to government practice or emerging issues, may be one of the top barriers to using EBSE. Although we believe that the participants do not know a great deal about research evidence (as expressed in the pre-lecture questionnaire), this limitation seems legitimate and has been reported in other disciplines as well (see Table 1). Initiatives to use grey literature (e.g., using multivocal literature reviews [45]) could help. Despite this, it seems necessary to also promote the generation of studies and evidence more suitable for industry needs.

Although the agency carries out internal research, for example, to evaluate emerging technologies, they raised the need to formalize these activities and the possibility of using primary study methods (Q07). We believe this could help them with their own research, but could also lead to doing research with other stakeholders. This seems an interesting approach since GAs often have the influence and respect needed to obtain participation in surveys or working groups from industry that would be difficult for academics to obtain.

It could be argued that the published literature does not reflect the body of knowledge (or does so in a biased way) and that therefore the quality of the evidence is not adequate for use by industry, government, or other non-academic stakeholders. To some extent, it appears that EBP has been better adopted in areas where the evidence is more quantitative and of direct application (e.g., medicine). However, it is

Table 4
Mapping of EBP reported barriers and EBSE limitations identified in this study.

Barriers to EBP adoption (2.2)	EBSE limitations discussed (Focus group)	Difficulties using EBSE concepts (Survey)
Relevance, applicability, availability, quality of research evidence	Lack of evidence from governments practices; Emerging issues	Relevance, applicability, availability of research evidence
Lack of time/workload pressure	Cost and opportunity	Lack of time
Lack of knowledge, training or inadequate skills	Team Profile	Lack of knowledge/training
Lack of resources (e.g., funding, staff, guidelines)	–	–
Lack of organizational support	Institutional support	–
–	Usual paths; Curb breakthrough ideas	–

true that currently EBP is being practiced in other areas more similar to SE (e.g., psychology, sociology, policy making). In these areas, the limitations of the existing evidence have also been detected (see, for example, [46]). These limitations can be addressed by generating new evidence and also using the evidence according to its nature and quality (e.g., for descriptive and correlational uses and not to support claims of causality). Finally, EBSE not only deals with the use of scientific evidence but also provides tools for the use of other types of material (e.g., guidelines and criteria to enable the use of grey literature).

Barriers to EBSE adoption. According to the participants, some barriers to EBSE adoption are the lack of evidence from governments, the cost of conducting systematic reviews, inadequate skills, lack of institutional support, and the lack of evidence from emerging issues. These issues are similar to those reported in relation to the adoption of evidence-based practice in general [9]. Table 4 shows a comparison of the results of this study and the summary presented about EBP barriers (see Section 2.2). Regarding the lack of resources, in Uruguay, access to scientific literature is largely provided by the government through the Timbó initiative.[6]

The cost of conducting secondary studies. Participants' perception of the high cost (in time and effort) of conducting secondary studies seems to be in agreement with that reported by researchers who conducted secondary studies [47]. There are two initiatives that address this issue. First, rapid reviews that aim to reduce the time needed to obtain an overview of existing evidence [11], and second, evidence briefings that are intended to make the results of reviews readily available to stakeholders [48]. Our results appear to support the value of these ideas.

It could be argued that practitioners should not conduct secondary studies but rather consult them. The fact that the participants considered the cost of secondary studies as negative could be related to a point that we will discuss below, and it is that they expressed some confusion between consulting and conducting secondary studies. Beyond that, we believe that the cost of applying EBSE (and not just conducting secondary studies) is an issue that needs further study.

Junior researcher participation. The role taken by the junior researcher had many characteristics in common with knowledge brokering [19]. Knowledge of both SE and the work of the agency allowed P01 to stimulate the other participants to a more in-depth discussion of EBSE-related issues. This suggests that more research is needed to evaluate the impact of knowledge brokers on EBSE adoption.

Misunderstandings of the use of evidence. On several occasions participants assumed that to use evidence it was necessary to conduct a secondary study, although it had been clarified several times before that the results of a pre-existing study could be used. Future EBSE training and awareness activities could take this into account.

Critical thinking and reflection. The focus group also seems to have contributed to the participants' reflection about their current practice and how to improve it. For example, some of them acknowledged that they do not find as much evidence about the practices adopted by similar government agencies in other countries as they would like. Then, later in the discussion, they proposed that they themselves could generate data on their experiences, something that could be useful to themselves and other governments.

Impact assessment. According to the validation survey, knowing about EBSE seems to have had some impact on the participants' practice. In general, they did not conduct or use secondary studies. Even so, it is possible that knowing about evidence-based practice has helped them to improve their research process, to be more aware of the value of (aggregating) evidence, and to improve their information literacy skills. The latter two are results similar to other reports of academic training in EBP and EBSE [49–51].

Grey literature. As well as investigating whether the participants were making use of EBSE concepts, we also sought to know what literature they commonly used. They reported using a wide variety of types of material, in particular, they use many forms of grey literature (e.g., government reports, and white papers). Due to this, it may be possible that practitioners can greatly improve their information evaluation skills by using quality assessment checklists for grey material, such as the one proposed by Garousi et al. [45].

8. Limitations

Focus groups and surveys rely on participants' reflection on, and description of, their attitudes and behaviors, and this can bias the results in several ways [52]. For example, people tend to remember events that are most important to them. We tried to mitigate its effects by requesting participants to fully explain their reflections and include examples when appropriate, however, our results must be considered with this limitation in mind.

Governments in different countries generally have different contexts and structures, and it may be difficult to find agencies similar to AGESIC in other countries. However, the positive and negative opinions about EBSE expressed by participants were very similar to those expressed by participants in other studies of EBSE and EBP.

Our work has two limitations with respect to introducing EBSE to participants. First, we had limited time to carry out the first stage. In particular, we were only able to give an EBSE class of an hour and a half. It could be argued that this is too short a time to introduce EBSE to (non-academic) participants. Their subsequent discussions suggested that the main impact was some confusion regarding the use of existing evidence and that most of the participants did understand the fundamentals of EBSE. Even so, we would be interested in carrying out longer EBSE dissemination activities (e.g., full-day or weekly workshops) and studying their effects. Second, in the interest of achieving a better response rate, we tried to use the principle of brevity [53] when creating the follow-up survey questionnaire. In particular, we omitted specific questions to assess whether the participants had correctly understood EBSE or to investigate in depth what EBSE concepts they were able to apply. Future initiatives should consider this, perhaps incorporating other research methods to assess understanding of EBSE (e.g., interviews) or better assessing their use (e.g., using observational methods).

Focus Group Limitations. Only one member of the Government Architecture division attended the focus group. Thus, the focus group results better reflect the viewpoints of staff in the other three divisions.

[6] The Timbó portal, created in 2009, enables online access to bibliography and scientific-technological literature from around the world throughout the country. It includes among others the collections of EBSEhost, IEEE, SpringerLink, ScienceDirect, and Scopus. http://www.timbo.org.uy/.

Although Tong et al. [32] suggest that the participants validate the transcripts, we opted instead for the validation of the synthesized results, which we were able to send only 16 months after the meeting. This certainly had some impact, for example, quantitatively: three participants initially did not remember having participated in the first stage of our study, two had left the agency, and we received seven responses from the nine remaining participants. However, the impact of the delay on the returned responses is difficult to assess. For similar future initiatives, we recommend using another type of validation or more steps, e.g., by incorporating validation of transcripts and then validation of synthesized results.

By using the location of the agency, one or two participants felt free to enter or leave the meeting at will. Despite this, we do not believe that this had a major influence on the results. The discussions were not interrupted or substantially changed by these events (as we checked in the analysis of the transcripts).

Survey Limitations. A major limitation of this survey is the long elapsed time between the questionnaire and the first stage which we believe has reduced the response rate and perhaps the quality of some responses. Regarding the response rate, we believe that it had no major impact, in fact, the response rate is 63%. The quality of the responses does not seem to be greatly affected either, in fact, during coding we only discarded two responses (from the same participant) for having content that did not answer the questions (P09 in S7 and S9). Future studies using the same approach should decouple the validation of the focus group results and the assessment of the impact on working practices, this would allow for the assessment to be performed in a more timely manner.

Another limitation is that, apart from P01, we cannot be sure that the use of evidence-based activities reported by other participants was actually related to EBSE (i.e., the use of evidence related to software engineering, IT, or computing issues) or that they were basically adopting techniques for searching and assessing information (e.g., searching electronic sources for information on work-related topics). However, even if they were only using techniques for searching and assessing information, their agency does undertake reports related to software engineering issues, and accepting the value of evidence is a prerequisite for adopting EBSE if a SE-related issue arises.

9. Recommendations for researchers

Research on non-academic adoption of EBSE is not a well-studied topic and as such poses certain challenges. Likewise, it is not trivial to work with GAs. Our experience in this study leads to the following recommendations.

RE1. *Consider different stakeholders and their needs.* Different stakeholders and contexts require different approaches to the introduction and promotion of EBSE. We suggest trying to understand in advance the participants' backgrounds and adapting introductory material to their needs.

RE2. *Emphasize the use of evidence rather than the production of SLRs.* Our study suggests that being aware of the use of evidence, and its implications and considerations, is more beneficial to practitioners than knowing the EBSE process or how to conduct SLRs. In particular, it seems very important to publicize the role of the evidence generated by secondary studies in relation to the other types of information available to non-academic stakeholders. For example, by indicating the differences (e.g., benefits and disadvantages) of evidence obtained by secondary studies over other types of information or available material (e.g., grey literature). We also believe that non-academic stakeholders would also benefit from learning methods to search for and appraise grey material. Another issue is that the focus group participants talked about using consultants, so it might be worth emphasizing that government agencies can commission secondary studies if there is a risk that available consultants have vested interests in particular solutions/methods.

RE3. *Value of a strong link with the stakeholder organization.* Our study was possible thanks to a strong relationship between our research group and some agency members. Given that many of them were not aware of our research nor EBSE concepts in advance, the trust and freedom they gave us were very important throughout the research process. In particular, it was very helpful to clarify the potential benefits and the confidentiality mechanisms of working together in the initial stages.

RE4. *Promote EBSE adoption research and the generation of evidence more suitable for non-academic stakeholders.* In other disciplines, there are numerous studies of EBP adoption. This allows for a greater understanding of the problems that arise in adopting evidence-based practice and how to address them. In addition, in order to allow better evaluation and aggregation, EBSE adoption studies should include detailed information on the context and research methods used. We also believe it is necessary to generate more evidence related to the needs of the software industry and government and regulatory bodies. These needs must be studied and addressed to improve the impact of EBSE and, more generally, of SE research.

RE5. *Present EBSE with reference to examples/topics of interest to participants.* A good way to introduce the evidence-based practice, already explored in other disciplines, can be to use examples and topics of interest to the target audience. This can include examples, e.g., one based on an SLR and one based on an industry study.

RE6. *Consider prior knowledge and follow-up evaluation.* Two aspects of this study could be useful in future EBSE adoption research. First, the focus group participant P01 added depth to the discussion due to their prior knowledge of EBSE. This was part of our context rather than our research design. Nonetheless, it suggests that EBSE dissemination interventions will benefit from the participation of participants with prior knowledge of EBSE who can stimulate deeper discussions. Second, our results also confirm that a follow-up survey is a useful way of gaining more reflective feedback on EBSE.

These recommendations, although useful, can be difficult to address in the context of SE, some due to their specific nature and others due to the characteristics of our discipline. To assist other researchers, Table 5 provides a list of references that support our recommendations and may help other researchers to select and implement the most appropriate recommendations for their own research context.

10. Concluding remarks

Our results indicate that participants perceived evidence resulting from SLRs as being more reliable than other types of information and suitable for certain cases (e.g., for managing innovation and risk, and to support the agency as a role model). Institutional support, lack of government practice reports, and team profile (i.e., skills and motivation) were the most discussed barriers to EBSE adoption. The barriers to EBSE adoption that we identified are very similar to those obtained in studies of EBP attitudes in other disciplines.

Related to EBSE adoption background, our analysis confirms certain assumptions or motivations that other researchers have had. First, practitioners believe the cost and timeliness of evidence are important. This has been the motivation for initiatives such as the introduction of rapid reviews in SE [11]. Second, the lack of evidence reported by the participants on certain issues (e.g., government practices and emerging issues) suggests that the use of grey literature (e.g., [45]) could have benefits for practitioners. Finally, the fact that one participant was also a junior researcher in our faculty could indicate that, as happens in other disciplines [19], the role of a knowledge broker could help improve the adoption of evidence-based practice.

Table 5
Previous studies related to our recommendations that can be used to improve the non-academic adoption of EBSE.

Ref.	Indications	Rec.
[10–12]	These three papers report studies of non-academic applications of EBSE. They can be useful to take as examples of EBSE applications or to know the current scope of non-academic use of EBSE.	RE1, RE2, RE4
[54]	It presents an in-depth analysis of the use of evidence, including its credibility, validity, relevance, fit-for-purpose, and rigor (in its creation and also in its use). Section 6 includes recommendations for researchers, as producers and consumers of evidence, and for practitioners who want to consume evidence. Information from this paper can be used when introducing EBSE to practitioners so that they can learn more about the use of evidence. For example, to help them properly consider the context of the use of evidence. In addition, the recommendations could help to improve the evidence produced in primary and secondary studies.	RE1, RE2, RE4
[55]	This paper presents 63 challenges and 127 good practices for collaboration between academia and industry from 33 articles obtained through an SLR covering the period 1996–2014. For each stage of the collaboration (i.e., problem formulation, planning, conduct & transfer, and dissemination) the paper reports challenges, good practices to address them (including references to the respective primary studies), and key findings. The recommendations could be used by researchers (with or without experience) to better carry out collaborations with other actors such as industry and government.	RE1, RE3, RE6
[56]	This paper reports a tertiary study on the use of grey literature in SE. It may be useful for researchers who want to know how grey literature is currently included in secondary studies and what challenges are faced. The authors report finding 446 secondary studies published in top venues (a minimum h5-index of 20 for conferences and 25 for journals) from 2011 up to 2018 (full list in [57]). It can be a helpful source of examples to present EBSE to practitioners or of evidence to address industry and government SE problems.	RE2, RE4, RE5
[58]	This paper includes an in-depth analysis of how to include grey material in secondary studies, including preservation issues, ethical issues, and representation of the practitioner's view. Based on this analysis, the authors recommend not including social media posts (e.g., blogs, tweets, vlogs) – material that some authors consider as grey literature – in a secondary study but rather in primary studies. In addition, it proposes a framework for how to incorporate different sources of information in the different steps of EBSE. This paper could be useful for researchers interested in considering grey material in their secondary studies, and for those who introduce EBSE to practitioners or support them in the use of evidence.	RE2, RE4
[45]	This paper presents a more permissive view (with respect to the paper presented above) as to what is considered grey literature and how to include it in secondary studies. Although we agree with the approach of how to manage grey literature of the paper presented above, we include in our review this other view that is currently recognized by other researchers. In particular, Table 7 can be used as a basis for assessing grey material, both by researchers and practitioners.	RE2, RE4
[59,60]	These papers investigate the use of blog-like documents in SE research. The first one reports a review of previous research and identifies alternatives (e.g., conducting primary or secondary studies using blogs) and resources (e.g., repositories of software-related blogs). The second one includes credibility criteria (see Table 4). It presents a methodology and tools to speed up the process of aggregating blog-like documents. This material can be helpful to researchers or practitioners looking to assess this kind of material.	RE2, RE4
[61]	This paper presents a list of 49 SLRs reporting practice-oriented recommendations published up to the end of 2015. The authors include a summary of each SR including their topic and findings. This can be helpful for practitioners looking for evidence from secondary studies on a particular topic. It can also be used for selecting examples of SLRs with which to better introduce EBSE to practitioners.	RE1, RE2, RE4, RE5
[62]	This paper presents and illustrates a set of guidelines for reporting quantitative and qualitative systematic reviews. It considers different reporting issues that may help to better consume secondary studies' results (e.g., certainty and limitations of the evidence, relationship with other evidence). These guidelines can be useful when reporting secondary studies since their application is expected to improve evidence consumption. It can also be helpful for researchers and practitioners to correctly consider the information reported in secondary studies.	RE4
[63]	This chapter presents guidelines and recommendations for conducting rapid reviews, which, as the authors indicate, are ''*lightweight secondary studies focused on delivering evidence to practitioners in a timely manner*''. In this way, it can be an interesting method to produce adequate evidence, but with certain limitations, for collaborations between academia and industry/government.	RE1, RE2, RE4
[48]	This paper introduces evidence briefings, which are one-page documents with findings drawn from secondary studies. This can be a useful format to improve the transfer of evidence to industry and government practitioners.	RE1, RE2, RE4
[51,64]	The first paper reports an SR of EBSE training initiatives reported up to the end of 2019 (including benefits and challenges), a proposal of more than forty learning outcomes for EBSE teaching, and the description and evaluation of a course based on those learning outcomes. In the second paper, a longitudinal study of the effects of training provided in three consecutive courses is presented. This material can be used as a basis to develop specific materials for dissemination and training of EBSE.	RE1, RE6

Our follow-up study confirmed that although the participants were not undertaking systematic reviews themselves, more than half of them reported improvements in how they searched for and evaluated information to support their work. These results confirm the value of evidence-based practice in the context of regulatory and government bodies. They could also indicate that promoting EBSE to non-academic stakeholders could obtain better results if it focuses on presenting the value of evidence, procedures for finding and assessing the reliability of evidence, and the use of existing SLRs.

We believe it is necessary to continue studying the non-academic adoption of EBSE. Possible future lines of research could be: studying the opinions of other key stakeholders (e.g., project and quality managers in larger established companies and software engineers involved in small start-up companies), using theories or models that allow a better understanding of evidence-based practice adoption (e.g., Rogers' diffusion of innovations theory [14]), investigate the use of knowledge brokering [19] applied to EBSE, and applying argumentation theory for a deeper understanding of the decision process of using evidence (e.g., by using the framework proposed by Rainer et al. [65]).

CRediT authorship contribution statement

Sebastián Pizard: Conceptualization, Methodology, Validation, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Fernando Acerenza:** Validation, Investigation, Writing – review & editing. **Diego Vallespir:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Barbara Kitchenham:** Conceptualization (Stage-2 only), Methodology, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

This research was partially funded by PEDECIBA, Uruguay. The authors would like to thank all the AGESIC members who participated in this study and the authorities who allowed the conduct of our study. We also want to thank Cecilia Apa for being a moderator and Natalia Milan for her translation work.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.infsof.2022.107101. It includes further information on the design, conduct of the research and characteristics of the data analysis process.

References

[1] C. Goues, C. Jaspan, I. Ozkaya, K. Stolee, Bridging the gap: From research to practical advice, IEEE Softw. 35 (2018) 50–57.
[2] B. Kitchenham, T. Dybå, M. Jørgensen, Evidence-based software engineering, in: International Conference on Software Engineering, IEEE Computer Society, USA, 2004, pp. 273–281.
[3] E. Ferlie, Conclusion: From evidence to actionable knowledge? in: Knowledge to Action?: Evidence-Based Health Care in Context, Oxford Scholarship Online, 2005, pp. 182–197, http://dx.doi.org/10.1093/acprof:oso/9780199259014.003.0009.
[4] Agency for Healthcare Research and Quality, Evidence-based practice centers (EPC) program overview, 2020, URL: https://www.ahrq.gov/research/findings/evidence-based-reports/overview/index.html.
[5] F. Bunn, D. Trivedi, P. Alderson, L. Hamilton, A. Martin, E. Pinkney, S. Iliffe, The impact of cochrane reviews: A mixed-methods evaluation of outputs from cochrane review groups supported by the national institute for health research, Health Technol. Assess. 19 (28) (2015) 1–99, v–vi.
[6] M. Unterkalmsteiner, T. Gorschek, Process improvement archaeology: What led us here, and what's next? IEEE Softw. 35 (4) (2018) 53–61, http://dx.doi.org/10.1109/MS.2018.227110005.
[7] M. Kim, Assessing attitudes towards evidence-based practice among social workers in Hong Kong, J. Soc. Work 19 (6) (2018) 769–788, http://dx.doi.org/10.1177/1468017318784077.
[8] L. Tacia, K. Biskupski, A. Pheley, R.H. Lehto, Identifying barriers to evidence-based practice adoption: A focus group study, Clin. Nurs. Stud. 3 (2) (2015) 90–95, http://dx.doi.org/10.5430/cns.v3n2p90.
[9] D. Upton, D. Stephens, B. Williams, L. Scurlock-Evans, Occupational therapists' attitudes, knowledge, and implementation of evidence-based practice: A systematic review of published research, Br. J. Occup. Ther. 77 (1) (2014) 24–38, http://dx.doi.org/10.4276/030802214X13887685335544.
[10] A. Kasoju, K. Petersen, M. Mäntylä, Analyzing an automotive testing process with evidence-based software engineering, Inf. Softw. Technol. 55 (7) (2013) 1237–1259.
[11] B. Cartaxo, G. Pinto, S. Soares, The role of rapid reviews in supporting decision-making in software engineering practice, in: International Conference on Evaluation and Assessment in Software Engineering, Association for Computing Machinery, New York, NY, USA, 2018, pp. 24–34.
[12] T. Lewowski, L. Madeyski, Code smells detection using artificial intelligence techniques: A business-driven systematic review, in: Developments in Information & Knowledge Management for Business Applications, vol. 3, Springer International Publishing, 2022, pp. 285–319, http://dx.doi.org/10.1007/978-3-030-77916-0_12.
[13] D. Budgen, P. Brereton, N. Williams, S. Drummond, The contribution that empirical studies performed in industry make to the findings of systematic reviews: A tertiary study, Inf. Softw. Technol. 94 (September 2017) (2018) 234–244, http://dx.doi.org/10.1016/j.infsof.2017.10.012.
[14] E.M. Rogers, Diffusion of Innovations, fifth ed., Free Press, New York, NY, 2003, p. 576.
[15] S. Pizard, F. Acerenza, D. Vallespir, B. Kitchenham, Supplementary materials for the paper 'assessing attitudes towards evidence-based software engineering in a government agency', 2022, http://dx.doi.org/10.5281/zenodo.6391579.
[16] L. Scurlock-Evans, D. Upton, The role and nature of evidence: A systematic review of social workers' evidence-based practice orientation, attitudes, and implementation, J. Evid.-Inform. Soc. Work 12 (4) (2015) 369–399, http://dx.doi.org/10.1080/15433714.2013.853014.
[17] L. Scurlock-Evans, P. Upton, D. Upton, Evidence-based practice in physiotherapy: A systematic review of barriers, enablers and interventions, Physiotherapy 100 (3) (2014) 208–219, http://dx.doi.org/10.1016/j.physio.2014.03.001.
[18] H. Sadeghi-Bazargani, J.S. Tabrizi, S. Azami-Aghdash, Barriers to evidence-based medicine: A systematic review, J. Eval. Clin. Pract. 20 (6) (2014) 793–802, http://dx.doi.org/10.1111/jep.12222.
[19] F. Bunn, K. Sworn, Strategies to promote the impact of systematic reviews on healthcare policy: A systematic review of the literature, Evid. Policy 7 (4) (2011) 403–428, http://dx.doi.org/10.1332/174426411X603434.
[20] J. Lomas, The in-between world of knowledge brokering, BMJ 334 (7585) (2007) 129–132, http://dx.doi.org/10.1136/bmj.39038.593380.AE.
[21] AGESIC, Notes on the creation and historical evolution of AGESIC, 2021, URL: https://www.gub.uy/agencia-gobierno-electronico-sociedad-informacion-conocimiento/institucional/creacion-evolucion-historica.
[22] Carlos Sabatino, Uruguay's Digital Development Policy, Technical Report, Global Delivery Initiative - World Bank Group, 2017, URL: https://www.effectivecooperation.org/content/uruguays-digital-development-policy.
[23] The World Bank, Expense (current LCU) - Uruguay, 2021.
[24] Instituto Nacional de Evaluación Educativa (INEEd), Total public spending in constant 2020 pesos (in millions of pesos). years 2007 to 2020, 2020, URL: http://mirador.ineed.edu.uy/indicadores/gasto-publico-total-en-pesos-constantes-de-2020-en-millones-de-pesos-28-1.html.
[25] AGESIC, Digital Government Strategy 2020, Technical Report, 2015, URL: https://www.gub.uy/agencia-gobierno-electronico-sociedad-informacion-conocimiento/sites/agencia-gobierno-electronico-sociedad-informacion-conocimiento/files/2019-01/DownloadDigitalGovernmentStrategy2020_0.pdf.
[26] G. Milano, D. Vallespir, A. Viola, A technological and innovative approach to COVID-19 in Uruguay, Commun. ACM 63 (11) (2020) 53–55, http://dx.doi.org/10.1145/3422826.
[27] J. Kontio, J. Bragge, L. Lehtola, The focus group method as an empirical tool in software engineering, in: Guide to Advanced Empirical Software Engineering, Springer London, London, 2008, pp. 93–116.
[28] K. Hannes, M. Leys, E. Vermeire, B. Aertgeerts, F. Buntinx, A.M. Depoorter, Implementing evidence-based medicine in general practice: A focus group based study, BMC Fam. Pract. 6 (2005) 1–13, http://dx.doi.org/10.1186/1471-2296-6-37.
[29] K. Hannes, D. Norré, J. Goedhuys, I. Naert, B. Aertgeerts, Obstacles to implementing evidence-based dentistry: A focus group-based study, J. Dent. Educ. 72 (6) (2008) 736–744, http://dx.doi.org/10.1002/j.0022-0337.2008.72.6.tb04539.x.
[30] K. Hannes, G. Pieters, J. Goedhuys, B. Aertgeerts, Exploring barriers to the implementation of evidence-based practice in psychiatry to inform health policy: A focus group based study, Community Ment. Health J. 46 (5) (2010) 423–432, http://dx.doi.org/10.1007/s10597-009-9260-1.
[31] D.L. Morgan, Focus groups, Annu. Rev. Sociol. 22 (1996) 129—152.
[32] A. Tong, P. Sainsbury, J. Craig, Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups, Int. J. Qual. Health Care 19 (6) (2007) 349–357, http://dx.doi.org/10.1093/intqhc/mzm042.
[33] B.C. O'Brien, I.B. Harris, T.J. Beckman, D.A. Reed, D.A. Cook, Standards for reporting qualitative research: A synthesis of recommendations, Acad. Med. 89 (9) (2014) 1245–1251.
[34] J. Kontio, L. Lehtola, J. Bragge, P.O. Box, Using the focus group method in software engineering: Obtaining practitioner and user experiences, in: International Symposium on Empirical Software Engineering, IEEE Press., New York, 2004, pp. 271–280.
[35] B.B.N. França, T. Ribeiro, P.M.D. Santos, G. Travassos, Using focus group in software engineering: Lessons learned on characterizing software technologies in academia and industry, in: Ibero-American Conference on Software Engineering, Curran Associates, Inc., New York, NY, USA, 2015, pp. 1–15.
[36] E. Zambrano, R.P.C. Reyes, J.W. Castro, E.R. C., Metrics that could be used in mobile applications development for people with visual impairments: A preliminary literature review, Rev. Iber. Sist. Tecnol. Inf. E17 (01/2019) (2019) 985–999.
[37] AGESIC, Web Quality - Accessibility Assessment (in Spanish), Technical Report, 2014, URL: https://www.gub.uy/agencia-gobierno-electronico-sociedad-informacion-conocimiento/comunicacion/publicaciones/evaluacion-de-accesibilidad.
[38] D. Silverman, Interpreting Qualitative Data, Sage, London, UK, 2015.
[39] V. Braun, V. Clarke, Using thematic analysis in psychology, Qual. Res. Psychol. 3 (2) (2006) 77–101, http://dx.doi.org/10.1191/1478088706qp063oa.
[40] C.B. Seaman, Qualitative methods in empirical studies of software engineering, IEEE Trans. Softw. Eng. 25 (4) (1999) 557–572.
[41] L. Birt, S. Scott, D. Cavers, C. Campbell, F. Walter, Member checking: A tool to enhance trustworthiness or merely a nod to validation? Qual. Health Res. 26 (13) (2016) 1802–1811.
[42] M. Kasunic, Designing an Effective Survey, Technical Report CMU/SEI-2005-HB-004, Carnegie Mellon University Software Engineering Institute, 2005.
[43] M. Torchiano, D. Fernández, G. Travassos, R. De Mello, Lessons learnt in conducting survey research, in: International Workshop on Conducting Empirical Studies in Industry, 2017, pp. 33–39.
[44] J. Adams, F.C. Hillier-Brown, H.J. Moore, A.A. Lake, V. Araujo-Soares, M. White, C. Summerbell, Searching and synthesising 'grey literature' and 'grey information' in public health: Critical reflections on three case studies, Syst. Rev. 5 (1) (2016) 164, http://dx.doi.org/10.1186/s13643-016-0337-y.

[45] V. Garousi, M. Felderer, M.V. Mäntylä, Guidelines for including grey literature and conducting multivocal literature reviews in software engineering, Inf. Softw. Technol. 106 (2019) 101–121, http://dx.doi.org/10.1016/j.infsof.2018.09.006.
[46] R.A. Maynard, The role of federal agencies in creating and administering evidence-based policies, Ann. Am. Acad. Political Soc. Sci. 678 (1) (2018) 134–144, http://dx.doi.org/10.1177/0002716218768742.
[47] B. Kitchenham, P. Brereton, A systematic review of systematic review process research in software engineering, Inf. Softw. Technol. 55 (12) (2013) 2049–2075.
[48] B. Cartaxo, G. Pinto, E. Vieira, S. Soares, Evidence briefings: Towards a medium to transfer knowledge from systematic reviews to practitioners, in: International Symposium on Empirical Software Engineering and Measurement, Association for Computing Machinery, New York, NY, USA, 2016, pp. 57:1–57:10.
[49] B. Aglen, Pedagogical strategies to teach bachelor students evidence-based practice: A systematic review, Nurse Educ. Today 36 (2016) 255–263.
[50] P. Brereton, A study of computing undergraduates undertaking a systematic literature review, IEEE Trans. Educ. 54 (4) (2011) 558–563.
[51] S. Pizard, D. Vallespir, B. Kitchenham, A longitudinal case study on the effects of an evidence-based software engineering training, in: 2022 IEEE/ACM 44th International Conference on Software Engineering: Software Engineering Education and Training, ICSE-SEET, 2022, pp. 1–13, http://dx.doi.org/10.1145/3510456.3514150.
[52] T.C. Lethbridge, S.E. Sim, J. Singer, Studying software engineers: Data collection techniques for software field studies, Empir. Softw. Eng. 10 (3) (2005) 311–341, http://dx.doi.org/10.1007/s10664-005-1290-x.
[53] E. Smith, R. Loftin, E. Murphy-Hill, C. Bird, T. Zimmermann, Improving developer participation rates in surveys, in: International Workshop on Cooperative and Human Aspects of Software Engineering, 2013, pp. 89–92.
[54] C. Wohlin, A. Rainer, Challenges and recommendations to publishing and using credible evidence in software engineering, Inf. Softw. Technol. (2021) 106555, http://dx.doi.org/10.1016/j.infsof.2021.106555.
[55] V. Garousi, M. Felderer, M.V. Mäntylä, The need for multivocal literature reviews in software engineering: Complementing systematic literature reviews with grey literature, in: International Conference on Evaluation and Assessment in Software Engineering, Association for Computing Machinery, New York, NY, USA, 2016, http://dx.doi.org/10.1145/2915970.2916008.
[56] F. Kamei, I. Wiese, C. Lima, I. Polato, V. Nepomuceno, W. Ferreira, M. Ribeiro, C. Pena, B. Cartaxo, G. Pinto, S. Soares, Grey literature in software engineering: A critical review, Inf. Softw. Technol. 138 (2021) 106609, http://dx.doi.org/10.1016/j.infsof.2021.106609.
[57] F. Kamei, I. Wiese, C. Neto, I. Polato, V. Nepomuceno, W. Ferreira, M. Ribeiro, C. Pena, B. Cartaxo, G. Pinto, S. Soares, Grey literature in software engineering: A critical review, 2022, http://dx.doi.org/10.5281/zenodo.6780520.
[58] B.A. Kitchenham, L. Madeyski, D. Budgen, How should software engineering secondary studies include grey material, IEEE Trans. Softw. Eng. (2022) 1, http://dx.doi.org/10.1109/TSE.2022.3165938.
[59] A. Rainer, A. Williams, Using blog-like documents to investigate software practice: Benefits, challenges, and research directions, J. Softw. Evol. Process 31 (11) (2019) e2197, http://dx.doi.org/10.1002/smr.2197, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/smr.2197. e2197 smr.2197.
[60] A. Williams, J. Buchan, Using the case survey methodology for finding high-quality grey literature in software engineering, in: The International Conference on Evaluation and Assessment in Software Engineering 2022, EASE 2022, Association for Computing Machinery, New York, NY, USA, 2022, pp. 1–9, http://dx.doi.org/10.1145/3530019.3530020.
[61] D. Budgen, P. Brereton, N. Williams, S. Drummond, What support do systematic reviews provide for evidence-informed teaching about software engineering practice? e-Inform. Softw. Eng. J. 14 (1) (2020) 7–60, http://dx.doi.org/10.37190/e-Inf200101.
[62] B.A. Kitchenham, L. Madeyski, D. Budgen, SEGRESS: Software engineering guidelines for reporting secondary studies, IEEE Trans. Softw. Eng. (2022) 1, http://dx.doi.org/10.1109/TSE.2022.3174092.
[63] B. Cartaxo, G. Pinto, S. Soares, Rapid reviews in software engineering, in: Contemporary Empirical Methods in Software Engineering, Springer International Publishing, Cham, 2020, pp. 357–384, http://dx.doi.org/10.1007/978-3-030-32489-6_13.
[64] S. Pizard, F. Acerenza, X. Otegui, S. Moreno, D. Vallespir, B. Kitchenham, Training students in evidence-based software engineering and systematic reviews: A systematic review and empirical study, Empir. Softw. Eng. 26 (3) (2021) 50, http://dx.doi.org/10.1007/s10664-021-09953-9.
[65] A. Rainer, Using argumentation theory to analyse software practitioners' defeasible evidence, inference and belief, Inf. Softw. Technol. 87 (2017) 62–80.

Supplementary materials for the paper *"Assessing attitudes towards evidence-based software engineering in a government agency"*

Sebastián Pizard[a], Fernando Acerenza[a], Diego Vallespir[a] and Barbara Kitchenham[b]

[a]*School of Engineering, Universidad de la República, Montevideo, Uruguay.*
[b]*School of Computing and Mathematics, Keele University, Keele, UK.*

ABSTRACT

Supplementary materials to the paper *"Assessing attitudes towards evidence-based software engineering in a government agency"*. It includes additional information on the design, conduct of the research as well as the characteristics of the data analysis process. March 2022.

This document presents supplementary material to our study *"Assessing attitudes towards evidence-based software engineering in a government agency"*. In our study, we investigated the potential value of EBSE in the context of the Uruguayan E-Government agency (AGESIC). To achieve this, our work involved three research instances subject to specific methodologies and of which a detailed report is necessary.

- In a preliminary stage, to better understand the challenges in using EBP we conducted a rapid review searching for secondary studies on its adoption.
- In a first stage, we conducted an EBSE awareness lecture to AGESIC members, followed by a focus group session to study their perceptions on the value and limitations of EBSE.
- In a second stage, we followed up our investigation with a survey that sought to establish whether the ideas presented in the EBSE lecture had had any impact on the agency working practices.

Due to the fact that the information required to fully report the conduct of a qualitative study (i.e., [12, 22]) like ours exceeds the size limits of the original paper, details of the design, conduct, and data analysis of our study are presented here. It is worth noting that ethical considerations are explained in detail in the original paper, so we do not repeat them here.

1. Rapid review on EBP adoption

Pizard planned and conducted a rapid review in June 2019, which was then updated in January 2020. The objective was to search for secondary studies on the adoption or impact of EBP or systematic reviews. It was conducted in SCOPUS with the following search string: *TITLE-ABS ((adoption OR transfer) AND "evidence based" AND ("systematic review "OR" mapping study "OR" systematic literature review "OR" slr "OR" scoping study "))*. As a result, 278 candidate studies were obtained in the initial search, and in the update 26 were added. From these 304, we selected (by reading title and abstract) the only five studies that were systematic reviews on the adoption or impact of EBP or systematic reviews. To aggregate the studies we used descriptive coding [11] and constant comparative method [19].

Rapid review results can be very limited, but we accepted this limitation since our goal was to gain some knowledge about EBP adoption in other disciplines in order to have more information to compare with our study results.

Table 1 presents the resulting systematic reviews on EBP adoption. For each systematic review we included: reference to its report (1st column), discipline to which it corresponds (2nd column), its objective (3rd column), the period it covers (4th column), the search engines used (5th column), and the papers finally selected from the total number of papers found by the search strategy (6th column).

2. Focus Group design and conduct

The focus group was conducted on August 26, 2019 at the agency's facilities. To design and conduct the focus group, we used the guidelines included in the literature [6, 8, 9]. The details of the research design, conduct and validation are presented below.

Motivation and conditions for collaboration The former director of the Information Technology area division is also a teaching assistant at Universidad de la República. In late 2018, he attended a talk our research group gave on technical debt, which included results from a secondary study. Afterwards, he suggested to us that the concepts of technical debt and evidence-based practice could be useful to the agency. Members of our research group told him that we were doing research on both topics, although he had no detailed knowledge of our research or interest in its results. Subsequently, we began conversations that led to the delivery of two lectures: one on technical debt and the one on EBSE. We also agreed that we would be allowed to study the participants' views of EBSE.

Participants' selection. A call for participation via email was made within the Information Technology area of AGESIC. Below we reproduce the lecture invitation briefing

spizard@fing.edu.uy (S. Pizard)
ORCID(s):

Study	Discipline	Purpose of study	Period	Search engines	Papers
[17]	Social workers	EBP orientation, attitudes and implementation	2003-2012	Academic Search Complete, Cumulative Index of Nursing and Allied Health Literature (CINAHL) Plus, American Psychological association databases (e.g., Psych info and Psych Articles), Medline, Journal Storage (JSTOR), and Science Direct	32/ 2302
[4]	Healthcare policy	Strategies to promote the impact of SLRs	up to 2/2010	Pubmed, CINAHL, The Cochrane Library (including the Cochrane Database of Systematic Reviews, DARE, Methods Studies, Technology Assessments and Economic Evaluations) and ISI Web of Science	11/768
[24]	Occupational therapists	Attitudes, knowledge, and implementation of EBP	2000-2012	Academic Search Complete, Cumulative Index of Nursing and Allied Health Literature Plus, PsycARTICLES, Ingenta, Medline, Science Direct, and Journal Storage	32/12990
[18]	Physiotherapy	EBP adoption - barriers, enablers and interventions	2002-2012	Academic Search Complete, Cumulative Index of Nursing and Allied Health Literature Plus, American Psychological Association databases, Medline, Journal Storage, and Science Direct	32/NS
[15]	Medicine	Barriers to EBP adoption	2000-2013	PubMed, Scopus, Cochrane library, Web of knowledge, Pro Quest, Magiran (Persian database) and SID (scientific information database – Persian database) databases	106/2592

Table 1: Some systematic reviews on EBP adoption on different disciplines

sent to the members of the Information Technology area of AGESIC.

Lecture. Introduction to evidence-based software engineering.

Places. 25 people

Description of the topic.

In general, knowledge is derived from evidence through a process of interpretation. This occurs, for example, when a scientist studies medical records to show that smoking tobacco causes lung cancer. In areas with empirical or experimental research, the evidence is obtained through observations and measurements, the results of which are reported in one or more scientific publications. In the relatively recent past, in order to have the most current knowledge on a subject, it was common for an expert to review the publications that he believed were most relevant and current. These traditional reviews have biases related to the experience of the researcher, their subjectivity, and they are not entirely reproducible.

To improve the aggregation of evidence, evidence-based practice (EBP) emerged for medicine in the 1970s. The EBP seeks to use an objective, rigorous and planned approach to select relevant studies and to synthesize the results of those studies. The methodological rigor makes the results more reliable since it is possible to study the procedure carried out to obtain it as well as to reproduce it. In medicine, EBP has been essential to help control risk factors for heart attack and stroke, to transform HIV from a fatal to a chronic infection, to test drugs for hepatitis C, and to improve treatments for some types of cancer.

The techniques used in EBP are called secondary studies, since they perform the aggregation of evidence from other studies (called primary studies). The main secondary study is the systematic literature review (SLR). SLRs make it possible to collect and synthesize evidence from different sources in a more objective and rigorous way.

The introduction of EBP in the area of software engineering began in 2004, called EBSE, and was widely accepted by researchers. It is estimated that more than 200 secondary studies have been published during the first ten years.

Meeting objective. To provide a brief introduction to evidence-based software engineering and the main practical aspects of using evidence. During the meeting, the main fundamentals of research in software engineering and a case of application of a SLR in software engineering will be presented. This lecture will help participants understand the report and results of a secondary study, they will see how it is possible to use the evidence obtained through this type of study.

Attendee profile. Practitioners of the software industry. It is not limited to professionals with university degrees in the area, but it is sufficient that they have been linked to at least one software development or implementation project.

Pre-lecture Questionnaire We created a non-compulsory pre-lecture questionnaire in Google forms and sent it to all those Agency staff who had confirmed their attendance. The translation of the questions (originally in Spanish) is shown below.

S1 What is your position or role within the organization? Explain very briefly what activities you do.*

S2 How do you obtain information to support decision making, eg. choose a technology?*

- Talking to colleagues.
- Reading blogs or internet forums.

  - Reading books.
  - Attending events or workshops.
  - Studying what others do (companies or agencies).
  - Reading specialized magazines.
  - Other...

S3 How old are you?*

S6 What is your highest completed level of education?*

  - Elementary School
  - High School
  - University Degree
  - Specialization or Master's Degree
  - PhD

S8 What is your reading comprehension level in English?*

  - Very bad, Low, Good, Very good, Excellent.

Technical initial briefing. The initial presentation had the following content:

- *Software Engineering Research* - Brief introduction to research, software engineering research, and scientific publications (e.g., purpose, parts and types of a research paper).
- *Evidence-based software engineering* - Brief introduction to EBSE, including: review process, methods within EBSE and types of issues that can be addressed with EBSE.
- *Adoption of EBSE* - Brief introduction to using an SLR. For this, we presented an SLR, published in Spanish (participants' mother tongue) and with a topic of potential interest to the participants [25]. Each stage of the review process was explained indicating its elements in the example. An example of EBSE in action applied to the industry was also briefly presented [5].

Discussion sessions design. Thirteen agency staff attended the lecture but only eleven attendees agreed to take part in the discussion, but one of them participated only in the combined session. In a first stage, the participants worked in **group sessions**. Two groups were formed (without any segmentation strategy), each one with five participants and a moderator. They were asked to discuss the benefits and limitations of EBSE. They were encouraged not only to say whether EBSE would be useful or not, but also to raise arguments or situations that would support their position. In a second stage, in a **combined session**, each group shared and discussed their reflections with the other group. In both sessions took aproximately 20 minutes.

Only the 11 participants and the three university researchers staff who ran the focus group were present during the discussion sessions.

Moderator roles Pizard (R1) made the initial presentation and gave the focus group instructions. Pizard's main PhD research concern is the use of evidence-based practice in SE and this study is part of his work. In particular, he has been studying barriers and facilitators in EBSE adoption. Recently, he authored a study of EBSE and systematic review training intended for university students [13].

Both moderators, Cecilia Apa and Fernando Acerenza (R2 and R3 in transcripts), were active researchers and were teachers of at least one of the three courses on EBSE and SLRs that were taught in recent years at our university. However, they had no vested interest in the success or failure of the present study.

All researchers had eight years or more experience of conducting lectures and tutorials. Both Apa and Pizard hold a Master's degree in Computer Science, while Acerenza has a Master's degree in Software Engineering.

Moderators had been advised to minimize intervening in the discussion so as not to influence the results. However, they could intervene when a participant made a very obvious misinterpretation of EBSE, or it was necessary to clarify something to continue the discussion.

In addition, the moderators were instructed to ensure that all the participants had the opportunity to express their opinion. This is important in order to mitigate three risks: the negative influence of authority relations [8] (i.e., some managers participated in the activity), that more assertive participants dominate the discussions, or that the audio recordings would discourage participation [10].

Discussions in both groups began with an open brainstorming session about EBSE pros and cons and ended with the moderator listing the points that were raised seeking to confirm them and expand the discussion. As requested, the moderators restricted their interventions to clarification, for example, asking a participant to provide a better explanation, or to give an example to support their statements. Also, in one group, the moderator had to explain the concept of mapping study. It was not necessary for the moderators to undertake any special actions to ensure individual participation. All participants were actively involved in the discussion.

Relevant study design decisions. Some important design decisions made relate to: participants' understanding of EBSE and study location.

As EBSE was a novel concept for most of the participants, there was a risk that they misinterpret some of its aspects or the purpose of the focus group. In order to mitigate this, we adopted two strategies, a simple presentation was prepared and specially adapted to the participants, and special attention was paid during the data analysis to identify possible misinterpretations.

The location choice is an issue with certain consequences [1]. We were invited to hold the meeting at the agency's facilities. We did not propose an alternative, since we believe that conducting the study at the agency's location would encourage a high level of participation. Any off-site location would involve moving participants around town and risk needing to conduct the study outside of the participants normal working hours, two issues that would discourage participation.

Participant validation. In December 2020, we contacted the agency and sent participants a summary of the study results[1]. Seven participants responded, five from the IT Optimization division, one from Emerging Technologies, and one from Governance Architecture. They agreed with the results summary and did not suggest any changes.

3. Focus Group Data analysis

Each moderator took notes during the discussions and all the sessions were recorded in audio.

In order to analyze the data we used qualitative thematic analysis. Thematic analysis is generally described as *'a method for identifying, analysing and reporting patterns (themes) within data'* [3], and is one recommended approach for analyzing focus group data [20].

As theoretical framework, we used an essentialist or realist approach. In this approach, it is assumed that there is a largely unidirectional relationship between meaning and experience and language, from which it is possible to study motivations, experiences and meanings directly from the expressions of the participants [3].

To carry out the analysis we followed a process adapted from thematic analysis [3, 20], whose main characteristics are: familiarization with the data, initial coding, and identification of themes together with quotes that illustrate them.

We started with a **(1) verbatim transcription** of all audio recordings (including each group session and combined sesion), generating a total of 19 pages of data.

We had three distinct data sets: group session transcripts, combined session transcripts, and moderators' notes. During the combined session the participants reviewed the opinions they had expressed in the group sessions. The moderators' notes also summarized the group sessions. So, we decided to conduct an in-depth analysis of the group session transcripts and then validated them with the information from the rest of the data set.

Subsequently, transcripts were uploaded to Saturate[2], a freeware software tool that allows coding and categorizing texts. Then, we used descriptive coding and the constant comparative method [19] to **(2) classify all the positive and negative opinions** of the participants in relation to EBSE made during the group session. Our unit of analysis (and coding) were phrases within each participants' interventions (or an entire intervention, if applicable). We consider an intervention as a fragment of conversation in which a single participant speaks and that can be preceded or succeeded by interventions made by other participants.

Later, we **(3) identified discussion themes from the codes used and we selected quotes to illustrate each theme**. At the end of this step we had an idea of what each theme was and the overall story the data tell about the participants' opinions about EBSE.

[1] This is a method of validating results sometimes referred to as member checking [2].

[2] http://www.saturateapp.com/

We **(4) verified the previous analysis results and extended it using the transcript of the combined session and notes by moderators**. In this step, more insightful quotes or additional elements were included. No new themes arose from this step.

To report the results, as the language used by the participants was Spanish, we worked with an experienced translator to create accurate English versions of the selected quotes.

Up to this point, all the analysis was carried out by Pizard. Subsequently, Acerenza carried out independent checking of the analysis process and its results. Acerenza's suggestions and comments on the analysis were discussed with Pizard until they reached agreement on all issues.

To improve reliability (e.g., to avoid misinterpretation of the data), we had a rigorous analysis process planned in advance. We also used three approaches to improve reliability [14]: systematic tracking of all data, triangulation (recording of audios and notes by moderators) and peer debriefing (i.e., the participation of two researchers instead of only one). In addition, to avoid misinterpreting the data, we used participant validation [14]. For this, as mentioned before, we sent a brief report of the results (including all the selected quotations) to the participants for their validation.

4. Survey Design and Conduct

The main purpose of the survey was to collect the participants' opinions about any impact that knowing EBSE had on their work practices. In addition, to better understand the context in which the participants could make use of the evidence generated by the SLRs, we sought to investigate what type of literature the participants usually utilize.

The survey was conducted simultaneously with the validation of the first stage-results. It was designed and conducted following recommendations and guidelines proposed in the literature [7, 23]. The survey design and conduct considerations are described below.

The participants were the unit of analysis for the survey and the group of AGESIC members, that participated in the first stage of our investigation, was the target population. Participation was optional but not anonymous.

Pizard created the questionnaire based on recommendations from Vallespir and Kitchenham. The final version was reviewed by Vallespir and a study participant (our contact at the agency at that time). In essence, the survey instrument was a self-administered questionnaire with eleven questions (some of which were open questions, to allow the participants to explain their answers more completely). The questionnaire was designed in Google Forms, and was available for one month (from December 23, 2020 to January 26, 2021). The translation of the questions (originally in Spanish) and the flow of the survey are shown below.

S1 Have you used any ideas seen in the EBSE lecture in your working practices?*

- Yes

- No *[Skip to question 3 and after to question 6]*

S2 Indicate what you have done

- Conducting a secondary study (systematic review, mapping study, or rapid review).
- Using an existing secondary study (systematic review, mapping study, or rapid review).
- Using search engines for scientific articles.
- Undertaking a critical appraisal of reports that compare different technologies.
- Other activities, indicate which.

S3 Indicate your first and last name. This information will be used to relate your answers to previous results.*

S4 Of the ideas you used, did you find anything particularly difficult?*

S5 Of the ideas you used or their results, did you find anything particularly useful?* *[Skip to question 8]*

S6 What is the main reason why you did not use anything seen in the EBSE lecture?*

- The ideas do not fit the way of working.
- I have had no chance or need.
- I tried to use the ideas but was unsuccessful.
- The lecture was not enough for me to understand how to apply the ideas of evidence-based practice.
- Other reasons, indicate which.

S7 Do you think that your knowledge about EBSE could be used in the future? For what?*

S8 Did knowing about EBSE MOTIVATE you to improve or change any activity in your working practices?*

- Yes *[Skip to question 9]*
- No *[Skip to question 10]*

S9 In which activities, tasks, problems or situations do you perceive that you have changed after learning about EBSE?*

S10 What kind of literature do you use in your practice at the agency? *

S11 Could you indicate titles or links (this is preferable) to two typical examples of the literature you use?*

To minimize the risk of participants misunderstanding what the survey was about or not remembering the EBSE lecture, we included a paragraph summarizing EBSE purposes and a very brief description of the meeting held in the first stage of this research.

To improve the survey capacity to correctly collect the information required to answer the research question, we took two actions. First, the questions were very specific and the most relevant to our research required participants to provide examples or justifications. In addition, in order to relate the participants' responses with their characteristics (e.g., agency division or academic degree) and their participation in the previous stage of this study, we requested that they identified themselves with their first and last names.

To improve survey participation, we used some principles listed by Smith et al. [21]. We applied brevity by including as few and as specific as possible questions, and designing a survey dynamic presentation (i.e., items and response options participants receive are based on their prior choices). We applied the authority and credibility principles, along with social benefit, by stating that their participation would contribute to our research and to gain a better understanding of EBSE adoption. In addition, as it is specifically recommended to improve the response rate in follow-up studies [16], we asked our contact at the agency to send two reminder emails (spaced approximately by two weeks).

5. Survey Data analysis

To analyze responses of S1, S2, S6, and S8 we counted the number of responses in each response category. In S4, S5, S7, S9, and S10 the textual responses were collated and we used descriptive coding and constant comparative method to classify the responses (without pre-existing codes [S5, S10], or use as initial set of codes based on the activities presented in subsection 5 [S7, S9] and the barriers presented in Table 2 and in subsection 5.2 [S4]). The responses from S11 (i.e., examples of publications used by participants) were listed and mapped with the results from S10 (i.e., types of publications used). The result of this mapping was also classified into the grey literature layers. This analysis was performed by Pizard. Later in a meeting, Pizard and Acerenza discussed the analysis process and its results, agreeing on all the decisions made.

For the open questions (those with textual responses) we sought to use the thematic analysis approach. Although in these cases the data sets were quite small, we decided not only to code but also to identify themes from the codes used, understand what each theme was about, review their consistency, and identify quotes that illustrate each of them.

References

[1] Barbour, R.S., 2005. Making sense of focus groups. Medical Education 39, 742–750. doi:10.1111/j.1365-2929.2005.02200.x.

[2] Birt, L., Scott, S., Cavers, D., Campbell, C., Walter, F., 2016. Member checking: a tool to enhance trustworthiness or merely a nod to validation? Qualitative health research 26, 1802–1811.

[3] Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. Qualitative Research in Psychology 3, 77–101. doi:10.1191/1478088706qp063oa.

[4] Bunn, F., Sworn, K., 2011. Strategies to promote the impact of systematic reviews on healthcare policy: A systematic review of the literature. Evidence and Policy 7, 403–428. doi:10.1332/174426411X603434.

[5] 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, Association for Computing Machinery, New York, NY, USA. pp. 24–34.

[6] França, B.B.N., Ribeiro, T., Santos, P.M.D., Travassos, G., 2015. Using focus group in software engineering: lessons learned on characterizing software technologies in academia and industry, in: Ibero-American Conference on Software Engineering, Curran Associates, Inc., New York, NY, USA. pp. 1–15.

[7] Kasunic, M., 2005. Designing an effective survey. Technical Report CMU/SEI-2005-HB-004. Carnegie Mellon University Software Engineering Institute.

[8] Kontio, J., Bragge, J., Lehtola, L., 2008. The focus group method as an empirical tool in software engineering. Springer London, London. chapter 4. pp. 93–116.

[9] Kontio, J., Lehtola, L., Bragge, J., Box, P.O., 2004. Using the Focus Group Method in Software Engineering: Obtaining Practitioner and User Experiences, in: International Symposium on Empirical Software Engineering, IEEE Press., New York. pp. 271–280.

[10] Marrelli, A.F., 2008. Collecting data through focus groups. Performance Improvement 47, 39–45. doi:https://doi.org/10.1002/pfi.201.

[11] Miles, M. B., Huberman, A. M., Saldana, J., 2014. Qualitative Data Analysis: A Methods Sourcebook. 3 ed., SAGE Publications, Ltd.

[12] O'Brien, B.C., Harris, I.B., Beckman, T.J., Reed, D.A., Cook, D.A., 2014. Standards for reporting qualitative research: a synthesis of recommendations. Academic Medicine 89, 1245–1251.

[13] Pizard, S., Acerenza, F., Otegui, X., Moreno, S., Vallespir, D., Kitchenham, B., 2021. Training students in evidence-based software engineering and systematic reviews: a systematic review and empirical study. Empirical Software Engineering 26, 50. URL: https://doi.org/10.1007/s10664-021-09953-9, doi:10.1007/s10664-021-09953-9.

[14] Runeson, P., Host, M., Rainer, A., Regnell, B., 2012. Case Study Research in Software Engineering: Guidelines and Examples. John Wiley & Sons, New Jersey, US.

[15] Sadeghi-Bazargani, H., Tabrizi, J.S., Azami-Aghdash, S., 2014. Barriers to evidence-based medicine: A systematic review. Journal of Evaluation in Clinical Practice 20, 793–802. doi:10.1111/jep.12222.

[16] Salkind, N., 2010. Encyclopedia of Research Design. SAGE Publications.

[17] 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, 369–399. doi:10.1080/15433714.2013.853014.

[18] Scurlock-Evans, L., Upton, P., Upton, D., 2014. Evidence-Based Practice in physiotherapy: A systematic review of barriers, enablers and interventions. Physiotherapy (United Kingdom) 100, 208–219. doi:10.1016/j.physio.2014.03.001.

[19] Seaman, C.B., 1999. Qualitative methods in empirical studies of software engineering. IEEE Transactions on software engineering 25, 557–572.

[20] Silverman, D., 2015. Interpreting qualitative data. Sage, London, UK.

[21] 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.

[22] Tong, A., Sainsbury, P., Craig, J., 2007. Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. International Journal for Quality in Health Care 19, 349–357. doi:10.1093/intqhc/mzm042.

[23] 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.

[24] 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, 24–38. doi:10.4276/030802214X13887685335544.

[25] Zambrano, E., Reyes, R., Castro, J., C., E., 2019. Metrics that could be used in mobile applications development for people with visual impairments: A preliminary literature review. RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao E17, 985–999.

Appendix C

Evidence-based Software Engineering Application in Industry

Empirical Software Engineering manuscript No.
(will be inserted by the editor)

Using Rapid Reviews to Support Software Engineering Practice: A Systematic Review and a Replication Study

Sebastián Pizard · Joaquín Lezama · Rodrigo García · Diego Vallespir · Barbara Kitchenham

the date of receipt and acceptance should be inserted later

Abstract Context: A few years ago, rapid reviews (RR) were introduced in software engineering (SE) to address the problem that standard systematic reviews take too long and too much effort to be of value to practitioners. Prior to our study, few practice-driven RRs had been reported, and none involved collaboration with practitioners lacking SE research experience.

Objective: To investigate practitioners' perspectives on the use of RRs in supporting SE practices, we aimed to validate and build upon the findings of the seminal RR in SE study, specifically considering practitioners without explicit SE research experience.

Method: First, we studied previously conducted RRs in SE through a systematic review. Second, we carried out an external replication of the first study that proposed the use of RRs in SE. Specifically, we conducted an RR for an agile software development team looking to improve its knowledge management practices.

Results: Most of the software development team's perceptions about RR results were positive and strongly consistent with previous research. In particular, RR results were considered more reliable than other sources of information and adequate to address the problems detected. Some months later they confirmed using some of the recommendations.

Conclusions: The results show that practitioners without explicit SE research experience appreciate the value of evidence and can make use of the results of RRs. However, SE research may need to be translated from broad recommendations to specific process change options. Our research also reveals that SE RRs reporting needs to be substantially improved.

Keywords rapid review · action research · replication study · industry–academia collaboration · systematic literature review

S. Pizard · J. Lezama · R. García · D. Vallespir
Engineering School, Universidad de la República, Montevideo, Uruguay

D. Vallespir
Pyxis Research, Pyxis, Montevideo, Uruguay

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

1 Introduction

Cartaxo et al. (2018) introduced the concept of rapid reviews (RRs) to software engineering (SE) researchers to address concerns that standard systematic reviews (SRs) take too long and require too much effort to be of significant value to the software industry. They explained that, *"RRs are lightweight secondary studies focused on delivering evidence to practitioners in a timely manner"*. To achieve this, reviewers simplify or skip steps from full SRs. For example, RR process options include only having one reviewer, omitting quality assessment of the studies, or limiting the literature search.

RRs were first used in health care, and have proved to be extremely useful, especially in emergency situations. For example, during the COVID-19 pandemic more than 3000 RRs were conducted (Tricco et al., 2022). They are also recognized as a good option for low-resource settings, in which there is no capacity to produce a full SR. Given this recent explosion of RRs in medicine, they have been studied in greater depth to reach agreement on their characteristics and methods (see, for example, Tricco et al. (2022); Mijumbi-Deve et al. (2022); Kelly et al. (2022); Wilson et al. (2021); King et al. (2022)).

RRs could become a valuable resource in SE for several reasons. In a similar way to what happened in the health field, they would make it possible to obtain recommendations to address problems or improve practice based on offering evidence in short time scales or for low-resource settings. Moreover, because they are designed to tackle issues collaboratively with practitioners, RRs could establish a connection between EBSE and professional practice in a manner that SRs have not yet accomplished (Rico et al., 2024). For example, Hassler et al. (2014) identified a significant barrier to SRs in the lack of industry connection, and Cartaxo et al. (2016)'s survey of Stack Exchanges users revealed that SRs often fail to address practitioners' specific questions. A tertiary study examining 120 SRs by Da Silva et al. (2011) found that only 32 providing recommendations for users, highlighting a gap in their practical applicability. Furthermore, Kitchenham et al. (2015) discussed only a single paper reporting a direct EBSE application in industry up to 2015 (Kasoju et al., 2013).

In their original paper, Cartaxo et al. (2018) evaluated the perceptions of some practitioners on the use of an RR to support decision-making in an industrial setting. The results were promising. Practitioners perceived that RRs offer reliable results and reduced decision-making time, and help better understood the problem and potential solutions. Two months later, the practitioners reported that they had adopted some of the evidence provided. In subsequent publications, Cartaxo et al. (2019, 2020) encouraged RR producers to publish their experiences and the feedback received from companies or practitioners to better inform SE researchers of the benefits and limitations of RRs.

We agree with Cartaxo and his colleagues, that SE researchers and practitioners would benefit from having more knowledge about the practical use of RRs, and this is the rationale for this replication study. In particular, our main focus is on determining whether RRs can assist practitioners in utilizing the findings of SE research, even in the absence of prior experience in SE research.

To fulfill this objective, we (1) conducted a systematic review on SE RR to explore the characteristics of RR adoption in our field, and (2) carried out a replication of the original study by Cartaxo et al. (2018).

Our replication involved assessing the use of RR in SE, with a particular emphasis on their potential to assist practitioners without experience in SE research. Specifically, it implied conducting an RR designed to assist a software company in addressing a real-world problem and evaluating practitioners' perceptions of the RR results. The systematic review, although smaller in scope, allows both positioning our replication and its results with respect to previous SE RR research and an overview of how RR studies are used in SE research.

The rest of this paper is structured as follows.

- In Section 2 we briefly outline the main characteristics of RRs both in the health field and in SE.
- In Section 3, we report a systematic review of the RRs published in SE.
- In Sections 4 to 7, we used Carver (2010)'s guidelines for reporting experimental replications and the principles and criteria for Action Research (the research method used both in the original study and in our replication) proposed by Davison et al. (2004). Specifically:
  - Section 4 presents the main characteristics of the original study by Cartaxo et al. (2018)
  - Section 5 presents the main aspects of our replication, this includes our research purpose (5.2), the research context (5.3), ethical issues (5.4), research steps (5.5), variations from the original study (5.6), and data collection and analysis procedures (5.7).
  - Section 6 presents details of the RR protocol and conduct.
  - Section 7 presents the results of our replication.
- Section 8 presents a discussion of the results of our systematic review and our replication.
- We present our concluding remarks in Section 9.
- Finally, Appendices I and II include additional information on the conducted RR and the data analysis of our replication study, respectively.

It is essential to point out that the replication was conducted before the systematic review. However, the systematic review is presented first in the paper to provide readers with more detailed information about RRs in SE and healthcare, enabling them to better evaluate the methodology and significance of our replication. The systematic review did not influence the conduct of the replication but provides additional information to assess the benefits and risks of RRs in SE.

2 Rapid Reviews

This section introduces RRs and their application in SE. First, we summarize the types of literature, and subsequently provide a brief overview of the definition and characteristics of RRs, as they are commonly used in the healthcare field. Second, we summarize the main characteristics of RR in SE, as proposed by Cartaxo et al. (2018). Finally, we outline the main differences between RRs and other types of literature reviews.

2.1 Types of Literature Reviews

Ralph and Baltes (2022) classified literature reviews in the following types.

Ad Hoc Reviews: These involve discussions of specific literature, e.g., commonly found in the related work section of most research papers. Researchers use purposive sampling, i.e., selectively choosing papers or studies that are useful, relevant, or supportive of their arguments. This unsystematic approach introduces sampling bias and challenges replication.

Systematic Reviews: These literature reviews follow a systematic, replicable process for selecting primary studies. The management and analysis of these primary studies define the type of systematic review. Apart from rapid reviews, which will be discussed in the next subsection, the types of SRs include:

- **Meta-analysis:** An archetypal SR that examines a set of randomized controlled experiments with the same independent and dependent variables.
- **Meta-synthesis:** Also known as thematic analysis, narrative synthesis, meta-ethnography, and interpretive synthesis, this approach involves aggregating qualitative studies. Meta-synthesis serves as the constructivist counterpart to meta-analysis.
- **Case Survey (aka Case Meta-analysis):** These reviews convert qualitative accounts (i.e., case studies) into a quantitative dataset for null-hypothesis testing.
- **Critical Reviews:** These reviews analyze a selection of primary studies to support an argument or critique. They mainly contribute to the meta-scientific discourse, which involves the internal discussions within a scientific community about research practices.
- **Scoping Reviews (aka Systematic Mapping Studies):** These reviews aim to understand the current state of research on a particular topic, typically by categorizing primary studies. Scoping reviews are generally mostly descriptive.

Ralph and Baltes (2022) defined RRs as a form of meta-analysis that *"makes methodological compromises to reduce completion time"*. They argue that RRs are only justifiable if waiting for a full review would be harmful, and that such circumstances are rare in software engineering.

2.2 RRs in Health Field

In the absence of a consensus, the Cochrane Rapid Review Methods Group proposed the following definition of an RR: *"A rapid review is a form of knowledge synthesis that accelerates the process of conducting a traditional systematic review through streamlining or omitting various methods to produce evidence for stakeholders in a resource-efficient manner"* (Hamel et al., 2021).

RR stakeholders (aka knowledge users) are individuals likely to use research evidence to make informed decisions across various fields such as health, education, policy-making, and industry (Garritty et al., 2024). This group includes public partners, practitioners, policymakers, educators, and industry leaders. Their involvement in conducting RRs aims to produce relevant, useful evidence for real-world applications, enhancing evidence integration into practice and decision-making. Moreover, RRs are often commissioned by decision-makers and organizations when urgent decisions are needed, typically due to resource constraints like limited funding (Garritty et al., 2024). Among the stakeholders, requesters are distinguished by their role in requesting the RR, defining the information needs,

participating in the RR process (e.g., validating results), and disseminating the results (King et al., 2022).

Several mechanisms can be employed to strike a balance between timeliness and resource constraints when conducting a rigorous knowledge synthesis process to inform practice. Many researchers concur that there is no universally applicable approach to conducting RRs, and often, the choice of mechanisms should be tailored to the specific topic under investigation and the stakeholders' needs (King et al., 2022; Watt et al., 2008). Recent studies surveyed the most commonly used practices in RRs (King et al., 2022; Garritty et al., 2021) and identified the following key characteristics of RRs.

The most fundamental key to success is maintaining early and continuous engagement with the research requester King et al. (2022). This collaboration helps focus the RR and ensures its alignment with the stakeholders' needs. It is also recommended to discuss with stakeholders their expectations for communication in advance and adapt the reporting and dissemination approach to prioritize practical needs (Kelly et al., 2022).

Methods can be streamlined at all stages of the review process, from the initial search to synthesis King et al. (2022). This can be achieved by limiting the search in terms of dates and language, restricting the number of electronic databases searched, assigning a single reviewer for study selection and data extraction (often with verification by another reviewer), limiting risk-of-bias assessment for the most important outcomes, and opting for a descriptive synthesis rather than a quantitative summary.

Like conventional SRs, the protocol serves as the starting point for the review, however, methodological decisions often evolve iteratively, involving requester participation Wilson et al. (2021). Any modifications to the protocol should be reflected in the final report. Researchers must transparently document their methodological choices, which should be communicated to stakeholders, to ensure that the evidence review is suitable for its intended purpose. Since the potential bias introduced by these choices may not be evident, transparency is indispensable King et al. (2022).

2.3 RRs in SE

Cartaxo et al. (2018) proposed RRs as a means to transfer knowledge from research to SE practice. They identified some key characteristics of RRs in SE, which, slightly adapted, are:

- **Timely results and reduced costs:** In general, SRs are produced in relatively long timelines (several months or years) by a team of reviewers. Instead, RRs have shorter timelines that seek a compromise between the needs of practitioners and methodological rigor, considering timelines between days, weeks, and months depending on the stakeholders' needs (Wilson et al., 2021). To achieve these reductions, various strategies can be used, e.g., participation of a single reviewer, not conducting quality assessment, using specific questions, or using tables to quickly map and summarize the findings (Cartaxo et al., 2018; Wilson et al., 2021; King et al., 2022).
- **Collaboration with practitioners:** A key aspect to achieve a successful uptake of the RR results is a close collaboration with the decision makers who

requested the RR (also called review requesters) (Tricco et al., 2022). This collaboration should start at an early stage in order to understand the requesters' information needs and expectations (Garritty et al., 2021; King et al., 2022).
– **Appealing media:** Cartaxo et al. (2018) highlight as another key aspect asserting that the results of an RR should be presented in formats that appeal to practitioners rather than the conventional research paper format. They specifically advocate for the use of evidence briefings, defined as one-page documents summarizing findings from secondary studies (Cartaxo et al., 2016), as a potential method to report the outcomes of an RR.

2.4 Differences between RRs and other types of reviews

First, ad hoc reviews are conducted on an as-needed basis to address specific, often immediate questions across various contexts. They are more flexible, less structured, and can vary widely in methodology and duration. While RRs balance speed with some level of systematic thoroughness, ad hoc reviews prioritize immediate relevance and flexibility over comprehensive rigor.

Second, while RRs focus on quickly synthesizing evidence to meet urgent needs (usually with a concise and targeted scope), often with less methodological rigor and resource investment, SRs generally aim for a thorough and exhaustive synthesis of the literature, ensuring high methodological rigor and comprehensive analysis (Ralph and Baltes, 2022). RRs provide a more limited analysis of the available evidence, focusing on key findings and major trends rather than in-depth synthesis, which is typical of SRs as they seek to compile robust and nuanced evidence.

Third, in the case of mapping studies (or scoping reviews), their aim is to offer a broad overview of the research landscape, identifying gaps and trends in a specific field (Ralph and Baltes, 2022). They typically categorize and map the literature without necessarily synthesizing detailed findings from primary studies. This differs from RRs, which focus on rapidly synthesizing evidence to address urgent decision-making on specific questions.

3 SE Research Using RRs

In order to assess take-up of RRs in SE, we conducted a systematic review of RR research based on citation analysis of two of Cartaxo's papers (Cartaxo et al., 2018, 2020). Our research questions were:

– RQI: What is the extent of take-up of RRs in the SE domain?
– RQII: What was the scope of these studies?
– RQIII: What are the methodological characteristics of the reported RRs?
– RQIV: Which studies contributed to assessing the value of RRs and what have they found?

We searched for citations of Cartaxo et al.'s papers on the publication sites of the two papers in their respective publication sites (ACM and Springer respectively) and Google Scholar. The process of our review had two iterations and its complete details an expanded results can be found in the supplementary material to this paper. The last search took place on November 27, 2023, and identified

150 citations. We found 23 publications of interest, including articles from conferences or journals and book chapters (Ponce et al., 2019; Radu, 2020; Baldassarre et al., 2023; Marchetto, 2023; Matalonga et al., 2022; Abdelfattah and Cerny, 2023; Păvăloaia and Necula, 2023; dos Santos et al., 2023; Hidalgo et al., 2024; Lonetti et al., 2023; Paes et al., 2023; Barletta et al., 2023; Baldassarre et al., 2021; Pizard et al., 2023; Fritzsch et al., 2023, 2022; Rufino Júnior et al., 2023; Furukawa et al., 2022; Song et al., 2022; Loli et al., 2020; Bjarnason et al., 2023; Motta et al., 2023; Rico et al., 2024). Below RQs are answered using tables and summaries. Cross-references to study characteristics are reported in the supplementary material.

3.1 RQI: Extent of take-up of RRs in the SE

We found 23 papers reporting RRs in SE, although one of them analyzed the process used by two previously published RRs (Rico et al., 2024). Since this study includes additional objectives and information to the original studies, we maintain it in our following analysis except in situations where otherwise indicated. 15 of these studies were published in 2023 (prior to Nov 23).

Researchers from 13 countries participated in the studies. Table 1 shows the number of papers and the number of researchers from the different countries of affiliation of the authors (some authors have more than one affiliation). A co-author of the original RR in SE studies participated in one of the studies (Loli et al., 2020).

Country	Number of Papers	Number of Researchers
Brazil	7	24
Italy	6	15
Sweden	3	9
Portugal	2	5
Germany	2	8
Romania	2	3
UK	2	2
Chile	2	4
France	1	2
Uruguay	1	3
USA	1	2
Spain	1	1
The Netherlands	1	1

Table 1 Number of papers and researchers from the countries of authors' affiliations.

3.2 RQII: Scope of the studies

Thirteen of the studies report an RR aimed at acquiring knowledge in a specific field, with some studies explicitly indicating this intent while others do not (though they also refrain from reporting any alternative use or motivation). The remaining 10 studies had broader goals, as shown in Table 2. For example, some of them

sought to complement or validate the RRs' results, while others used the RRs' results to develop a model, catalog, or artifact.

Table 2 Scope of SE RRs.

	Number of Studies
Purpose of the study[1]	
Conducting an RR.	13
Creating a model/catalog/artifact using the RR results as input.	3
Conducting an RR and complementing/comparing its results with stakeholders' opinions.	6
Investigating the RR methodology.	1
Type of stakeholders[2]	
Software industry	6
University students	2
Government agency	1
Other industry	1
Stakeholder participation[2]	
RR results are reported as starting point for future collaboration with stakeholders.	2
Stakeholders participated throughout the RR process.	3
Stakeholders validated RR results.	2
Stakeholders validated the model, catalog, or artifact created.	3[3]
Stakeholders participation was used to complement RR results.	2
Stakeholders participation's results were compared to RR results	2

[1] Including all 23 papers reporting SE research using RRs.
[2] Not including (Rico et al., 2024) as it analyzes two previously published RRs.
[3] Two of these studies involved conducting families of RRs.

Nine studies report conducting an RR and complementing/comparing its results with stakeholders' opinions (refer to Table 2). Additionally, in another study, the RR served as a starting point for collaboration with stakeholders, but neither their participation nor feedback is reported.

Table 3 shows topics addressed by RRs in the 22 studies excluding (Rico et al., 2024). The number of topics covered is greater than the number of studies because some studies reported multiple RRs on different topics. Studies that reported multiple RRs on the same topic were counted once only. Table 3 confirms that 13 of the most commonly addressed topics related to the software process (i.e., software construction, software testing and software maintenance), and technology impact was another frequently addressed topic.

The vast majority of RRs included only white literature. However, three RRs include grey literature and another one was conducted in conjunction with a grey literature review. Two studies aimed to search for software tools (the selection ends in Gitlab or Github) and another study searched for both models and tools (both within white literature).

3.3 RQIII: Methodological characteristics of published RRs

Table 4 summarizes the reporting limitations and process changes that we found in the papers reporting RRs.

Topic	Number of Studies
Software construction	5
Technology impact	5
Software testing	4
Software maintenance	3
Software tools	2
SE professional practice	2
SE models and methods	2
Software design	1
Software process	1
Evidence-based practice	1

Table 3 Topics of the RRs conducted in SE, not including (Rico et al., 2024) as it analyzes two previously published RRs.

RR Process Issues	# of Studies
RR Reporting Issues Decreasing Transparency and Reproducibility	
Did not report synthesis methods adequately	18
Did not include the full date of the search	15
Did not report number and roles of reviewers	9
Only reported number and roles of reviewers for some stages	8
Did not cite primary studies	7
Did not mention limitations	6
Reported results only via an Evidence Briefing	2
Process Changes that could Bias RR Recommendations	
Omitted risk of bias evaluation (i.e., quality assessment)	21
Used a single search engine	12
Used a single researcher for 1 or more stages	8
Used a subset of the studies found in searches	4
Included additional studies without explanation	3
RR Risk Reduction Processes	
Used a single search engine complemented with snowballing	4
Used an Evidence Briefing with additional commentary about RQs	2
Used only primary studies cited in related SRs	1
Used tools to assist analysis and classification	1

Table 4 Reporting issues and process characteristics of the RRs conducted in SE, not including (Rico et al., 2024) as it analyzes two previously published RRs.

In most studies, adequately considering the RR process or its results is challenging due to insufficiently reported information. In general, detailed information is provided about the sources used and the information search stage, less information about selection and extraction. But the synthesis stage is the worst reported (only four studies report it adequately).

Additionally, inadequate reporting of synthesis methods hinders reproducibility in most SE RR studies. Barletta et al. (2023) and Motta et al. (2023) are the only exceptions, offering comprehensive details crucial for reproducibility, including details such as the date of search, the list of primary studies, and synthesis methods used.

Finally, there is confusion about the term "protocol". Some studies use it correctly for the artifact that guided their RR conduct, while others use the term to refer to the generic RR process. Only five studies mention using a protocol (Furukawa et al., 2022; Matalonga et al., 2022; Motta et al., 2023; Rufino Júnior et al., 2023; Song et al., 2022), and only one makes the protocol available to readers(Motta et al., 2023).

3.4 RQIV: Studies assessing the value of RRs

Six studies confirmed the value of RRs by validating their outcomes (or the models created from them) through collaboration with stakeholders outside the review team.

- Furukawa et al. (2022) conducted an opinion survey with IT professionals (75% out of 20 participants having postgraduate level education) to validate the RR outcomes.
- Song et al. (2022) conducted an RR as a starting point in a collaboration with a software company.
- Loli et al. (2020) created a catalog of object-relational mapping code smells in java using results of an RR. The researchers undertook an opinion survey to validate the results (97% out of 86 participants with, at least, a bachelor's degree).
- Bjarnason et al. (2023) worked collaboratively with a company. In particular, three RRs were conducted to create a software selection model, which was validated with a focus group and an application in the company.
- Motta et al. (2023) conducted seven RRs to develop a roadmap for IoT development. To validate it, an experimental study was carried out where undergraduates used and evaluated the roadmap.
- In Rico et al. (2020), the authors analyzed the artifacts of the RR reported in Song et al. (2022) and one of the RRs reported in Bjarnason et al. (2023). They also conducted interviews with the RR review teams, aiming to gain a better understanding of how RRs were conducted.

Although the feedback reports are not extensive or detailed, in all studies, stakeholders had a positive attitude towards the results confirming the value of RRs. Except for the study with undergraduates, the others included practitioners with education achievements of at least degree level.

In the three studies in which researchers collaborated with companies throughout the RR process (Bjarnason et al., 2023; Song et al., 2022; Rufino Júnior et al., 2023), the practitioners were technology experts who specialized in topics related to RR questions. In two of those studies in which the results were validated with practitioners (Bjarnason et al., 2023; Song et al., 2022), their perceptions of the results were positive, and their feedback was used to improve the RRs results. Despite the fact that in one of the studies, the results were not directly applicable for practitioners (Song et al., 2022), both studies support the view that RRs are useful in industrial contexts.

Rico et al. (2024) analyzed the artifacts of the RR reported in (Song et al., 2022) and the first of the RRs reported in (Bjarnason et al., 2023) to evaluate the

application of recent guidelines for conducting RRs in collaboration with practitioners, as well as to comprehend the benefits and challenges associated with RRs. The results confirmed that the guidelines were adequate. Conducting RRs collaboratively benefited the relationship between researchers and practitioners, fostering an understanding of expectations and establishing a common terminology. The main challenges included divergent needs of the review team and the industry collaborators, inadequacy of the evidence found (necessitating the use of broad questions), and concerns about short timelines (RR conduct took a few months but with low weekly effort).

3.5 Conclusions

The majority of papers (13 of 23) did not comment on the validity of the RR process or the risks entailed by process changes made to reduce timescales and/or effort neither did they appear to have a specific knowledge user. Of the remaining papers, the outcomes of the RR were used by the reviewers, or the value of the RR outcomes was assessed by industry collaborators, or the validity of the RR outcomes were assessed by other participants.

Current SE procedures for RR conflict with Healthcare recommendations, as described by Kelly et al. (2022). In particular current RRs often fail with respect to targeting a specific knowledge user and communicating directly with them, working from a protocol, reporting in sufficient detail to support reproducibility, and identifying the methodological strengths and weaknesses of the review. With respect to targeting a specific knowledge users, Rico et al. (2020) proposed guidelines for interactive RRs by revising Cartaxo et al. (2018)'s RR process to include practitioners. His guidelines also include producing a protocol, and reporting to the practitioners involved and disseminating the results to other practitioners and academic audiences. In our opinion, the guidelines identify by Rico et al. (2020) address most of the problems in current SE RRs, but they could be improved by including an assessment of the methodological strengths and weaknesses of the review.

The studies by Bjarnason et al. (2023); Song et al. (2022); Rico et al. (2024) confirm that RRs can support collaborations between industry and academia. However, like the RR conducted by Cartaxo et al. (2018), the industry collaboration was in the context of advanced R&D projects and involved practitioners with experience in SE research —a limitation already identified by Cartaxo et al. (2018). Our study addresses this research gap by undertaking an RR aimed at collaboration with industry participants who had no previous experience with SE research or the use of evidence.

4 Original study on RR in SE

In their study, Cartaxo et al. (2018) not only proposed the use of RRs in SE but also reported an empirical evaluation of the perceptions of practitioners about the outcomes of an RR conducted to address a problem identified in their practice.

Research Question. The researchers asked: *What are practitioners' perceptions on using Rapid Reviews to support informed decision-making in software engineering practice?*

Context & Participants. To answer this question, the researchers introduced RRs to offer empirical evidence aimed at enhancing customer collaboration in agile software development projects conducted by an applied research institution in Brazil, which offers services such as software development, applied research, and consultancy. The project aimed to develop a monitoring system for reusable packages during the production chain of the automotive industry. The practitioners who participated in the research were the project coordinator (who was the leader of all project managers) and one of the project managers. They both had a master's degree in computer science.

Research Design. Action research was used as a research method. In particular, a cycle of the canonical action research process was conducted, as proposed by dos Santos and Travassos (2011a). This process has the following five stages: Diagnosis, Planning, Intervention, Evaluation, and Reflection.

- In the first stage the researchers explored and established the problem practitioners would help address. As an instrument, they used semi-structured face-to-face interviews.
- In the planning stage they decided to use an RR and developed the protocol in collaboration with the requesters.
- The intervention consisted of conducting the RR and transferring the results to the practitioners. The latter was carried out through the preparation of an evidence briefing and a workshop to disseminate the evidence.
- The evaluation stage consisted of two semi-structured interviews, one during the results dissemination workshop and a second interview carried out two months later.
- The reflection stage included reflecting and reasoning about the previous steps.

To report and disseminate the results of their RR, Cartaxo et al. (2018) utilized an evidence briefing. Subsequently, their study examined practitioners' perceptions of both RRs and the utilization of evidence briefings as a means of disseminating their findings. The participants often used scientific papers and were, therefore, able to compare the use of evidence briefings as an alternative means of receiving scientific evidence.

Results. The results show that practitioners had a positive perception of RRs information. They reported benefits such as learning new concepts, reducing the time and cost of decision-making, and improving their understanding of the problems they faced. Two months after transferring the RR results, practitioners had adopted the evidence provided. The study also identified some improvements to the RR process, such as the need for discussing the findings of the RR, avoiding printing the RR report in black-and-white, and including graphical information in the report. However, not all the evidence provided by the RR was found to be useful by the participants, as some strategies were already in place or could not be implemented. Overall, the study demonstrated the potential of RRs in transferring research knowledge to SE practice.

5 Replication Information

As stated above, our study replicates the research conducted by Cartaxo et al. (2018). This involved conducting a Rapid Review (RR) to provide information for a software company addressing a real-world problem. We assessed practitioners' perceptions of considering RR results to enhance their practice, both immediately after sharing them and a few months later.

Reviewing Cartaxo et al. (2018) results, we found that their RR's recommendations were SE concepts rather than direct support for decision-making. For example, a recommendation is *"Customer Proxy: Some agile teams use a customer proxy —a member of the development team coordinating with the customers—to secure requirements and feedback"*. In particular, the recommendations were not presented as a comparative evaluation of a set of alternative methods to address the requesters' problems. So, we consider both the original study and this replication to be an evaluation of the use of RRs to support practice in SE, not an evaluation of RR support for decision-making.

Below, we present the key features of our study, which include the motivation behind conducting the replication, our research objectives and questions, the context and research design, and finally, a summary of the differences from the original study.

5.1 Motivation

Replication of empirical studies is a fundamental activity in the construction of knowledge in all empirical sciences (Da Silva et al., 2014). The significance of replications can be viewed from two angles: first, to validate or broaden the findings of previous studies, and second, to understand the effects of new variables, including those introduced by changes in the environment (Kitchenham, 2008).

However, despite the growing trend in the number of published replications, the actual number of replication reports remains small (Cruz et al., 2020). Consequently, many techniques and methods are proposed for SE based on single empirical studies. This creates difficulties in persuading other researchers, and also practitioners, to adopt our techniques since they have not found substantial empirical evidence and the available evidence may not align with their specific contexts (Weyuker et al., 2011).

Our motivation for undertaking this replication is primarily focused on further investigating RRs. In light of the positive outcomes reported by Cartaxo et al. (2018), we became intrigued by the potential of RRs to support practitioners wanting to address problems that they themselves identified. In particular, we were interested in collaborating with practitioners who did not have experience in scientific research. Our research goals were viable because Pizard and Vallespir worked at the Universidad de la República in Montevideo, Uruguay and, therefore, had links with the Uruguayan software industry, which has the highest per capita turnover in Latin America (Escalante and Fagúndez de los Reyes, 2022). It includes a substantial number of small and medium-sized enterprises, staffed practitioners with different levels of education and training.

5.2 Research Goals and Method

As stated above, this is a replication of Cartaxo et al. (2018)'s study and we have adopted a similar methodology to that used by the original authors, in order to facilitate comparisons of our results. However, we concentrate primarily on the issues arising from collaboration with industry, not those concerned with the format of evidence briefings. This means our research question is limited to the benefits and limitations of RRs as perceived by the review requesters. Arising from this goal our research questions are:

- RQ1: What are the perceptions and attitudes about using a rapid review to support software engineering practice in a software company?
- RQ2: Are there any problems using RR information when collaborator have no SE research experience?

With RQ1 we sought to replicate the objective of the original RR study in SE. Meanwhile, we proposed RQ2 to further study the gap that we identified in the systematic review.

To address these questions, like Cartaxo et al. (2018), we used action research, a method that integrates research with exploratory actions that promote change (Denzin and Lincoln, 2017). In particular, researchers and participants (e.g., company representatives) perform an action or tackle a problem by working collaboratively, and then evaluate and reflect on the results (dos Santos and Travassos, 2011b). Action research supports addressing problems in a pragmatic way without sacrificing scientific rigor. It also promotes reflection and knowledge generation. This makes it particularly useful for conducting field research.

We can define our replication as follows (Ralph et al., 2021):

- A methodological replication - same objectives and research method but a different practitioner problem and a different context (Dennis and Valacich, 2015). Our context, notably, includes practitioners without explicit experience in SE research.
- A partial replication - addresses a subset of the original research question (Carver, 2010). Although the research question seeks the same in both studies, as we mentioned before the original study also sought to evaluate the use of evidence briefings, which we did not replicate.
- An external replication - the replication team does not share members with the original study team (Da Silva et al., 2014). It is also necessary to clarify that we did not have any direct contact with the group of researchers who carried out the original study.

5.3 Research Context

Here we describe certain aspects of the research context.

The company: A UK company specializing in digital out-of-home (DOOH) advertising[1]. The company's IT department, located in Uruguay, was responsible for developing and maintaining a platform to manage advertising campaigns. More information about the company and the software development team is included below (see Table 5.5.1).

The requesters: The technical product leader and the project manager accompanied all the stages of the project, from the diagnosis of the problems to the dissemination of evidence. They answered questions, carried out intermediate validations, and received the recommendations obtained from the rapid review. For the purposes of this study, we have considered them as review requesters. The review requesters were the staff members responsible for identifying and introducing any required changes to the software development process. During the first meeting, the reviewers asked them about their educational level and use of information sources. Their educational level was Intermediate[2] (one with *upper secondary education* and the other one with *post-secondary non-tertiary education*). As sources of information for supporting practice they usually talked to colleagues, read technology forums or blog articles, and watched technology videos (e.g., from the Microsoft youtube channel). Neither of them consulted scientific literature.

Other stakeholders: Three other members of the development team also participated in the dissemination workshop of RR results. At the beginning of the workshop, they were asked what sources of information they usually used, and they indicated that they neither knew nor consulted scientific literature.

The reviewers: Reviewers were Lezama, García and Pizard. Pizard has ten years of industry experience as a technical lead and software quality manager, and twelve years as a teaching assistant at the university. This study is part of his doctoral research that focuses on investigating EBSE adoption. García and Lezama were about to finish their computer science degrees. Participating in the RR was part of their capstone project. Also, both of them had full-time jobs related to software development. In particular, Lezama was also part of the company's development team and was so during the first half of RR conduct. Both of them were trained in EBSE, specifically, in the planning and conduct of SRs. The training was led by Pizard and based on an EBSE and SRs course he teaches (Pizard et al., 2021, 2022).

5.4 Ethical Issues

Our university did not require our study to be approved by an ethics committee. However, given the participatory nature of action research, ethical aspects must be carefully considered, and, in particular, it is imperative to ensure that the processes are transparent to all participants (Stringer, 2007). Both the company members and the students (i.e., García and Lezama) were informed of the purpose and nature of this research prior to their consent to participate.

In addition, two other ethical considerations were:

[1] Digital out-of-home advertising (DOOH) is advertising designed to reach consumers when they are not at home and that is also dynamically and digitally displayed. This includes digital transit, digital billboards, and digital place-based displays.

[2] According to UNESCO's ISCED 2011 classification. https://ilostat.ilo.org/resources/concepts-and-definitions/classification-education/

– Ensuring that the students' education experience was not adversely impacted by the study: (1) Students should not be required to undertake tasks beyond their capabilities. This concern was addressed by appropriate training and supervision. (2) The students were assured that the outcome of the study in terms of whether or not it was favorable to the use of RRs would not impact their capstone project marks.
– Ensuring that the interests of the company are not adversely affected by the study: (1) The company would receive the best scientific information to help them address their process issues. This was assured by the personal experience and supervisory role of the first author. (2) Commercially or personally sensitive information would be kept confidential or anonymous as appropriate. Specifically, only the company and roles are identified, and specific comments are not attributed to specific individuals.

5.5 Research Steps

Our study used the five steps recommended for action research in SE (dos Santos and Travassos, 2011b). These were diagnosis, planning, intervention, evaluation, and reflection.

5.5.1 Diagnosis

In this step, we sought to better understand the company and the knowledgement management (KM) problems they had. We did this by conducting a face-to-face meeting at the company's offices. The meeting, which followed a script[3], had the following parts:

a. We explained the purpose of the study;
b. We briefly presented EBSE and secondary studies.
c. We asked for information from each participant, e.g., their position and seniority in the company.
d. We asked about the project and the team's characteristics.
e. We asked about the KM problems they were facing.
f. Finally, we agreed on how to continue working.

The meeting lasted ∼1h45min and later its recording was transcribed to facilitate our analysis.

In summary, requesters felt comfortable with their software development methodology (Scrum with two-week sprints). They reported that although they had certain KM practices, they faced several problems with the management of knowledge and documentation used in the software development and maintenance process. Furthermore, they had tried some solutions without any significant success.

The company used a continuous improvement process that involved identifying small improvements that were to be introduced in future sprints. Specifically, they told us that the results of our project were going to be incorporated into that improvement process, so they preferred a set of small specific recommendations rather than a single large-scale change.

[3] For a version of the script, in English, contact the first author.

As a validation of this diagnosis activity, we shared a summary of the context of the company and its KM problems with the requesters, who formally approved it. This document is reproduced in Table 5.

Context
The company, a spin-off of a UK advertising agency specializing in digital out-of-home (DOOH), is responsible for the entire life cycle of four products that make up a platform whose purpose is to facilitate the management of the advertising agency's campaigns. The staff is geographically distributed. Outside of Uruguay are the CEO (US) and four account executives (UK), who are the platform users and maintain contact with the end customers. The software development team is located in Montevideo: product owner, project manager, architect, three back-end developers, one front-end, a DevOps manager, and a QA. They use Scrum with two-week sprints. There are daily stand-up meetings and at the end of each sprint, the progress is validated through demos with the stakeholders (the CEO, the account executives, and some employees of the advertising agency). Planning and retrospectives are also carried out. Sprint by sprint, improvements identified during the retrospective meetings are introduced and evaluated experimentally. The team produces and stores a lot of documentation using a variety of different online tools. - GitHub is used to store the code and installation instructions. - Lucidcharts are used to document architecture diagrams, flows, roadmap planning, processes (e.g., service integration), and retrospectives. - To document test cases and test scenarios, the Visual Studio Online (VSO) wiki is used. - To keep the backlog, the VSO board is used. - To track tasks such as integration stages, product discovery, and technical debt, Trello boards are used. - To document the architecture (architecture decision record, ADR) or make spreadsheets (e.g., notification management) Google Docs is used. The responsibility for maintaining each document or tool change according to the needs. The only document that has a pre-defined structure is the ADR. To share knowledge, one or two team members usually prepare internal workshops, which sometimes include coding dojos.
Problems related to KM
Difficulties related to the management of the documentation already created. - It is hard to find the right document (there are duplicates or different ones but with the same purpose, in different media, and from different dates). - It is also complicated to keep the documentation up-to-date and to deprecate, or directly eliminate, unnecessary or out-of-date documents. The decision of what type of document to create usually depends on the person who is going to do it. There are no standard definitions, so it is hard to decide what documentation to produce. An exception to this is ADRs which do have a defined structure and the team agrees that had positive results. The company perceives as a problem the centralization of knowledge in some roles covered by a single person in the team. QA and DevOps manager roles are perceived as riskier, both roles with only one member.

Table 5 Results of the Diagnostic Stage: Context and problems related to KM.

5.5.2 Planning

At this stage, we defined the RR protocol (see Section 6 for more details). In particular, we performed several preliminary searches on Scopus and other search engines to check that appropriate evidence existed to address the issue. As a result of these searches, we selected Scopus as our search engine, refined our search string, and specified our selection criteria. We did not agree on a timeline since the students would only be available on a part-time basis and it was the first time that they would conduct an RR. We stipulated carrying out an intermediate validation so that the requesters could validate a sample of the evidence. We also specified that the dissemination would be done through an evidence briefing and a workshop.

5.5.3 Intervention

In this step, we conducted the RR and disseminated the results. We carried out an intermediate face-to-face validation of an evidence sample and undertook several consultations via email. In addition, because Lezama worked in the company for a while, the requesters answered his queries and asked him questions informally. The RR was done in three and a half months and the total time spent by all reviewers was ∼150hs, including team meetings and knowledge dissemination transfer activities with practitioners.

5.5.4 Evaluation

We circulated two questionnaires to assess requesters' perceptions of the RR results and the challenges faced during its conduct, one at the end of the dissemination workshop (in which company software developers also participated) and another follow-up questionnaire eight months later. We also recorded the workshop in audio (with prior approval) to analyze the attendees' initial attitudes. In addition, the reviewers held a retrospective meeting to reflect on the RR process and its results, in which, for example, the stages and activities carried out were recalled, and the main challenges and facilitators were discussed. As the meeting did not directly address the RQS, the insights gathered from this meeting were employed as input to the discussion of results (refer to Section 8).

5.5.5 Reflection

As stated by dos Santos and Travassos (2011b), this step is intended to investigate two issues 1) whether the results of the action delivered the results reported in the literature, and 2) the learning experience of the participants and its impact on the organization.

Because the goal of our study was to assess the value of the RR to support practice, we investigated attitudes of the company staff to the evidence and recommendations provided by the RR, and compared our results with those reported by Cartaxo et al. (2018). Thus, in this study, we only considered the second issue mentioned by dos Santos and Travassos (2011b).

5.6 Variations from the original study

We consider that the most significant difference with the original study by Cartaxo et al. (2018) is that our study considers practitioners lacking experience in SE research. However, there are several differences with the original study, primarily arising from the specific conditions and context of our research. Table 6 provides a concise summary of the disparities between the two studies, with detailed information on each point provided earlier in this section.

5.7 Data Collection & Analysis

Data Collection. Data were collected about the organization, the roles of all participants, the actions of the participants, and the outcomes of the action research process.

In particular, the opinions of requesters and software development team members on the RR results were captured at different instances:

- During the workshop, when one of the requesters and members of the development team openly discussed RR information.
- At the end of the workshop, through a questionnaire circulated to the requesters and members of the development team, which included closed questions and space for additional comments.
- Eight months later, through a follow-up questionnaire to requesters, asking about the use of the RR results.

Additionally, as previously stated, the reviewers held a retrospective meeting in which they revisited the activities carried out during the research, reflected on the usefulness of the RR results, and discussed the challenges encountered during both the RR process and the workshop aimed at disseminating the results.

Throughout the research, audio recordings were made of all meetings, and all emails and communications among reviewers, requesters, and other stakeholders were collected. In addition, internal documents, reports for requesters, and other artifacts (e.g., documents for intermediate validation of evidence) were collected. Finally, the report of the students' capstone project and the video recording of the capstone project defense presentation (including the discussion with committee members) were included in the dataset. All the material is in Spanish.

Data Analysis. Data analysis was strongly based on thematic analysis with a realistic approach (including coding, theme identification, and selection of illustrative quotes) (Braun and Clarke, 2006).

Initially, the audio recordings of the meetings with requesters, the results transfer workshop, and the reviewers' retrospective meeting were completely transcribed.

In the first stage, Pizard carried out an inductive analysis to identify (i) all stages and activities carried out during the research, (ii) the requesters and stakeholders' opinions on the RR information, and (iii) the challenges faced by the reviewers. First, a subset of the data was coded (i.e., requesters and stakeholders' responses to questionnaires, transcription of the workshop and the retrospective meeting, and the final report of the students' capstone project). Subsequently, all elements of the dataset were reviewed in chronological order to confirm and

	Original Study (Cartaxo et al., 2018)	Our Replication
The company	An applied research institute in Brazil.	A UK company specializing in digital out-of-home advertising with an IT department located in Uruguay.
The project	Developing a monitoring system for reusable packages during the production chain of the automotive industry.	Developing and maintaining a platform to manage out-of-home advertising campaigns.
Participants (requesters)	The project coordinator, the leader of project managers, and one of the project managers, both held master's degrees.	The technical product leader and the project manager, both lack graduate degrees and have no experience in SE research.
Team members (reviewers)	A Ph.D. candidate supervised by two professors.	A Ph.D. candidate and two undergraduate students finishing their 5-year computer engineering program, supervised by a professor.
Issues addressed with RR	Practitioners face issues due to low customer collaboration in their agile software development projects. They seek evidence to enhance customer relationships and mitigate these problems.	Practitioners face challenges in managing knowledge and documentation in software development despite having certain KM practices. They sought proven strategies or ideas to address these issues.
Research goals	To evaluate the perceptions from practitioners on the use of RR to support decision-making in SE practice. Perceptions about using evidence briefings for disseminating scientific results obtained from RR were also evaluated.	To evaluate the perceptions on the use of RR to support SE practice. We did not evaluate participants' perceptions about the use of evidence briefings as an alternative means for disseminating scientific evidence since they did not consume this type of evidence.
Research design	A single cycle of action research was conducted, encompassing diagnosis, planning, intervention, evaluation, and reflection phases. The intervention involved conducting an RR and sharing its results with practitioners through a workshop. The evaluation comprised two semi-structured interviews: the first was conducted during the workshop for disseminating RR results and the second two months later.	A single action research cycle, with the same stages, was undertaken, incorporating an RR as part of the intervention, with results shared among practitioners through a workshop. The evaluation involved observing and analyzing initial perceptions during the workshop and circulating two self-administered questionnaires (one at the workshop's end and another eight months later). A large part of the software development team took part in the workshop, which involved an activity to assess and prioritize the recommendations derived from the RR.

Table 6 Differences between the original RR study in SE (Cartaxo et al., 2018) and our replication.

update the codes. ATLAS.ti was used to assist the analysis (Hecker and Muhr, 2022). A preliminary report was prepared with the identified themes and illustrative quotes, which was translated into English and validated and reviewed by Vallespir and Kitchenham. At their suggestion, Pizard re-examined the dataset to confirm or expand the analysis. This process was repeated until no new themes were identified, achieving the *inductive thematic saturation* applicable to this type of analysis (Saunders et al., 2018). The results of this stage of the analysis were used to prepare sections 6, 5, 8.3, and 8.4 of this document and as input for the next stage of the analysis.

In the second stage, a deductive analysis was conducted to confirm the themes identified by Cartaxo et al. (2018). For this, requesters and stakeholders' responses to questionnaires and representative fragments of the transcription of the workshop were translated to English. Using the list of themes identified by the original study by Cartaxo et al. (2018), Pizard and Kitchenham checked which were confirmed in the replication and identified illustrative quotes. This activity continued until no more themes could be confirmed from the original study's list, thus achieving the *a priori thematic saturation* applicable to this type of analysis (Saunders et al., 2018). Themes identified in the previous stage that were not present in the original list by Cartaxo et al. (2018) (e.g., company situation) were added to the list of themes confirmed by our study. The result of this stage of the analysis was used to answer the research questions. An expert translator helped create the most accurate English versions of the selected quotes.

To increase transparency, the cross-reference between identified themes and the supporting data according to the collection stage is reported in Appendix I.

As suggested by Runeson et al. (2012), we tried to improve reliability by systematically tracking all data. Pizard also kept a detailed journal of the decisions made during the investigation, including the data analysis process. We used the O'Brien et al. (2014) checklist for reporting qualitative research to enhance clarity and completeness. We also considered the eight criteria for the quality of qualitative research proposed by Tracy (2010).

6 Rapid Review Protocol & Conduct

This section presents the RR protocol and the main characteristics of the RR conduct. This includes the details of the problem addressed and the knowledge dissemination activities carried out to share the results of the RR with the company's staff.

The RR was conducted by García and Lezama with supervision and support from Pizard.

6.1 Aim and Research Question

Together with the requesters, we determined that the purpose of the RR would be to find practical and applicable recommendations on KM for the company. For these recommendations to be applicable, the context of the studies must be similar to that of the company, and ideally, the recommendations must have been experimentally validated.

The question that guided our RR was: ***What are some empirically validated recommendations for knowledge management for software development companies?***

In preliminary searches, we encountered some difficulty finding studies on KM with practical recommendations. Thus, we decided to use a broad question and consider a general software development industry context, so we did not restrict the context of our primary studies to small or midsize companies using agile methods.

6.2 Search Strategy

We used a single search engine, *Scopus*, in order to reduce time and effort. We choose Scopus because it is one of the largest repositories of peer-reviewed publications and has been used for other SE RRs (Baldassarre et al., 2021; Cartaxo et al., 2018). We did not consider grey literature to maintain consistency with the original study by Cartaxo et al. (2018).

The search string we used (reproduced below) was obtained after trialing several preliminary searches. These confirmed that our search would retrieve studies with experimentally validated recommendations in industry settings.

> TITLE-ABS-KEY((“software engineering” OR “software development”) AND (“knowledge management” OR “knowledge sharing”) AND (industry OR company OR enterprise OR organization OR organization) AND (“case study” OR “systematic review” OR slr OR “scoping study” OR “mapping study” OR “lessons learned” OR recommendations OR survey))

The four sets of terms correspond to the following concepts: software engineering, knowledge management, industry context, and experimental validation. The search string is quite complex, but, unlike many digital search engines, Scopus respects the rules of Boolean algebra, so does not generally produce large numbers of spurious results.

6.3 Study Selection

The inclusion criteria used were: (1) studies in English, (2) with practical recommendations, (3) full text available, (4) about KM in software development companies. In addition, the exclusion criteria were: (5) publications that were not articles (books, technical reports, etc.), and (6) studies about KM theoretical models or frameworks.

Initially, we attempted to incorporate selection criteria to exclude studies that lacked applicability. Specifically, we sought to limit the results to industrial evaluations of small companies as needed by the requesters, but there were insufficient papers available.

We do not explicitly exclude secondary studies. We identified them just in case the primary studies did not include useful evidence. In this situation, we planned to use information from the secondary studies or consider them for snowballing.

This was not necessary and we were able to obtain practical recommendations from the selected primary studies.

Each reviewer assessed the eligibility status of half of the candidate primary studies obtained from Scopus. To validate an adequate level of agreement, they reviewed the first 30 candidate primary studies together, and the kappa statistic was calculated (the obtained value was 0.618, which indicates a good level of agreement). Subsequently, the reviewers used two rounds: (1) checking titles and abstracts and (2) reading the full text. From the 425 studies returned by Scopus, we identified 21 primary studies (Birk and Dingsøyr, 2005; Viana et al., 2013; Yglesias, 1998; Matturro and Silva, 2010; Viana et al., 2015; Vasanthapriyan et al., 2017; Gervigny and Nagowah, 2017; Soini et al., 2007; Santos et al., 2014; Smite and Dingsøyr, 2012; Khan et al., 2012; Kammani et al., 2013; Pulkkinen. et al., 2007; Soini, 2008; Heredia et al., 2017; Šmite et al., 2017; Heredia et al., 2014; Humayun et al., 2013; Jurado et al., 2015; Milovanović et al., 2012; Chen et al., 2018).

6.4 Data Extraction

In this step, each reviewer extracted data from half of the selected studies and validated the data extracted by the other reviewer. We used a data form that summarized: (i) the context of the study (software development method, year of the study, type of company), (ii) the research method used, (iii) the results of the validation or evaluation, and (iv) recommendations for practice or lessons learned.

6.5 Assessment of the relevance of the evidence

We assessed the relevance of the evidence from the viewpoint of the RR requesters. Our appraisal of the evidence included understanding the context in which the evidence was generated and the research methods used. For example, for organizations that participated in a primary study, we identified in which country they carried out their activity, and their size. This made it possible to validate the evidence with the requesters (as mentioned before) and to select evidence most appropriate to their context. It is important to note that the requesters indicated that these contextual factors should be used. While other factors such as organizational domain or maturity could have been considered for a more nuanced assessment of relevance, only the factors mentioned earlier were employed.

6.6 Synthesis Procedure

Although we explicitly looked for primary studies with experimental validations, most did not include recommendations for practice but rather offered lessons learned, reflections, or certain observed behaviors. We analyzed those observations, as well as the context, and developed practice-oriented recommendations. We understand that this is similar to the process carried out by Budgen et al. (2020). As an example, the extract: *"The findings showed that the motivation for knowledge sharing, a time-consuming and demanding activity, is highly related to the*

awareness that managers and developers have of the benefits associated with this professional practice." (Chen et al., 2018) was translated into *"It is recommended to emphasize the benefits of sharing and reusing knowledge."* This activity involved continuing with the critical appraisal of evidence since we used our previous analysis of the studies when preparing the final text of the recommendations.

Subsequently, we used content analysis with an inductive approach, adapted from (DeFranco and Laplante, 2017), to synthesize these recommendations. The stages carried out were: labeling of the different types of recommendations using open coding, grouping, and categorization of the fragments according to their codes (e.g., grouping similar or complementary codes), and, finally, creation of descriptions. As a tool to facilitate the coding, we used the freeware tool Saturate (http://www.saturateapp.com/).

In a similar way to the previous steps, each reviewer worked on one-half of the studies and validated the other half. In particular, they both did some initial coding. Subsequently, in two meetings with Pizard, all the recommendations and their coding were reviewed, and the categories and descriptions of the final recommendations were created.

The recommendations to improve KM obtained as outcomes of this stage are presented below in the Report/Diffusion section. The cross-reference between recommendations and primary studies from which they emerged is reported in the Appendix I.

6.7 Intermediate Validations

When starting the study selection, we carried out two validations with the requesters to assess the evidence. They approved the sample evidence we presented to them. Their main concern was to obtain recommendations from contexts similar to their own. They also indicated that some recommendations of the sample (e.g., a recommendation on defining simple and clear KM processes) seemed useful to them but they did not know how to put them into practice. Based on this, we decided that we would include if necessary, examples or brief guidelines to put the recommendations into practice. In addition, we agreed not to consider evidence on KM frameworks or models, as it was preferable to propose recommendations that could be introduced into their improvement process.

6.8 Report/Diffusion

We developed an evidence briefing with the RR results (see Figure 1) and conducted a workshop with the software development team. In the workshop, which lasted ∼1h30min, we presented the recommendations obtained in the RR and proposed a practical exercise in which the participants studied and prioritized the recommendations. The project manager and three other members of the software development team participated in the meeting. Table 7 details the activities of the workshop.

Results of the practical exercise are shown in Figure 2. After placing all the recommendations in the quadrant, the participants discussed which of them they could implement and how. They all agreed that the definition of a KM strategy had

KNOWLEDGE MANAGEMENT IN SW PROJECTS

This briefing presents scientific evidence on strategies to improve KM in software development projects.

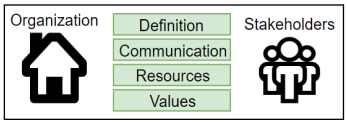
RECOMMENDATIONS

All the recommendations presented in this report are a synthesis of 21 scientific studies. They correspond to different case studies, interviews, surveys, and focus groups in which professionals belong to soft. dev. and telecommunications companies (e.g, IBM, NOKIA, ABB, and Ericsson) participated, some with distributed teams, and the research was Made in Brazil, Uruguay, Sri Lanka, Mauritius, Finland, Sweden, USA, India, Norway, Pakistan, China, and Russia.

KM IN SOFT. DEV. ORGANIZATIONS

Software development organizations may have various problems in KM. Currently, software development projects handle large volumes of information and are integrated by professionals from different areas with different knowledge and skills. Although knowledge is a success factor in software development, many times development methodologies, e.g., agile ones, do not have activities or mechanisms for their effective management.

Recommendations to improve KM found in the literature can be classified into 4 dimensions.



DEFINITION

Certain definitions can help knowledge management in an organization. It should be taken by the senior or middle management.

- **KM strategy:** a defined strategy with objectives allows to align the rest of the KM activities. It is recommended that the strategy seeks to centralize knowledge, and facilitate its access and search.
- **Simple & clear processes:** that establish the objectives of KM, what kind of knowledge is valuable, how it will be stored and how those involved should proceed when interacting with knowledge.
- **Defining responsibilities:** clearly establish who are the referents in each area of knowledge.
- **Metrics:** define and use metrics on the creation, use, and usefulness of knowledge. For example, using a score system, recording the number of accesses or allowing those involved to qualify, respectively.

COMMUNICATION

In order to facilitate the stakeholders' alignment with the established definitions, effective communication is recommended.

- **Communicate strategy:** management must be transparent with its objectives so that everyone involved understands the value of KM to the business.
- **Emphasize benefits:** communicate openly the benefits and profits that are expected and obtained from KM
- **KM Training:** It allows all stakeholders (especially the new ones) to understand how to better perform KM activities.

VALUES

Fostering and developing certain values makes it easier to achieve better KM.

- **Commitment:** A high degree of commitment can be achieved by showing the usefulness of knowledge and using simple KM processes.
- **Communication:** foster communication and cooperation within the organization, in order to increase the trust of stakeholders and promote teamwork.
- **Cooperation:** To facilitate the dissemination of knowledge throughout the organization, it is recommended to generate teams that manage that knowledge with members of various areas of the organization. Highlighting the importance of teamwork and organizational learning is recommended.
- **Trust:** to increase trust, face-to-face project initiation meetings can be organized to help the team to know each other and build trust.
- **Horizontality:** The management of KM should not be vertical but in all directions.
- **Motivation:** Favor the motivation of workers to get involved with KM. A motivated employee will make more and better contributions than one who is not, so the motivational factor is one of the keys to a successful business KM.

RESOURCES

To implement the above dimensions there are several alternatives.

- **Tools:** good criteria for choosing tools includes: an easy-to-use interface, robustness, ease of access, and powerful search functionalities. In addition, the organization must protect and promote its use.
- **Social software:** its use allows to share knowledge in a simple way and encourages socialization among stakeholders. Examples are: wikis, mailing lists, project tracking tools, intranets, and blogs.
- **Categories:** having categorized knowledge facilitates and encourages its access.
- **Gamification:** Using recreational dynamics encourages participation in KM. Examples of this are: having a scoring system, medals, leaderboards, etc.
- **Training:** the training carried out by the organization facilitates knowledge generation because it allows knowledge consolidation by stakeholders and encourages the exchange between them.
- **Rotation of stakeholders:** between different projects or areas it allows to disseminate and generalize knowledge.
- **Rewards:** for stakeholders who perform tasks related to KM. This can be implemented with an increase in their remuneration or by taking into account their contributions to the KM in their performance evaluation.
- **Record of lessons learned:** the use a record of lessons learned, where stakeholders can access to consult about previous experiences. The lessons can be registered during the project, not only after completion.

¿Who is this briefing for?
SE professionals who want to make decisions about KM based on scientific evidence.

¿Where do the discoveries come from?
All the findings of this report were taken from a rapid review.

¿What is included in this briefing?
Recommendations to improve KM with experimental validation in companies or soft. development teams.

¿What is not included in this briefing?
Other information or guarantee of the results of applying the recommendations for non-conformity of the context or variants of its application.

Fig. 1 Evidence briefing with RR results (adapted for inclusion here).

Workshop Activities
The workshop had the following characteristics. - A copy of the evidence briefing (see Figure 1) was given to participants with a reasonable time to read it. - The reviewers present the main concepts of EBSE, listed the characteristics of the RR (mentioning the problem and the methodology used), and finally presented the recommendations obtained. - The reviewers led a practical exercise to promote the analysis of the recommendations. The exercise had the following steps: 1. All the separately printed recommendations were placed on the table. As the only exception to this, all recommendations from the 'Values' category were grouped into one. 2. A quadrant was drawn, with the X-axis representing the ease of implementing a recommendation and the Y-axis its potential benefit. 3. Each participant had a turn (in a pre-established order) to grab a recommendation from the table and put it in a place in the quadrant or to rearrange a recommendation that was already located in the quadrant. In any case, the participant had to explain their action. 4. The previous step was continued until all the recommendations were located in the quadrant and a general agreement had been reached. 5. Finally, the team was asked to choose recommendations that were suitable for implementation.

Table 7 Workshop for the dissemination of results.

the highest priority, and they commented that not currently having KM objectives did not allow them to focus well or obtain good results. They associated the definition of strategy with other recommendations that should also be implemented jointly, these were to communicate the strategy, define categories of knowledge, use social software, and define those responsible (although this was previously out of their initial selection, see the rectangle in the upper right part of the figure).

At the time of the workshop, the company had been recently acquired by a larger company and the participants had started using Atlassian Confluence[4], a KM tool already used by the acquiring company. We believe that the participants associated the recommendation of social software with high benefits and ease of implementation due to this current situation, and because they were beginning to perceive the benefits of this particular tool.

6.9 Threats to RR Validity

The main threats are: (i) we used a single search engine for scientific articles, (ii) although we defined and tested criteria on a sample of studies, the reviewers performed the selection, extraction, and synthesis on a disjoint subset of studies, (iii) we did not perform a formal quality assessment of the studies which could limit

[4] Confluence is a web-based corporate wiki and collaboration tool developed by the Australian company Atlassian. According to its creators, it can be used by multiple teams and organizations to generate and consume content (in order to better manage their knowledge) and allow for better collaboration. It can also be integrated with recognized productivity and tracking tools (e.g., GitHub, Jira). Their website is https://www.atlassian.com/software/confluence.

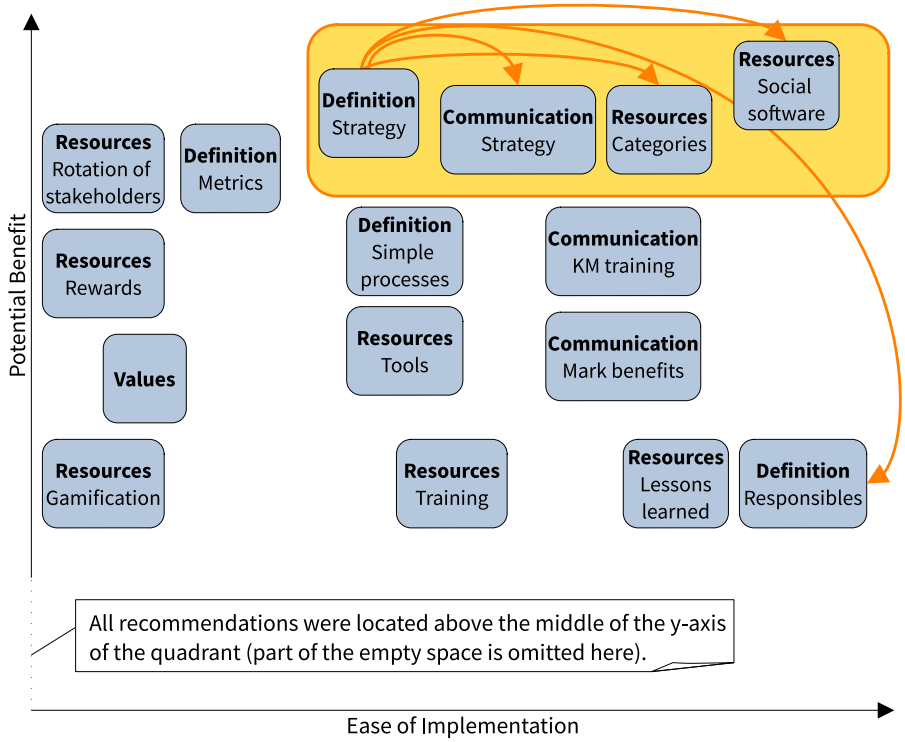


Fig. 2 Result of the workshop exercise: Recommendations placed by the soft. dev. team according to their potential benefit and ease of implementation.

the reliability of the evidence, and (iv) although closely trained and supervised, the reviewers were inexperienced in conducting reviews.

7 Results

In this section, we present the answers to the research questions of our replication.

7.1 RQ1: Perceptions and attitudes about using RR information

Perceptions and attitudes about the RR results were obtained from analyzing the audio recordings of the knowledge transfer workshop and the responses to the questionnaire we circulated that day (which we also sent later to the architect), as well as the responses to the follow-up questionnaire that we conducted eight months later with the requesters. All these activities were conducted in Spanish, thus we present translations of quotes that were reviewed by a translator to ensure they remain faithful to the originals.

We compared our findings with those of Cartaxo et al. (2018), who reported that they were evaluating of the use of RRs as a practice support method in SE, by identifying issues that supported the requesters identify practice changes that could address their specific KM problems and issues that hindered the adoption of the process changes. We present our findings in the same sequence as they were

obtained and relate them, where possible, to concepts already proposed by the aforementioned study (indicated in bold).

7.1.1 Early Attitudes

During and after the presentation of the recommendations, some attendees made comments or posed questions about them. The results of the analysis of these interventions seem to support the following previous results on RRs:

Offer reliable content. Some attendees considered that the information was more reliable than others that they used frequently. For example, the project manager said:

> *- Recommendations are good, they are clearly understood and based on scientific articles, which were approved, reviewed, and followed their process. That gives them veracity, another weight, has another value. This has to be a guide for us.*

Applicable to SE practice. All agreed that the recommendations were applicable and seemed to help solve their problems. This was also expressed in the practical exercise since all the recommendations were located above the middle of the y-axis, that is, all of them were perceived as having a medium or high potential benefit if they were applied.

Fostered the learning of new concepts. During the exercise, some attendees asked questions seeking to know about some of the concepts mentioned in the recommendations. These questions were answered by reviewers or other attendees, suggesting that all participants had learned new concepts related to the problems the company was facing.

7.1.2 First Questionnaire

Table 8 shows the answers to the questionnaire the five company attendees completed at the end of the workshop. Some clarifications are necessary. First, although all participants responded to the survey, they answered only the closed questions and did not include any additional comments. Second, regarding E3, only the requesters who participated in the project agreed with us on what we were going to consider as a similar context and were also aware of the literature that we were finding and its characteristics. Furthermore, the focus of the requesters was to find some strategies to address their problems. We could only expect them to identify approaches they felt were plausible solutions (or solution approaches) and practical in their own circumstances.

Even considering this, we believe that the responses confirmed the adequacy of the definition of the problem we addressed, the research question, and the studies selected (based on what we explained to the workshop attendees). They also confirmed the following results on the RRs: **Problem-oriented** (i.e., the RR provides evidence to address the problem they are facing), **Improve problem comprehension** and the need to **Discuss the findings of the RR** (i.e., use face-to-face transfer activities and not just to present results in reports).

Question ID	Totally agree	Somewhat agree	Neutral / Disagree
E1	5	0	0
E2	3	2	0
E3	1	4	0
E4	3	2	0
E5	1	4	0
E6	3	2	0

E1: The defined problem corresponds to a real problem that we have in knowledge management (KM).
E2: The question that guided the investigation is suitable to help us address the problems in KM
E3: The studies found in the literature review seem adequate, that is they have contexts and problems similar to ours.
E4: The recommendations that were found to improve KM seem useful to help solve our problems.
E5: The project helped me to understand problems from another perspective or understand approaches to their treatment.
E6: Today's meeting to present the results of the review was very good.

Table 8 Opinions of the participants of the dissemination workshop.

7.1.3 Follow-up Questionnaire

Eight months after the workshop, we tried to find out if the recommendations obtained through the RR had been put into practice and if they had helped to solve the problems. Below are the findings obtained by analyzing the responses of both requesters. Two of the findings (indicated with *) do not correspond to direct comparisons with the work of Cartaxo et al. (2018) but they are relevant.

***Company situation.** One requester gave us more explanations about the situation of the company. Certainly, this information should be considered when interpreting the results.

> *- Unfortunately, while this investigation was being carried out and even after it ended, the company underwent many very important changes. It was bought by another company, several important roles were changed or eliminated, the hierarchy of the organization changed, the priorities changed, and finally, the company closed.*

Applicable to SE practice & **Problem-oriented**. One of the requesters considered that the results seemed adequate and evidenced a mature understanding of KM.

> *- I think KM is a complex issue. But looking at the recommendations and comparing them with companies I have previously worked for, those that were good at KM and those that were not. I believe that this research has really managed to capture and show the critical parts of the problem and presents a very interesting approach. It's easy to understand and, at the same time open to different specific solutions that the organization and/or team may want to adopt.*

Offer reliable Content & **Novel approach to support decision-making**. Both requesters highlighted the value of scientific evidence and its differences from other types of information. Their comments also support the view that the evidence provided by RRs provides a better method for supporting decisions.

> *- I think that lately, we've become used to solving problems with the first two Google results, with the first two most popular papers that we find, we simply want to apply something that Netflix did, or we want to use the same thing as Facebook. Also, the industry almost never allows us to really do an exhaustive search on how to fix a problem. The evidence-based approach is absolutely the opposite and its conclusions are indisputable. I believe that the cost/benefit of applying it depends only on the situation of the company.*

***Results in a timely manner.** We received a negative opinion about our RR conduct. One requester considered that shorter deadlines would have allowed for better use of the results.

> *- Although the project had a certain dynamic proposed by you, I think that given the particular situation of the pandemic that we faced and hence the definitive closure of the company... I have the feeling that if we had shorter execution times we could all have had a different view of the results. Currently, any development team using agile methodologies runs work iterations of one or two weeks maximum, with which it should be possible to run short experiments and analyze their results in less time than we currently have.*

We had not agreed on a timeline for the RR with the requesters nor discussed whether we needed to consider any deadlines for delivering our recommendations. However, their comment makes it clear that short timeframes are important in industry collaboration. In addition, this comment supports Cartaxo et al. (2018)'s view that reducing timescales, and not only reducing effort (an issue that we did prioritize given the reviewers' restrictions), is an important benefit of using RRs (*Reduce time and cost to conduct decision-making process*.

Use of the results. The requesters confirmed that they appreciated the results and had started to use part of them. In particular, they indicate having (1) started the definition of processes and responsibles for the areas of QA and DevOps, (2) encouraged certain values that they already had, (3) decided to stop using Google Drive, (4) begun to use better knowledge categorizations, (5) considered using social software. Unfortunately, the major changes the company underwent prevented it from continuing in that direction. Even so, the following comments show that the RR results helped to make some of those changes less traumatic.

> *- [...] the acquisition of our company by a larger one forced our team to use Atlassian Confluence as the main tool for KM. [...] Given this scenario and the opportunities for improvement that our team had detected during the workshop, the use of Confluence was naturally adopted, thus covering several of the points mentioned in the evidence briefing (e.g., the definition of simple and clear pro-*

cesses, the definition of metrics and responsibles, the values of cooperation and horizontality, and the resources of tools and record of lessons learned)

7.2 RQ2: Using RR outcomes without experience in SE research

During the intermediate validations, practitioners noted that while the identified recommendations appeared beneficial, they did not know how to implement some of them. In response, we chose to refine the recommendations to enhance clarity. Moreover, during the workshop, we shared examples with practitioners that illustrated how some of these recommendations could be practically implemented. Given the current state of evidence provided by SE scientific publications, it appears that understanding the reported recommendations and possessing the knowledge required for their implementation may pose challenges for practitioners without academic experience in SE.

We also implemented some other actions to facilitate practitioners' participation in the RR process and ensure they understood the results appropriately. These measures included:

- Delivering a concise introduction to research and scientific articles, focusing on empirical SE and EBSE during the initial meeting with the requesters and at the beginning of the dissemination workshop.
- Providing condensed information at all times to eliminate the necessity for direct handling of scientific articles by practitioners. For instance, during the intermediate validation, we supplied a summary of the context of selected papers and their recommendations.

8 Discussion of Findings

It is necessary to discuss various aspects of our study. This section presents broader reflections on the answers to the research questions (Section 8.1). Additionally, we provide a comparison of the results of our replication with those of the original study conducted by Cartaxo et al. (2018) (Section 8.2). Furthermore, we delve into the significant challenges encountered by the review team and outline the strategies employed to mitigate them (Section 8.3). Subsequently, we present certain issues related to the RR process (Section 8.4). It is noteworthy that many insights presented in this section, in particular in the latter two points, stem from the retrospective meeting conducted by RR review team. In all mentioned subsections, we have included recommendations for the use of RR in SE to complement our reflections (see light blue boxes). Finally, we discuss on the weaknesses of our work (Section 8.5).

8.1 Attitudes about the use of RR results

Both our review requesters and other team members highlighted that RRs offer reliable content and that this differentiates them from other types of information

sources that they commonly used. This is also consistent with the results of our recent study on attitudes towards EBSE and SRs at a government agency (Pizard et al., 2023).

Requesters also agreed that the recommendations provided by the RR were applicable to their problem and that putting them into practice would bring benefits. Finally, collaborating in an RR (or at least participating in the activities of dissemination of its results), helped the team improve their understanding of the problem and learn new concepts.

Despite not stipulating it beforehand, one of the requesters would have preferred a shorter timeline. His argument was that with results in less time, they could have applied more recommendations before the closure the company. However, we were in a low-resource setting, Lezama and García could only dedicate a limited effort per week. For that reason, and also because we did not understand that it was a priority to obtain results in a short time, our RR was conducted in three months. The current effort used would have been mapped to about two weeks if reviewers worked full-time. Therefore, two aspects are important to highlight: that practitioners do appreciate short processes to obtain results and that all the requesters' needs are important, not just those about the problem to be addressed.

Recommendation 1. Consider adequately the needs of RR requesters and the context in which the evidence will be used.

8.2 Comparison with Cartaxo et al. (2018) study results

The findings from our study about the perception of the use of RRs align closely with those reported by Cartaxo et al. (2018) (refer to Table 9). Notably, despite differences in the educational backgrounds of our requesters, our results remain consistent with the original study, whose participants held master's degrees and were employed in an applied research institute. In contrast, our requesters lacked academic education, had no prior experience in SE research, and were employed in a software development company. This contrast underscores that the value of RR is not confined to practitioners with a background in research, highlighting its adequacy as a valuable tool for enhancing collaboration between academia and industry.

Our findings seem to indicate that collaborating in the conduct of an RR, or at least participating in the dissemination activities, allows practitioners to approach scientific evidence with a positive attitude, even when they are not used to using it or lack an academic educational level. When we undertook the study, there were no other studies than Cartaxo et al. (2018)'s that described the use of RRs to support practitioners. More recently papers by Song et al. (2022) and Bjarnason et al. (2023) have confirmed the values of RRs in the context of R&D collaboration with industry experts, however, our paper is the only example of using RRs in collaboration with non-expert requesters.

Recommendation 2. Conducting collaborative RR is recommended as an effective method to introduce practitioners to scientific evidence.

Cartaxo et al. (2018) Results	Confirmation
Benefits of the RR	
Applicable to software engineering practice	Yes
Novel approach to support decision-making	Yes
Fostered the learning of new concepts	Yes
Offer reliable content	Yes
Problem-oriented	Yes
Improve problem comprehension	Yes
Increase team confidence	No
Reduce time and cost to conduct decision-making process	No, but of importance[1]
Fast and easy way to find information	Evidence briefing (EB) issue[2]
Avoid reading multiple sources	EB issue / No
Flexible knowledge transfer medium	EB issue
Non-applicable evidence can support other problems	No
Interest to receive briefings regularly	EB issue
Recommend Rapid Reviews to other practitioners	No
Improvements to the Rapid Review	
Discussing the findings of the RR is needed	Yes
Present the primary studies' context near their findings	EB issue
Avoid printing the RR report in black-and-white	EB issue
Graphical information is needed	EB issue

[1] Discussed and clearly an issue of importance to requesters, even though they didn't specify it upfront.
[2] *Evidence briefing issue*: As we stated above, our study focuses on issues related to RR conduct and not those about evidence briefings.

Table 9 Results of Cartaxo et al. (2018) confirmed by our study.

Based on the recommendation of Cartaxo et al. (2018) to incorporate discussions of the results as dissemination activities, we not only prepared an evidence briefing with the RR findings, but we also held a workshop in which the attendees began to discuss the evidence. This combination worked very well for the dissemination of our results. From the results of both studies, it seems clear that a single-page evidence briefing report is insufficient to ensure that the results can be used in practice. In our experience, it might be useful to include (1) information on how to put each recommendation into practice, (2) dependencies between recommendations, (3) indications of the strength of evidence supporting each recommendation. In our study, the requesters were (slightly) interested in knowing more about point 3, and we discussed this with them. We suggested ideas or examples for points 1 and 2 during the workshop, and the development team discussed them by choosing suitable alternatives for their context.

Recommendation 3. To ensure that evidence can be effectively used in practice, it is essential to plan and implement appropriate dissemination activities with practitioners. Relying solely on a single evidence briefing appears to be insufficient.

8.3 Major Challenges and Mitigation Strategies

The biggest challenges encountered by the review team were: (i) the lack of guidelines and examples of using an RR in the software industry and (ii) the difficulty in finding adequate evidence. The implications of these issues are discussed below.

At the time of our study, Cartaxo et al. (2018)'s study was the only example of the application of RR to support SE practice. Since we had no experience in conducting this type of secondary study, we considered that there was a high risk of not being able to produce recommendations based on scientific evidence that would effectively help requesters to address their KM issues. To mitigate this, we worked cautiously at each stage, e.g., verifying each step of the process and its results. For example, to put together the research question and the protocol, we carried out several preliminary literature searches and verified that some of the retrieved studies had adequate recommendations for our RR. We also carried out intermediate validations with the requesters to verify that the evidence we were finding was adequate. The second issue was detected precisely in those preliminary searches, in which very few studies had adequate evidence to address our problem. To mitigate this, and knowing that very few SE studies usually include recommendations for practice (Da Silva et al., 2011), we decided to also consider those that included lessons learned or experimentally validated observations.

The limitations of SE evidence, also noted by Rico et al. (2024), could impact the RRs as follows:

1. Lack of empirical studies in SE means we are forced to rely on the use of less reliable results such as opinions and lessons learned.
2. Lack of clear recommendations in SE papers means that the results of individual papers may need to be restructured or re-analysed in order to deliver well-specified process change recommendations.
3. Lack of common research approaches and standardized research questions means that results from different studies may require qualitative aggregation.

Point 1 reduces the strength of evidence. Points 2 and 3 are risky exercises for an RR that relies on a single researcher because there is a danger that personal biases and prejudices may adversely affect the recommendations.

First, the challenges we faced suggest that researchers need to lead RRs conduct, a view consistent with the studies of Cartaxo et al. (2018) and Rico et al. (2024).

Recommendation 4. Researchers need to take the lead in conducting RRs.

Second, both challenges and the strategies we adopted make us reflect on whether it would be appropriate to use a multi-stage strategy to conduct RRs in SE. In their recent work on selecting RR methods for complex questions in the health field, Wilson et al. (2021) proposed a two-stage process that consists of *(1) scoping the literature*, i.e., understanding the needs of the requesters and conduct preliminary searches to understand the available literature, and *(2) selection of an optimal approach*, i.e., further consultations are made to the requesters to tailor questions and identify relevant studies. This approach, the authors pointed out, makes it possible to consider the available literature together with the timelines required to choose an appropriate RR strategy.

One issue that affected our timescales was that we had to go back to earlier stages of the RR several times, e.g., when trying to perform the synthesis it was necessary to extract more information from the primary studies. In this sense, we agree with King et al. (2022) that *"the customized and iterative nature of rapid reviews means that some flexibility may be required"*. Although of course, as these authors also indicated, changes made beyond the protocol and the rationale for making them must be transparent and adequately reported. Moreover, the aim of RRs to be applicable in low-resource settings also implies that they can be suitable for participation by non-experts in their conduct.

Recommendation 5. RRs require some flexibility (e.g., a preliminary stage to evaluate the existing literature, or some form of iteration in the stages), but it is important to be transparent and detailed during the dissemination of their results.

Finally, as recommended by King et al. (2022), we engaged with the requesters early and throughout the review process. This was crucial in achieving results that met the requesters' needs. It involved detailed surveying information needs, validating problems and research questions, and confirming selection criteria with the requesters. We also conducted an intermediate validation where we not only validated the evidence collected up to that point but also aligned the recommendations with the client's expectations, which influenced, alongside available primary studies, the synthesis approach. Lezama's dual role brought several advantages, including a deeper understanding of the company context that could simplify certain stages of the process, particularly the evaluation of evidence relevance. His involvement with both the research team and the organization receiving the evidence appears to have facilitated the conduct of an RR and the application of its results, reflecting effects similar to those observed in our previous study on EBSE within government agencies Pizard et al. (2023).

Recommendation 6. Engage with the requesters in the early stages and throughout the review process to understand their needs and discuss decisions about the RR process.

8.4 Reflections on Rapid Review Nature and Process

Some issues about the RR nature and process deserve reflection.

8.4.1 Rationale for conducting RRs

As Ralph and Baltes (2022) point out, RRs should not be used to *"legitimize bad systematic reviews when there is no pressing need for immediate results"*, something our results suggest is happening in SE (refer to Section 3). In our opinion, timely collaboration with industry practitioners is a reasonable justification for conducting an RR. However, 13 out of the 23 papers we found did not appear to have direct industry collaborators. We disagree with Ralph and Baltes (2022)'s conclusion that fewer RRs are needed. If we limit the use of the term RRs to studies that involve collaboration with industry (which is consistent with Rico et al. (2020)'s guidelines) and ensure that any identified process changes are monitored by the review team, this may lead to more industry case studies reporting evaluations

of process changes introduced by scientific evidence. More industry case studies might improve the availability of primary studies for case surveys, which is critical in systematic reviews aiming to influence industry practice (Kitchenham et al., 2023a).

Recommendation 7. Limit the use of RRs to collaborations with a defined requester or knowledge user (i.e., to use the evidence in a particular situation) in low-resource settings, and ensure that any identified process changes are monitored by the review team.

Finally, with regards to the decision between conducting an RR or an SR, opting for a full SR should be considered when stronger evidence is necessary. However, this might be deemed inappropriate if the requesters consider it crucial to minimize timescales or are in a low-resource setting. In such cases, it would be prudent to assess which SR processes can be restricted (see Section 8.4.4). Finally, even in the context of collaboration with requesters or other stakeholders, it is possible to consider conducting an RR as a preliminary stage to a full systematic review.

8.4.2 Low-resource settings and use of existing secondary studies

Although existing literature about RRs in SE emphasizes that they mainly seek to address a requirement within a short timeframe (Cartaxo et al., 2018; Rico et al., 2024), it could be argued that its primary goal is to aggregate evidence in a low-resource setting. This encompasses scenarios such as our replication or one of the RRs reported by Rico et al. (2024), where the RRs were conducted over several months but with very limited dedicated effort throughout that period. However, the need for scientific evidence in a resource-constrained setting cannot be solely addressed by conducting a RR.

In our case, as our intention was to conduct a replication as close as possible to the original study by Cartaxo et al. (2018), we directly considered conducting a RR without explicitly searching for possible previous SRs conducted on the topic of interest. The secondary studies we found in our search did not directly answer the research questions but we planned to consider them if we did not find suitable primary studies to answer the research questions. However, after analyzing the results of our study and previous studies on RRs in SE, with the aim of minimizing effort and timescales, we suggest researcher utilize secondary studies if relevant ones are available, as we explain below.

From the viewpoint of practitioners, they have a question or problem they need to address. An RR can provide an answer in situations where low cost and speed are critical. However, if the initial search, finds a good quality, and relevant, SR or mapping study, that may be sufficient to address the problem directly or at least identify some relevant primary studies. It may also present a mature (i.e. well-understood, well-tested answer) which would be of particular value to small IT companies (although the results might need to be refined and interpreted to provide actionable process change recommendations). For R&D organizations who want "state of the art" results, updating a good quality SR might be quicker and might provide more valuable information than conducting a RR.

For inexperienced researchers it may be difficult to consider how to use an already published SR or mapping study. In our opinion this could be done in a variety of ways:

- To answer questions or problems directly if the SR answers the same questions and has adequate quality.
- To reduce the effort needed for the RR by basing the selection on the set (or a subset) of the primary studies of the previous SR. This was the method used in the RR conducted by Song et al. (2022).
- As a means of improving or validating the RR:
  - Improving the search process, since a previous SR can provide validated search strings and/or ideas for keywords.
  - Validating the search strategy, i.e., by checking whether the RR search process found all the relevant primary studies found by the previous SR.
  - Validating the results, i.e., assessing whether the previous SR raised any issue that the RR has not? Assessing whether results agree, and if investigating why they do not. This type of analysis presuppose that both secondary studies are addressing very similar research questions.
- Justifying the publication of the RR, i.e., identifying the additional results that the RR provides compared with the previous SR(s).
- In general only in the case of updating an existing SR, can the results from a new search process be fully integrated with the existing results. An example is the study by Da Silva et al. (2011).

Recommendation 8. Before deciding to conduct an RR, consider using an existing SR (or other secondary study) if available.

8.4.3 Quality Assessment

None of the RRs previously reported in the field of SE (refer to Section 3), nor Cartaxo et al. (2018)'s original paper, nor the recent guidelines by Rico et al. (2020), included quality assessment of primary studies. However, Cartaxo et al. (2019)'s study indicates that some researchers believe that RRs without quality evaluation are not useful. In our study, the requesters were interested in evidence from studies of companies similar to their own in terms of staffing levels and use of SE methods. More generally, we would expect practitioners to be interested in information that provides insights into (1) the methods and limitations of the studies, to better determine the reliability of the evidence and the extent of support for specific process changes, and (2) the applicability of the evidence to the specific context in which they are working. To address the latter, incorporating selection criteria to exclude studies that do not show applicability could be a more direct approach, however, the limitations of the evidence in SE research can often make that option infeasible (as happened in our case).

Recommendation 9. Do not rule out quality assessment stage prematurely. Requesters might be interested in the strength and applicability of evidence for recommendations, which requires evaluating methodological weaknesses of primary studies.

8.4.4 Process shortcuts, deviations and risks

Several researchers indicate that minimum standards are necessary for RRs (Cartaxo et al., 2019). We agree with this and believe that Rico et al. (2020)'s guidelines for conducting RRs collaboratively with practitioners are a good reference.

The specific RR process needs to be designed to meet the specific requirements of requesters. No step in the SR process can be omitted without risking the reliability of the RR conclusions. We agree with Rico et al. (2020) and King et al. (2022) that producing a protocol is important because it defines the roles and responsibilities of the review team and the company participants. However, regarding other SR processes, the review team and requesters need to discuss the risks associated with any proposed deviation from the standard SR process.

Recommendation 10. The specific RR process should be tailored to meet the specific requirements of requesters. Adhering to Rico et al. (2020)'s guidelines may be helpful. Maintaining a protocol is essential. Any shortcut or deviation in the SR process must be carefully studied and discussed with requesters, as they introduce risks to the reliability of the RR conclusions.

Our systematic review (refer to Section 3) suggests that the most common process changes involve omitting quality evaluation of the literature, searching a single digital library, or using a single researcher for one or more of the search, selection, and aggregation processes. In terms of risk minimization, searching a single source such as SCOPUS might often be the best option, as SCOPUS covers most of the important SE journals. Furthermore, it would be relatively easy for academics wishing to update the RR to broaden the search process. However, if requesters require the most up-to-date information, the review team may need to search for articles published in specialist workshops or archive sites, which may not be indexed by SCOPUS.

Recommendation 11. To minimize risks, opting to search a single source such as SCOPUS can be the best approach to streamline the SR process.

Finally, regarding rigor, we believe that for RRs, the issue is not completeness but reproducibility and transparency in reporting. An RR could serve as a starting point for further research if it addresses issues likely to affect other companies. If the RR proposes options to address a problem, the success or failure of the adopted process change could provide valuable insights for other companies. Moreover, an RR could serve as a valuable starting point for a full SR if its scope, limitations, and results are well-documented, potentially extending beyond initial constraints to provide comprehensive insights. Reporting on the process changes and their impacts would make for an interesting case study on specific process change recommendations.

Recommendation 12. When it comes to rigor in an RR, the primary concern is not completeness but rather ensuring reproducibility and transparency in its report, with consideration for the future use of its results.

8.4.5 Dissemination of RR results

The dissemination of the results of a RR is an essential stage of the process. Requesters do not necessarily need a full academic-style report, but they do need to understand the results and recommendations well enough to make informed decisions. Rico et al. (2020) suggests that the results of the review must be reported to the requesters, and that a separate activity should consider dissemination to other stakeholders and academics. We agree and note that both Cartaxo et al. (2018) and our studies found that a meeting with the requesters and other company staff was useful. Such a meeting allows company representatives to discuss the recommendations with the review team and identify the best options for any required process change. This suggests that dissemination efforts such as workshops or hands-on activities facilitate the reception and incorporation of evidence from RRs. Additionally, it would seem sensible for the review team to provide a preliminary report of the main conclusions and recommendations of the review prior to the meeting (though we consider an evidence briefing is insufficient).

Cartaxo et al. (2018) argue for the use of appealing media to present results to practitioners, this perspective is applicable mainly to RRs that are formally requested or conducted in close collaboration with the practitioners utilizing the results. In such cases, requesters may not necessarily need access to detailed information such as search methods, search terms, selection processes, etc. In these cases, where requesters actively participate in the entire process and are familiar with and have validated the decisions of the reviewing team, the additional information may not be essential.

However, unless this more detailed information about the SR process is reported somewhere, regardless of the importance of the results, their value to other researchers or practitioners will be limited. For example, researchers will not know which recommendations require further research, hindering the ability to update or extend the results systematically. Practitioners will not be aware of the extent to which RR results have been tailored to the specific context of the original requesters, potentially omitting important findings that might not be relevant in the requester's specific situation. An overemphasis on evidence briefings for dissemination, without a recommendation to report basic RR process information in ancillary materials, could reduce the scientific and practical value of RRs. Although it should be relatively simple to increase its value by including the additional information (or a reference to it) in any externally published RR report.

Despite a recent increase in the publication of RRs in SE, our systematic review indicates that these reports are often incomplete and lack the necessary information for a comprehensive understanding of the RR process or its results. To address this, it may be beneficial to consider the use of SEGRESS guidelines for reporting secondary studies (Kitchenham et al., 2023b) and the list of core reporting & dissemination principles for rapid review prepared by Kelly et al. (2022). Importantly, it should be noted that the expedited nature of RRs, where certain stages are omitted or shortened to provide quicker results, does not justify inadequate or incomplete reporting.

Recommendation 13. The dissemination of RR results should consider the target audiences. Activities that facilitate discussions on evidence and its practical application are most suitable for the RR requesters. Meanwhile, an academic report must be sufficiently comprehensive to enable appropriate use of the results by those not involved in the RR conduct.

8.5 Trustworthiness of this study

It is necessary to reflect on the weaknesses of our work. For this, we used the proposal made by Krefting (1991), which is based on Guba (1981)'s model of trustworthiness of qualitative research. We evaluated the following aspects:

— ***Credibility*** is a measure of whether the study's findings are correct and accurate. It relies on the credibility of the researchers themselves, as well as their research methods and the reflexivity used to evaluate their research.

From this perspective, we detected one issue related to the ability of the research team to conduct an RR. Specifically, although they were previously trained and worked closely with Pizard, for Lezama and Garcıa this was their first participation in conducting an RR or SR. To mitigate possible deviations from the RR methodology, we worked following the RR guidelines (including, among other things, developing a protocol prior to conducting), and we consulted doubts and decisions with Vallespir throughout the process. Pizard took the lead in the aggregation process, one of the most difficult RR stages for novices to understand (Pizard et al., 2021), by supervising the elaboration of the process recommendations from the primary studies.

In addition, we identified two issues that might have introduced bias into the study outcomes:

- Lezama worked in the company and the requesters and members of the company knew that the RR conduct was part of his capstone project. Thus, they could have been tempted to give positive comments. We sought to mitigate this by repeating many times to them that positive or negative results were equally important to our study and obtaining their perceptions of the RR results on different occasions and by different methods.
- Pizard sought to complete his research on the adoption of EBSE, which could be thought of as more valuable with positive results. To mitigate the risk of possible bias in this direction, Pizard kept a detailed journal of his decisions and actions, he also reported and consulted his decisions with Vallespir and Kitchenham at different times throughout the planning and the conduct of the study, and analysis and reporting of the results.

— ***Transferability*** refers to the degree to which the results of qualitative research are applicable in other contexts or settings. From a qualitative perspective, the context in which the results were obtained, along with any atypical factors, events, or behaviors of the participants and their possible influence on the transfer of results to other settings, must be adequately analyzed and reported.

Thus we attempted to assessed the limitations of various aspects of study that might impact transferability:

- We cannot adequately assess the efficacy of the recommendations for process change identified by the RR because the situation in the company did not allow us to follow up how well the process changes introduced by the requesters worked out. This is a major limitation of our study and it limits the overall research value of our study. It should also be noted that we did not plan in advance how to assess the impact of any process changes which is an essential issue for researchers and practitioners who would like to know how effective specific recommendations were at addressing the specified company problems. The only approximation to an evaluation of the results was the follow-up questionnaire conducted a few months after the closure of the company with the requesters. During the time the requesters indicated using the results of the RR, they also reported using the Confluence tool for KM activities (as requested by the firm that had acquired the company), which may also have influenced their opinions about the results of the RR.
- With respect to collaboration between academics and small companies staffed by practitioners without SE research experiences, our results are encouraging, but it is possible that Lezma's personal relationship with the company had a significant contribution both to the collaboration process and to the attitude of the company staff to the RR method and its recommendations.
- With respect to the use of RRs to support industry-academia collaboration, our results generally support Cartaxo et al. (2018)'s conclusion that RRs can support practitioners to select process changes to address SE process problems. Our results are also consistent with the experiences of other researchers that have used RRs in such collaborations, see (Rico et al., 2024). However, in the context of collaboration with companies with little research experience, we believe it is important that the researchers take care to present the RR recommendation carefully, and provide explanation of how the recommendation can lead to specific process changes. Also, the company's positive attitude to the RR results may have been partly due to the fact that they were accustomed to embedding process change in their two-week sprint cycle and asked us to deliver recommendations that suited that process.

— ***Confirmability*** of qualitative data is assured when data are checked and rechecked throughout data collection and analysis in order to ensure results are likely to be auditable by others. This can be documented by a clear description of the analysis process, including approach strategies, roles of researchers, coding schemes that were applied, and strategies for verifying results. We have reported our data collection, analysis and synthesis processes including our internal validation processes with the goal of being as transparent as possible. Of particular note are the following validation processes:

- Validating the proposed information sources with the requesters during the RR process.
- Obtaining feedback on the RR conduct from Fernando Acerenza (a researcher with some EBSE knowledge and experience).
- Assessing the attitudes of company staff to the RR recommendations both by analysis of the comments made in the dissemination meeting and the post-meeting questionnaire.
- Vallespir and Kitchenham providing a critique of Pizard's thematic analysis process and results.

In addition, to provide readers with additional confirmation of the qualitative results mentioned in this report, as suggested by Braun and Clarke (2006), we provide specific quotes from our meetings with the company staff to support our analysis and interpretations of their comments.

However, the uses that the two requesters claim to have made of the results were obtained solely from the follow-up questionnaire carried out a few months after the closure of the company.

— ***Dependability*** is important to trustworthiness because it establishes the research study's findings as consistent and auditable. Thus, dependability has some overlap with confirmability. However, dependability also requires that researchers should aim to verify that their findings are consistent with the raw data they collected.

Our validation procedures were intended to ensure that our findings remained consistent with our raw data. In addition, we report specific quotes from our meetings with the company staff to confirm consistency between our findings and the raw data.

9 Concluding Remarks

Our study provides support for the value of rapid reviews in the software industry. Specifically, our results confirm that the benefits of RRs are not limited to practitioners with research expertise. Most of the practitioners' perceptions about RR outcomes were positive and strongly consistent with previous research (Cartaxo et al., 2018). All practitioners agreed that the RR results were more reliable than the information they usually used. They also highlighted the recommendations seemed useful and beneficial to address the problem they were facing, and, some months later, they reported having used some of the recommendations. In our application, we prioritized reducing the effort although due to the availability of the reviewers the RR took three months. One of the requesters has preferred shorter time scales, which confirms that practitioners appreciate results in a short time and that it is also necessary to carefully prioritize all requesters' needs.

The major novelties of our study are that it:

- Assesses the previous research on RRs in SE. This includes an investigation into the scope of RR adoption, main shortcuts and deviations used in the review process, and the reporting of RRs. Furthermore, it includes a detailed analysis of studies contributing to understanding the value of RRs in our field.
- Evaluates practitioners' perceptions on using RRs to support software engineering practice in the context of a software company. Unlike previous studies, the practitioners who participated in our research had no prior experience in SE research and they were not experts on the topic of RR.
- Provides evidence that RRs can support the SE industry practice. Specifically, it validates various benefits outlined by Cartaxo et al. (2018) during the initial assessment of RRs in SE, and confirms the need to translate broad RR recommendations into actionable process changes and the importance of ensuring practitioners comprehend the results, e.g., through face-to-face activities like dissemination workshops.

Our main recommendations on the use of RR in SE are:

- Following practices in healthcare Kelly et al. (2022), recommendations from Ralph and Baltes (2022), and guidelines by Rico et al. (2020), the term Rapid Review should be restricted to cases where researchers have a defined requester/knowledge user (or are themselves the knowledge user in a specific situation) and face requirements with short timescales or are in low-resource settings.
- SE researchers should adhere to Rico et al. (2020) guidelines, but should not arbitrarily dismiss the idea of quality assessment of primary studies.
- SE researchers should remain flexible regarding specific methods to streamline the SR process. In particular, they should utilize existing SRs wherever possible.

10 Declarations

- Funding: Pizard is pursuing his doctorate as part of the Basic Sciences Program of Uruguay (PEDECIBA). Pizard and Vallespir 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 known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- Availability of data and material: An extended report of the systematic review can be found as supplementary material. Data sets generated during the current study (without sensitive information from organizations or participants) are available from the corresponding author on reasonable request.
- 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.

Acknowledgments

We thank all the members of the company with whom we worked (whose names we keep anonymous for confidentiality reasons) and the middle and senior management who allowed the conduct of our study. We also want to thank Fernando Acerenza for his validation of our progress during the RR conduct and Natalia Milan for her translation work. We thank the anonymous reviewers for their constructive comments on the initial draft of this paper.

References

Abdelfattah AS, Cerny T (2023) Roadmap to Reasoning in Microservice Systems: A Rapid Review. Applied Sciences 13(3)

Baldassarre MT, Caivano D, Dimauro G, Romano S, Scanniello G (2021) On Internet-of-things Devices in Ambient Assisted Living Solutions. International Conference on Information Systems Development

Baldassarre MT, Caivano D, Fernandez Nieto B, Gigante D, Ragone A (2023) The Social Impact of Generative AI: An Analysis on ChatGPT. GoodIT '23, p 363–373

Barletta VS, Caivano D, Gigante D, Ragone A (2023) A Rapid Review of Responsible AI Frameworks: How to Guide the Development of Ethical AI. In: International Conference on Evaluation and Assessment in Software Engineering, EASE '23, p 358–367

Birk A, Dingsøyr T (2005) Trends in learning software organizations: Current needs and future solutions. In: Althoff KD, Dengel A, Bergmann R, Nick M, Roth-Berghofer T (eds) Third Biennial Conference, WM 2005 - Professional Knowledge Management, pp 70–75

Bjarnason E, Åberg P, bin Ali N (2023) Software selection in large-scale software engineering: A model and criteria based on interactive rapid reviews. Empirical Software Engineering 28(2):51

Braun V, Clarke V (2006) Using thematic analysis in psychology. Qualitative Research in Psychology 3(2):77–101

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 (ESEM), 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 (EASE), pp 24–34

Cartaxo B, Pinto G, Fonseca B, Ribeiro M, Pinheiro P, Baldassarre MT, Soares S (2019) Software Engineering Research Community Viewpoints on Rapid Reviews. In: International Symposium on Empirical Software Engineering and Measurement (ESEM), pp 1–12

Cartaxo B, Pinto G, Soares S (2020) Contemporary Empirical Methods in Software Engineering, chap Rapid Reviews in Software Engineering, pp 357–384

Carver JC (2010) Towards reporting guidelines for experimental replications: A proposal. In: 1st international workshop on replication in empirical software engineering, vol 1, pp 1–4

Chen H, Baptista Nunes M, Ragsdell G, An X (2018) Extrinsic and intrinsic motivation for experience grounded tacit knowledge sharing in Chinese software organisations. Journal of Knowledge Management 22(2):478–498

Cruz M, Bernárdez B, Durán A, Galindo JA, Ruiz-Cortés A (2020) Replication of Studies in Empirical Software Engineering: A Systematic Mapping Study, From 2013 to 2018. IEEE Access 8:26773–26791

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

Da Silva FQ, Suassuna M, França ACC, Grubb AM, Gouveia TB, Monteiro CV, dos Santos IE (2014) Replication of empirical studies in software engineering

research: a systematic mapping study. Empirical Software Engineering 19:501–557

Davison R, Martinsons MG, Kock N (2004) Principles of canonical action research. Information Systems Journal 14(1):65–86

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

Dennis AR, Valacich JS (2015) A replication manifesto. AIS Transactions on Replication Research 1(1):1

Denzin N, Lincoln Y (2017) The SAGE Handbook of Qualitative Research, 5th edn. SAGE Publishing, United States

Escalante MV, Fagúndez de los Reyes OG (2022) La Industria del Sofware: de Uruguay al mundo. PhD thesis, Universidad de la República (Uruguay), URL https://hdl.handle.net/20.500.12008/36229

Fritzsch J, Bogner J, Haug M, Franco da Silva AC, Rubner C, Saft M, Sauer H, Wagner S (2022) Adopting microservices and DevOps in the cyber-physical systems domain: A rapid review and case study. Software: Practice and Experience

Fritzsch J, Correia F, Bogner J, Wagner S (2023) Tools for Refactoring to Microservices: A Preliminary Usability Report. In: International Conference on Microservices, pp 1–6

Furukawa CA, Soares S, Cagnin MI, Paiva B (2022) Support for Accessible Software Coding : Results of a Rapid Literature Review. CLEI electronic journal, 25(3):1–13

Garritty C, Gartlehner G, Nussbaumer-Streit B, King VJ, Hamel C, Kamel C, Affengruber L, Stevens A (2021) Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews. Journal of Clinical Epidemiology 130:13–22

Garritty C, Tricco AC, Smith M, Pollock D, Kamel C, King VJ (2024) Rapid Reviews Methods Series: Involving patient and public partners, healthcare providers and policymakers as knowledge users. BMJ Evidence-Based Medicine 29(1):55–61

Gervigny MLI, Nagowah SD (2017) Knowledge sharing for agile distributed teams: A case study of Mauritius. In: International Conference on Infocom Technologies and Unmanned Systems - Trends and Future Directions (ICTUS), pp 413–419

Guba EG (1981) Criteria for assessing the trustworthiness of naturalistic inquiries. Educational Communication and Technology 29(2):75–91

Hamel C, Michaud A, Thuku M, Skidmore B, Stevens A, Nussbaumer-Streit B, Garritty C (2021) Defining rapid reviews: a systematic scoping review and thematic analysis of definitions and defining characteristics of rapid reviews. Journal of Clinical Epidemiology 129:74–85

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 (EASE), Association for Computing Machinery, pp 31:1–31:10

Hecker J, Muhr T (2022) ATLAS.ti Scientific Software Development GmbH. URL https://atlasti.com

Heredia A, Garcia-Guzman J, Amescua-Seco A, Serrano A (2014) Study of Factors Influencing the Adoption of Agile Processes When Using Wikis. International Journal of Software Engineering and Knowledge Engineering 24(06):859–885

Heredia A, Colomo-Palacios R, Soto-Acosta P (2017) Tool-supported continuous business process innovation: a case study in globally distributed software teams. European Journal of International Management 11(4):388–406

Hidalgo M, Astudillo H, Castro LM (2024) Challenges to use Role Playing in Software Engineering Education: A Rapid Review. In: Applied Informatics, pp 245–260

Humayun M, Gang C, Masood I (2013) An Empirical Study on Investigating the Role of KMS in Promoting Trust within GSD Teams. In: International Conference on Evaluation and Assessment in Software Engineering (EASE), p 207–211

Jurado JL, Fernandez A, Collazos CA (2015) Applying Gamification in the Context of Knowledge Management. In: International Conference on Knowledge Technologies and Data-Driven Business

Kammani A, Aljahdali S, Date H (2013) KM Capability for Software Development: A Case Study of the Indian Software Firms. International Journal of Business Information Systems 12(1):44–67

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

Kelly SE, McGowan J, Barnhardt K, Straus SE (2022) Paper 4: a review of reporting and disseminating approaches for rapid reviews in health policy and systems research. Systematic Reviews 11(1):1–10

Khan H, Ahmad A, Alnuem M (2012) Knowledge Management: A Solution to Requirements Understanding in Global Software Engineering. Research Journal of Applied Sciences, Engineering, and Technology 4:2087–2099

King VJ, Stevens A, Nussbaumer-Streit B, Kamel C, Garritty C (2022) Paper 2: Performing rapid reviews. Systematic Reviews 11(1):1–10

Kitchenham B (2008) The role of replications in empirical software engineering—a word of warning. Empirical Software Engineering 13:219–221

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, London

Kitchenham B, Madeyski L, Budgen D (2023a) How should software engineering secondary studies include grey material? IEEE Transactions on Software Engineering 49(2):872–882

Kitchenham B, Madeyski L, Budgen D (2023b) SEGRESS: Software Engineering Guidelines for REporting Secondary Studies. IEEE Transactions on Software Engineering 49(3):1273–1298

Krefting L (1991) Rigor in qualitative research: the assessment of trustworthiness. The American journal of occupational therapy : official publication of the American Occupational Therapy Association 45(3):214–222

Loli S, Teixeira L, Cartaxo B (2020) A Catalog of Object-Relational Mapping Code Smells for Java. SBES '20, p 82–91

Lonetti F, Bertolino A, Di Giandomenico F (2023) Model-based security testing in IoT systems: A Rapid Review. Information and Software Technology 164:107326

Marchetto A (2023) A rapid review on fuzz security testing for software protocol implementations. In: Bonfanti S, Gargantini A, Salvaneschi P (eds) Testing Software and Systems, Springer Nature Switzerland, Cham, pp 3–20

Matalonga S, Amalfitano D, Doreste A, Fasolino AR, Travassos GH (2022) Alternatives for testing of context-aware software systems in non-academic settings: results from a Rapid Review. Information and Software Technology 149:106937

Matturro G, Silva A (2010) ReBEC: A Method for Capturing Experience during Software Development Projects. In: Cimiano P, Pinto HS (eds) International Conference on Knowledge Engineering and Knowledge Management - Knowledge Engineering and Management by the Masses, pp 524–533

Mijumbi-Deve RM, Kawooya I, Kayongo E, Izizinga R, Mamuye H, Amog K, Langlois EV (2022) Paper 1: Demand-driven rapid reviews for health policy and systems decision-making: lessons from Lebanon, Ethiopia, and South Africa on researchers and policymakers' experiences. Systematic Reviews 11(1):1–14

Milovanović M, Minović M, Štavljanin V, Savković M, Starčević D (2012) Wiki as a corporate learning tool: case study for software development company. Behaviour & Information Technology 31(8):767–777

Motta RC, de Oliveira KM, Travassos GH (2023) An evidence-based roadmap for IoT software systems engineering. Journal of Systems and Software 201:111680

O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA (2014) Standards for reporting qualitative research: a synthesis of recommendations. Academic Medicine 89(9):1245–1251

Paes VdC, Pessoa CHM, Pagliusi RP, Barbosa CE, Argôlo M, de Lima YO, Salazar H, Lyra A, de Souza JM (2023) Analyzing the Challenges for Future Smart and Sustainable Cities. Sustainability 15(10)

Pizard S, Acerenza F, Otegui X, Moreno S, Vallespir D, Kitchenham B (2021) Training students in evidence-based software engineering and systematic reviews: a systematic review and empirical study. Empirical Software Engineering 26(3):50

Pizard S, Acerenza F, Vallespir D, Kitchenham B (2022) Materials for an Evidence-based software engineering & Systematic reviews course. DOI 10.5281/zenodo.7464649, URL `https://doi.org/10.5281/zenodo.7464649`

Pizard S, Acerenza F, Vallespir D, Kitchenham B (2023) Assessing attitudes towards evidence-based software engineering in a government agency. Information and Software Technology p 107101

Ponce F, Márquez G, Astudillo H (2019) Migrating from monolithic architecture to microservices: A Rapid Review. In: 2019 38th International Conference of the Chilean Computer Science Society (SCCC), pp 1–7

Pulkkinen M, Mazhelis O, Marttiin P, Meriluoto J (2007) Support for knowledge and innovations in software development: Community within Company: Inner Source Environment. In: International Conference on Web Information Systems and Technologies, pp 141–150

Păvăloaia VD, Necula SC (2023) Artificial Intelligence as a Disruptive Technology: A Systematic Literature Review. Electronics 12(5)

Radu LD (2020) Disruptive Technologies in Smart Cities: A Survey on Current Trends and Challenges. Smart Cities 3(3):1022–1038

Ralph P, Baltes S (2022) Paving the way for mature secondary research: the seven types of literature review. In: ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering, p 1632–1636

Ralph P, bin Ali N, Baltes S, Bianculli D, Diaz J, Dittrich Y, Ernst N, Felderer M, Feldt R, Filieri A, de França BBN, Furia CA, Gay G, Gold N, Graziotin D,

He P, Hoda R, Juristo N, Kitchenham B, Lenarduzzi V, Martínez J, Melegati J, Mendez D, Menzies T, Molleri J, Pfahl D, Robbes R, Russo D, Saarimäki N, Sarro F, Taibi D, Siegmund J, Spinellis D, Staron M, Stol K, Storey MA, Taibi D, Tamburri D, Torchiano M, Treude C, Turhan B, Wang X, Vegas S (2021) Empirical Standards for Software Engineering Research. 2010.03525

Rico S, Ali NB, Engström E, Host M (2020) Guidelines for conducting interactive rapid reviews in software engineering – from a focus on technology transfer to knowledge exchange. URL https://doi.org/10.5281/zenodo.4327725

Rico S, Ali NB, Engström E, Höst M (2024) Experiences from conducting rapid reviews in collaboration with practitioners — Two industrial cases. Information and Software Technology 167:107364

Rufino Júnior R, Classe TMd, Santos RPd, Siqueira SWM (2023) Current Risk Situation Training in Industry, and Games as a Strategy for Playful, Engaging and Motivating Training. Journal on Interactive Systems 14(1):138–156

Runeson P, Host M, Rainer A, Regnell B (2012) Case Study Research in Software Engineering: Guidelines and Examples, 1st edn. Wiley Publishing

dos Santos PSM, Travassos GH (2011a) Action Research Can Swing the Balance in Experimental Software Engineering, vol 83

dos Santos PSM, Travassos GH (2011b) Action Research Can Swing the Balance in Experimental Software Engineering, vol 83

dos Santos RP, Fachada N, Beko M, Leithardt VRQ (2023) A Rapid Review on the Use of Free and Open Source Technologies and Software Applied to Precision Agriculture Practices. Journal of Sensor and Actuator Networks 12(2)

Santos V, Goldman A, Filho H, Martins D, Cortés M (2014) The Influence of Organizational Factors on Inter-team Knowledge Sharing Effectiveness in Agile Environments. In: Hawaii International Conference on System Sciences, pp 4729–4738

Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, Burroughs H, Jinks C (2018) Saturation in qualitative research: exploring its conceptualization and operationalization. Quality & Quantity 52(4):1893–1907

Smite D, Dingsøyr T (2012) Fostering Cross-site Coordination through Awareness: An Investigation of State-of-the-Practice through a Focus Group Study. In: Euromicro Conference on Software Engineering and Advanced Applications, pp 337–344

Soini J (2008) Managing information and distributing knowledge in a knowledge-intensive business environment. In: Portland International Conference on Management of Engineering & Technology (PICMET), pp 889–894

Soini J, Makinen T, Tenhunen V (2007) Managing and Processing Knowledge Sharing between Software Organizations: A Case Study. In: Portland International Conference on Management of Engineering & Technology (PICMET), pp 1108–1113

Song Q, Borg M, Engström E, Ardö H, Rico S (2022) Exploring ML Testing in Practice: Lessons Learned from an Interactive Rapid Review with Axis Communications. In: Proceedings of the 1st International Conference on AI Engineering: Software Engineering for AI, CAIN '22, p 10–21

Stringer ET (2007) Action research third edition, 3rd edn. Sage Publications, Inc.

Tracy SJ (2010) Qualitative Quality: Eight "Big-Tent" Criteria for Excellent Qualitative Research. Qualitative Inquiry 16(10):837–851

Tricco AC, Straus SE, Ghaffar A, Langlois EV (2022) Rapid reviews for health policy and systems decision-making: more important than ever before. Systematic Reviews 11(1):1–4

Vasanthapriyan S, Xiang J, Tian J, Xiong S (2017) Knowledge synthesis in software industries: a survey in Sri Lanka. Knowledge Management Research & Practice 15(3):413–430

Viana D, Rabelo J, Conte T, Vieira A, Barroso E, Dib M (2013) A qualitative study about the life cycle of lessons learned. In: International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE), pp 73–76

Viana D, Conte T, Marczak S, Ferreira R, de Souza C (2015) Knowledge Creation and Loss within a Software Organization: An Exploratory Case Study. In: Hawaii International Conference on System Sciences, pp 3980–3989

Watt A, Cameron A, Sturm L, Lathlean T, Babidge W, Blamey S, Facey K, Hailey D, Norderhaug I, Maddern G, et al (2008) Rapid reviews versus full systematic reviews: An inventory of current methods and practice in health technology assessment. International Journal of Technology Assessment in Health Care 24(2):133–139

Weyuker EJ, Bell RM, Ostrand TJ (2011) Replicate, replicate, replicate. In: 2011 Second International Workshop on Replication in Empirical Software Engineering Research, pp 71–77

Wilson MG, Oliver S, Melendez-Torres GJ, Lavis JN, Waddell K, Dickson K (2021) Paper 3: Selecting rapid review methods for complex questions related to health policy and system issues. Systematic Reviews 10(1):1–8

Yglesias K (1998) IBM's reuse programs: knowledge management and software reuse. In: International Conference on Software Reuse, pp 156–165

Šmite D, Moe NB, Šāblis A, Wohlin C (2017) Software teams and their knowledge networks in large-scale software development. Information and Software Technology 86:71–86

I Complementary Information on the RR

For the purpose of increasing traceability and reproducibility, this section includes additional information on the rapid review presented in Section 6.

- Studies supporting RR recommendations. Table 10 report the primary studies associated with the recommendations for improving KM practices obtained as RR results.

II Complementary Information on the Data Analysis

To enhance traceability, this section provides additional details on the data analysis described in Section 6. Specifically, Table 11 reports the themes resulting from the analysis and the data from which each theme was identified.

Recommendations	Studies that support each recommendation
Definition	
KM strategy	(Yglesias, 1998; Santos et al., 2014)
Simple & clear processes	(Birk and Dingsøyr, 2005; Khan et al., 2012; Heredia et al., 2014)
Defining responsibilities	(Smite and Dingsøyr, 2012; Šmite et al., 2017)
Metrics	(Soini, 2008; Jurado et al., 2015)
Communication	
Communicate strategy	(Yglesias, 1998)
Emphasize benefits	(Soini, 2008; Heredia et al., 2014; Chen et al., 2018)
KM Training	(Matturro and Silva, 2010; Viana et al., 2015; Vasanthapriyan et al., 2017)
Values	
Commitment	(Khan et al., 2012; Heredia et al., 2014)
Communication	(Khan et al., 2012)
Cooperation	(Khan et al., 2012; Pulkkinen. et al., 2007; Soini, 2008; Šmite et al., 2017; Humayun et al., 2013)
Trust	(Khan et al., 2012; Soini, 2008; Humayun et al., 2013)
Horizontality	(Yglesias, 1998)
Motivation	(Soini et al., 2007; Khan et al., 2012; Chen et al., 2018)
Resources	
Tools	(Birk and Dingsøyr, 2005; Gervigny and Nagowah, 2017; Santos et al., 2014; Smite and Dingsøyr, 2012; Kammani et al., 2013)
Social software	(Santos et al., 2014; Heredia et al., 2017)
Categories	(Milovanović et al., 2012)
Gamification	(Jurado et al., 2015)
Training	(Viana et al., 2015; Khan et al., 2012)
Rotation of stakeholders	(Khan et al., 2012)
Rewards	(Vasanthapriyan et al., 2017)
Record of lessons learned	(Viana et al., 2013; Matturro and Silva, 2010)

Table 10 Studies supporting RR recommendations.

Theme	**Early attitudes** *Comments from the workshop participants (one requester and three members of the development team) during the discussion of the evidence.*	**First Questionnaire** *Participants' responses to the questionnaire circulated at the end of the workshop.*	**Follow-up Questionnaire** *Response of the requesters to the questionnaire circulated eight months after the workshop.*
Applicable to software engineering practice[d]	Mentioned by all participants.		Mentioned by one of the requesters
Novel approach to support decision-making[d]			Mentioned by both requesters.
Fostered the learning of new concepts[d]	Expressed in attitudes of all participants.		
Offer reliable content[d]	Mentioned by two participants.		Mentioned by both requesters.
Problem-oriented[d]		Question E4: Do the recommendations seem useful for solving the problems? - 3 participants totally agreed - 2 participants somewhat agreed	Mentioned by one of the requesters
Improve problem comprehension[d]		Question E5: Did the project help to better understand the problems? - 1 participant totally agreed - 4 participants somewhat agreed	
Discussing the findings of the RR is needed[d]		Question E6: Was the workshop to present the results good? - 3 participants totally agreed - 2 participants somewhat agreed	
Company situation[i]			Mentioned by one of the requesters
Results in a timely manner[i]			Mentioned by one of the requesters
Use of the results[i]			Mentioned by both requesters.

[d] Themes identified during the deductive stage of analysis.
[i] Themes identified during the inductive stage of analysis.

Table 11 Data by collection stage supporting the themes identified in the analysis.

Software Engineering Research Using Rapid Reviews: A Systematic Review

Sebastián Pizard[1] and Barbara Kitchenham[2]

[1]*School of Engineering, Universidad de la República, Montevideo, Uruguay. spizard@fing.edu.uy*
[2]*School of Computing and Mathematics, Keele University, Keele, UK.*

Supplementary materials to the paper *"Using Rapid Reviews to Support Software Engineering Practice: A Systematic Review and a Replication Study"*. It includes an extended version of the systematic review of rapid review research in SE presented in the referred paper. August 2024.

I. GOAL AND METHOD

In order to assess take-up of RRs in SE, we conducted a systematic review of RR research based on citation analysis of two of Cartaxo's papers ([7] and [8]). Our research questions were:

- RQI: What is the extent of take-up of RRs in the SE domain?
- RQII: What was the scope of these studies?
- RQII: What are the methodological characteristics of the reported RRs?
- RQIV: Which studies contributed to assessing the value of RRs and what have they found?

We conducted searches for citations of Cartaxo et al.'s works on the respective publication sites (ACM and Springer) and Google Scholar. The first search took place on February 1, 2023, during which we examined 117 citations and identified 12 publications of interest. A second search was conducted on November 27, 2023, resulting in the identification of 150 citations. After removing duplicates, we found 23 publications of interest, including articles from conferences or journals and book chapters [1, 3–6, 9–12, 15–19, 21–23, 25, 26, 28–31].

The analysis was performed by Pizard under the supervision of Kitchenham. Both authors discussed the results of the analysis, identifying new information of interest that was subsequently extracted by Pizard. The exception to this process was Table VI, which was developed by Kitchenham and reviewed by Pizard.

II. RESULTS

A. RQI: Extent of take-up of RRs in the SE

We found 23 reporting RRs in SE, although one of them analyzed the process used by two previously published RRs [28]. Since this study includes additional objectives and information to the original studies, we maintain it in our following analysis except in situations where otherwise indicated. 15 of these studies were published in 2023 (prior to Nov 23).

Researchers from 13 countries participated in the studies. Table I shows the number of papers and researchers from the different countries of affiliation of the authors (some authors have more than one affiliation). A co-author of the original RR in SE studies participated in one of the studies [15].

B. RQII: Scope of the studies

13 of the studies report an RR aimed at acquiring knowledge in a specific field, with some studies explicitly indicating this intent while others do not (though they also refrain from reporting any alternative use or motivation). The remaining 10 studies had broader goals, as shown in Table II. For example, some of them sought to complement or validate the RRs' results while others used the RRs' results were used to develop a model, catalog, or artifact.

Nine studies report conducting an RR and complementing/comparing its results with stakeholders' opinions (refer to Table II). Additionally, in another study, the RR served as a starting point for collaboration with stakeholders, but neither their participation nor feedback is reported.

Table III shows topics addressed by RRs in the 22 studies excluding [28]. The number of topics covered is greater than the number of studies because some studies reported multiple RRs on different topics. Studies that reported multiple RRs on the same topic were counted once only. Table III confirms that 13 of the most commonly addressed topics related to the software process (i.e., software construction, software testing and software maintenance), and technology impact was another frequently addressed topic.

The vast majority of RRs included only white literature. However, three RRs include grey literature [4, 5, 30] and another one was conducted in conjunction with a grey literature review [15]. Two studies aimed to search for software tools (the selection ends in Gitlab or Github) [10, 30] and another study searched for both models and tools (both within white literature) [11].

C. RQIII: Methodological characteristics of published RRs

Before analyzing the review processes used to conduct RRs, it is necessary to consider the quality of their reports. The studies report the RRs with varying levels of detail and completeness. Table IV presents our assessment of these reports against the SE Guidelines for

Country	Papers	# of Papers	# of Researchers	Researchers
Brazil	[11, 15, 18, 19, 21, 29, 30]	7	24	Alan Lyra, Andrea Doreste, Bruno Cartaxo, Carlos Eduardo Barbosa, Cassio Andrade Furukawa, Clinton Hudson Moreira Pessoa, Débora Maria Barroso Paiva, Guilherme H. Travassos, Herbert Salazar, Jano Moreira de Souza, Leopoldo Teixeira, Maria Istela Cagnin, Matheus Argôlo, Michele dos Santos Soares, Rebeca C. Motta, Roberto Rufino Júnior, Rodrigo Pereira dos Santos, Rodrigo Pereira Pagliusi, Rogério P. dos Santos, Samuel Loli, Sean Wolfgand Matsui Siqueira, Tadeu Moreira de Classe, Vítor de Castro Paes, Yuri Oliveira de Lima
Italy	[3–5, 16–18]	6	15	Alessandro Marchetto, Anna Rita Fasolino, Antonia Bertolino, Azzurra Ragone, Berenice Fernández Nieto, Danilo Caivano, Domenico Amalfitano, Domenico Gigante, Felicita Di Giandomenico, Francesca Lonetti, Giovanni Dimauro, Giuseppe Scanniello, Maria T. Baldassarre, Simone Romano, Vita Santa Barletta
Sweden	[6, 28, 31]	3	9	Elizabeth Bjarnason, Emelie Engström, Håkan Ardö, Markus Borg, Martin Höst, Nauman bin Ali, Patrik Åberg, Qunying Song, Sergio Rico
Portugal	[10, 30]	2	5	Filipe Correia, Marko Beko, Nuno Fachada, Rogério P. dos Santos, Valderi R. Q. Leithardt
Germany	[9, 10]	2	8	Ana Cristina Franco da Silva, Carolin Rubner, Horst Sauer, Jonas Fritzsch, Justus Bogner, Markus Haug, Matthias Saft, Stefan Wagner
Romania	[25, 26]	2	3	Laura-Diana Radu, Sabina-Cristiana Necula, Vasile-Daniel Păvăloaia
UK	[18, 22]	2	2	Barbara Kitchenham, Santiago Matalonga
Chile	[12, 23]	2	4	Francisco Ponce, Gastón Márquez, Hernán Astudillo, Mauricio Hidalgo
France	[19]	1	2	Rebeca C. Motta, Káthia M. de Oliveira
Uruguay	[22]	1	3	Diego Vallespir, Fernando Acerenza, Sebastián Pizard
USA	[1]	1	2	Amr S. Abdelfattah, Tomas Cerny
Spain	[12]	1	1	Laura M. Castro
The Netherlands	[10]	1	1	Justus Bogner

TABLE I. Number of papers and researchers from the countries of authors' affiliations.

Reporting Secondary Studies (SEGRESS) [14].

Table V summarizes the reporting limitations and process changes that we found in the 22 papers reporting RRs (i.e, not including [28] as it analyzes two previously published RRs).

In most studies, adequately considering the RR process or its results is challenging due to insufficiently reported information. In general, detailed information is presented about the sources used and the information search stage, less information about selection and extraction. But the synthesis stage is the worst reported (only four studies report it adequately).

Table VI presents the motivation for the RR conduct and the reported use of protocols by the authors. Looking at the primary studies, there appears to be some misunderstanding about the term "protocol" within the context of an RR. Some studies acknowledge that an RR protocol is utilized to outline the organization of their specific RR, while others seem to use the term "RR protocol" to refer to the generic RR process.

It is noteworthy that only the authors of five of the studies reporting RRs explicitly indicate having used a protocol to guide the RR conduct and only one study makes it available.

Reproducibility is a desirable characteristic of secondary studies. The lack of proper reporting poses a threat to achieving it. Specifically, the failure to report synthesis methods hinders reproducibility in most studies reporting RRs in SE. Only two studies, namely [5, 20], include the date of search, the list of primary studies, and details of the synthesis methods used. Furthermore, only the latter study included the full protocol. Additionally, [5] utilized Google as a search engine, which may pose challenges for reproducibility due to its inherent characteristics. Despite this, the authors of [5] have provided supplementary material detailing the outcomes of the different stages. To our understanding, these two studies are the only ones offering characteristics that enable their reproduction.

D. RQV: Studies assessing the value of RRs

Six studies confirmed the value of RRs by validating their outcomes (or the models created from them) through collaboration with stakeholders outside the review team.

- In [11], the authors conducted an opinion survey with IT professionals (75% out of 20 with postgraduate studies) to validate the RR outcomes.
- In [31], an RR is conducted as a starting point in a collaboration with a software company.
- In [15], a catalog of object-relational mapping code smells in java was created using an RR results. An opinion survey is done to validate the results (97% out of 86 with a bachelor's degree or more).
- In [6], researchers worked collaboratively with a company. In particular, three RRs were conducted

Study	[23]	[26]	[4]	[17]	[18]	[1]	[25]	[30]	[12]	[16]	[21]	[5]	[3]	[22]	[10]	[9]	[29]	[11]	[31]	[15]	[6]	[19]	[28][1]
Publication year	19	20	23	23	22	23	23	23	23	23	23	23	21	23	23	22	23	22	22	20	23	23	23
Purpose of the study																							
Conducting an RR.	●	●	●	●	●	●	●	●	●	●	●	●	●										
Creating a model/catalog/artifact using the RR results.[2]																				●	●	●	
Conducting an RR and complementing/comparing its results with stakeholders' opinions.														●	●	●	●	●	●				
Investigating the RR methodology.																							●
Stakeholder participation																							
Type of stakeholders (P-Software industry practitioners, S-students, O-Practitioners from other industry)													O	O	S	P	P	P	P	P	P	S	P[1]
RR results were a starting point for collaboration with stakeholders.													●	●									
Stakeholders participated throughout the RR process.																	●		●		●		●[1]
Stakeholders validated RR results.																		●	●				●[1]
Stakeholders validated the model/catalog/artifact created.																				●	●	●	●[1]
Stakeholders participation was used to complement RR results															●			●					
Stakeholders participation's results were compared to RR results														●		●							

[1] This study investigates the RR reported in [31] and the first RR of the three reported in [6].
[2] Two of these studies involved conducting families of RRs [6, 19].

TABLE II. Research Using RR in SE: Citation Analysis of [7, 8] conducted in Nov 2023.

Topic	# of Studies	Studies
Software construction	5	[1, 9–11, 15]
Technology impact	5	[3, 4, 21, 25, 26]
Software testing	4	[16–18, 31]
Software maintenance	3	[9, 10, 23]
Software tools	2	[10, 30]
SE professional practice	2	[12, 29]
SE models and methods	2	[5, 6]
Software design	1	[1]
Software process	1	[19]
Evidence-based practice	1	[22]

TABLE III. Topics of the RRs conducted in SE (not including [28] as it analyzes two previously published RRs).

to create a software selection model, which was validated with a focus group and an application in the company.

- In [19], seven RRs were conducted to develop a roadmap for IoT development. To validate it, an experimental study was carried out where undergraduates used and evaluated the roadmap.

- In [27], the authors analyzed the artifacts of the RR reported in [31] and one of the RRs reported in [6]. They also conducted interviews with the RR review teams, aiming to gain a better understanding of how RRs were conducted.

Although the feedback reports are not extensive or detailed, in all studies stakeholders had a positive attitude towards the results confirming the value of RRs. Except for the study with undergraduates, the others included practitioners with education achievements of at least degree level.

In the three studies in which researchers collaborated with companies throughout the RR process [6, 29, 31], the practitioners were technology experts who specialized in topics related to RR questions. In two of those studies in which the results were validated with practitioners [6, 31], their perceptions of the results were positive, and their feedback was used to improve the RRs results. Despite the fact that in one of the studies, the results were not directly applicable for practitioners [31], both studies support the view that RRs are useful in industrial contexts.

In [27], the authors analyzed the RR reported in [31] and the first of the RRs reported in [6] to evaluate the application of recent guidelines for conducting RRs in collaboration with practitioners, as well as to comprehend the benefits and challenges associated with RRs. The results confirmed that the guidelines were adequate. Conducting RRs collaboratively benefited the relationship between researchers and practitioners, fostering an understanding of expectations and establishing a common terminology. The main challenges included divergent needs, inadequacy of the evidence found (necessitating the use of broad questions), and concerns about short timelines (RR conduct took a few months but with low weekly effort).

III. SUMMARY OF STUDIES, RR SHORTCUTS, & COMMENTS ON REPORTING

This section presents an individual analysis of each paper that includes a very brief summary of the study, comments on the RR process conducted and comments on the report that complement the information presented in Table IV.

A. Migrating from monolithic architecture to microservices: A Rapid Review [23]

The paper presents an RR conducted to learn techniques to migrate monolithic architecture to microservices.

RR shortcuts. Scopus is used as sources and Google Scholar for snowballing. Selection was done by a single reviewer, extraction was done by two reviewers (no process or agreement statistics are indicated). There is no quality assessment of the studies.

Comments on the report.

- It has six pages, not including those that contain references..
- The abstract is not structured but includes the necessary information.
- The problem and the rationale for the study are described, but not so much why the study is necessary (knowledge gap).
- It includes supplementary material with information on the results of the selection and data extraction.
- The synthesis carried out is more in the form of a mapping study. It is easy to clearly trace the presented results back to the extracted data (included in the supplemental material).

B. Disruptive Technologies in Smart Cities: A Survey on Current Trends and Challenges [26]

The paper presents a RR conducted to understand disruptive technologies used for the development of smart cities.

RR shortcuts. They are not explicitly indicated nor is it clear what is omitted or adapted from the SR process. Search is done in IEEE Xplore, Web of Science and Scopus (which in total return less than 100 papers). The quantity or roles of the reviewers are not explained. But it is a single author. There is no quality assessment of the studies.

Comments on the report.

- The problem is described, but not so much why the study is necessary (knowledge gap) nor is the rationale explained well.
- Only the quantity of studies obtained in each search/source is included.

Study	[23]	[26]	[4]	[17]	[18]	[1]	[25]	[30]	[12]	[16]	[21]	[5]	[3]	[22]	[10]	[9]	[29]	[11]	[31]	[15]	[6]	[19]	[27][1]
Title	●	◐	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	●	◐	●	●	●	●	●	●
Abstract	●	◐	◐	●	●	◐	◐	●	●	●	◐	●	◐	●	●	●	◐	◐	●	●	●	●	●
Introduction	◐	◐	●	●	●	◐	◐	●	◐	●	◐	●	●	●	●	●	◐	●	●	●	●	●	●
Methods																							
Eligibility criteria	●	○	●	●	●	◐	●	●	●	●	◐	●	●	◐	○	●	◐	●	●	●	●	●	-
Information sources	●	●	●	●	●	◐	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	-
Search strategy	●	◐	●	●	●	◐	●	●	●	●	◐	●	●	●	◐	●	●	●	●	●	◐	●	-
Selection process	●	◐	◐	◐	◐	◐	◐	◐	◐	●	◐	●	●	◐	◐	●	◐	●	◐	◐	◐	●	-
Data collection process	●	○	●	○	●	◐	○	◐	◐	○	◐	●	○	○	◐	●	◐	●	◐	○	◐	●	-
Data items	●	○	○	○	●	○	○	●	●	○	●	●	○	○	○	○	○	●	●	○	◐	●	-
Study risk of bias assessment[2]	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	◐	○	○	-
Effect measures	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Analysis and synthesis methods	○	○	●	●	◐	○	◐	◐	◐	○	○	●	◐	◐	○	◐	○	○	◐	○	◐	●	-
Reporting bias assessment[2]	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	-
Certainty assessment[2]	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	-
Results																							
Study selection	●	◐	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	-
Study characteristics	●	○	○	●	●	○	○	●	●	●	○	●	○	●	●	●	●	●	●	◐	○	●	-
Risk of bias in studies[2]	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	-
Results of individual studies	●	○	○	●	●	○	○	●	●	●	○	○	○	◐	●	●	○	●	●	○	○	●	-
Results of analyses and synthesis	●	◐	●	●	●	◐	◐	●	◐	●	●	●	◐	●	●	●	●	●	●	●	○	●	-
Reporting bias[2]	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	-
Discussion																							
Discussion	●	◐	◐	●	●	◐	◐	●	◐	●	◐	●	◐	◐	●	●	●	●	●	●	○	●	-
Registration and protocol[3]	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	-
Support/Funding	●	●	●	○	●	●	●	●	●	●	●	○	○	●	○	○	●	○	●	●	●	●	-
Availability of data, code and other materials	●	○	○	●	●	◐	○	○	○	○	○	●	○	●	○	●	○	○	●	●	○	●	-

Categories: ●indicates "Acceptably reported", ◐indicates "Poorly/partially reported", ○indicates "Not reported" and - is used for "Not applicable".

[1] This study analyzes two previously reported RRs [6, 31]. Therefore, many items of the SEGRESS guidelines do not apply and must be read in those studies previously presented.

[2] Its report depends on the stages actually carried out in the RRs. Usually their omission is considered within the shortcuts in the RR process.

[3] There is usually no secondary studies protocol registration in SE, although this is starting to change with the "Registered Protocols" tracks at several conferences.

TABLE IV. Verification of studies against SEGRESS guidelines for reporting secondary studies.

RR Process Issues	# of Studies	Studies
RR Reporting Issues Decreasing Transparency and Reproducibility		
Did not report synthesis methods adequately	18	[1, 3, 6, 9–12, 15, 16, 18, 21–23, 25, 26, 29–31]
Did not include the full date of the search	15	[1, 3, 6, 9–12, 15, 17, 18, 21, 25, 26, 29, 31]
Did not report number and roles of reviewers	9	[1, 10, 12, 15, 17, 18, 21, 25, 26]
Only reported number and roles of reviewers for some stages	8	[3, 4, 6, 9, 11, 29–31]
Did not cite primary studies	7	[3, 4, 6, 15, 21, 25, 26]
Did not mention limitations	6	[4, 6, 21, 22, 25, 26]
Reported results only via an Evidence Briefing	2	[3, 15]
Process Changes that could Bias RR Recommendations		
Omitted risk of bias evaluation (i.e., quality assessment)	21	all studies except [15]
Used a single search engine	12	[3, 4, 6, 15–17, 19, 21–23, 25, 29]
Used a single researcher for 1 or more stages	8	[3, 9, 11, 19, 22, 23, 29, 30]
Used a subset of the studies found in searches	4	[5, 9, 15, 25]
Included additional studies without explanation	3	[6, 12, 26]
RR Risk Reduction Processes		
Used a single search engine complemented with snowballing	4	[16, 19, 21, 23]
Used an Evidence Briefing with additional commentary about RQs	2	[19, 29]
Used only primary studies cited in related SRs	1	[31]
Used tools to assist analysis and classification	1	[25]

TABLE V. Processes adopted in SE RRs (not including [28] as it analyzes two previously published RRs).

Study	Rationale for RR conduct	Specific client	Using a protocol is reported
[1]	RR applies to practical problems	No	No
[3]	Fast delivery of results	Project researchers	No
[4]	No specific discussion	No	No - Generic process
[5]	RR benefits explained	No	No - Generic process
[6]	Supported close collaboration	Yes	No
[9]	Industry relevance and efficiency	No	No - Generic process
[10]	Fast delivery of results	No	No
[11]	RR benefits explained	No	Yes - Protocol not included
[12]	Fast delivery of results	No	No[1]
[15]	RR related to decision making & practical problem	Project researchers	No
[16]	Speed up knowledge transfer	No	No - Generic process
[17]	Fast delivery of results	Project researchers	No
[18]	Quick & resource efficient	No	Yes - Protocol not included
[19]	No specific discussion	Project researchers	Yes - Protocols included in supplementary materials
[21]	Fast delivery of results	No	No - Generic process
[22]	Fast delivery of relevant evidence	Project researchers	No
[23]	RR benefits explained	No	No - Generic process
[24]	RR benefits explained	No	No
[26]	RR benefits explained	No	No
[29]	Fast delivery of results	Yes	Yes - Protocol not included
[30]	Scoping review	No	No
[31]	Practitioners context & knowledge exchange	Yes	Yes - Protocol not included

[1] Although the authors indicate using a protocol, they do not appear to have produced it prior to conducting the RR steps.

TABLE VI. Rationale for RR and use of protocols (not including [28] as it analyzes two previously published RRs).

- The quantity or roles of the reviewers are not clear.
- It is not indicated the date of the searches. It is said that information from other sources (without more details) is added to complement information.
- The limitations of the evidence or the process are not discussed.

C. The Social Impact of Generative AI: An Analysis on ChatGPT [4]

The paper presents a RR conducted to understand the social impacts of ChatGPT. The RR considered two different searches, one on grey literature (blogs and news articles) and another on white literature (using Google Scholar).

RR shortcuts. The shortcuts considered are not explicitly indicated. But only consider a search engine and do not make quality assessment of the studies.

Comments on the report.

- It is not explicitly stated that it is a rapid review, although it says "we followed the protocol proposed in (Cartaxo, 2018), and we complemented the review process with the strategies presented in (Kitchenham and Charters, 2007)".
- No quantity or roles of reviewers for selection or synthesis are indicated. Data extraction was carried out by two reviewers and a third resolved differences.
- The authors include a link for supplementary material on extracted data but it does not work (accessed 1/9/2024).
- Limitations of the evidence or the process are not

discussed.

D. A Rapid Review on Fuzz Security Testing for Software Protocol Implementations [17]

The paper reports a RR conducted to study fuzz security testing for software implementations of communication protocols.

RR shortcuts. The shortcuts considered are not explicitly indicated. But they only considered one search engine (Scopus). It is not known if they made shortcuts in the selection because the process is not explained. There is no quality assessment of the studies.

Comments on the report.

- The abstract is not structured but includes the necessary information.
- The search returns 48 studies and the accessible full papers, which are 45, are selected. The selection process is not explained, although the criteria are.
- No roles or more elaborate selection process are indicated.
- Limitations of the RR process are included.
- Includes supplementary material with a list of selected studies and their classification.
- There are no references to primary studies in the results (but they can be traced in the supplementary material). It is rather the analysis of a mapping study.

E. Alternatives for testing of context-aware software systems in non-academic settings: results from a Rapid Review [18]

The paper reports a RR conducted to understand how non-academic software projects deal with content variation when testing Context-Aware Software Systems. There were three interactions of the protocol conduction in 2019, 2020 and 2022 respectively.

RR shortcuts. Shortcuts are not explicitly stated. Although the authors used only Scopus and ACM Library. They say they have invested time in achieving consistency in the selection of papers, but it is not clear if this was a stage prior to the selection and then each reviewer had a group of papers or if they did it simultaneously. There is no quality assessment of the studies. However, study limitations reported by the authors of the studies are extracted and analyzed to answer one of the RQs.

Comments on the report.

- The process report is quite detailed, but does not indicate the number or roles of the reviewers. Although details of how they improved the consistency of their criterion are explained, that is, that the RR was done by more than one.
- The results begin with a narrative summary of the studies. Then, some questions are answered with categorization of the papers while others include themes that emerge from the studies (although it is not specified what method is used to obtain them).

F. Roadmap to Reasoning in Microservice Systems: A Rapid Review [1]

The paper presents a RR conducted to understand the microservice-based reasoning process (i.e., analysis and reasoning of microservice-based systems).

RR shortcuts. Shortcuts are not explicitly stated. Although only Scopus is used. There is no quality assessment of the studies.

Comments on the report.

- The background and need for the RR, nor its rationale, are not fully described.
- The RQs are composed and complex, they cover 8 or 9 items instead of 3.
- The number of reviewers or their roles during the stages of the process are not indicated.
- It is said that snowballing was done but not of what type or on what papers. The authors also indicate that they supplemented with ad hoc searches that are not explained.
- Collaboration from experts is reported but it is not indicated which ones or their role.
- There is a spreadsheet as supplementary material but it only seems to have search and selection results.
- Ten papers were found before snowballing, but although additional information was included in the snowballing, it is not detailed.
- Results are grouped into themes and stages of a map but no method or process of synthesis is explained. The list of selected papers is not indicated. And the answers to the RQs have entire paragraphs without references that it is not clear what evidence supports them.
- An analysis of the limitations of the study is included but not a discussion of the results (o this could be mixed with the results, it is not clear).

G. Artificial Intelligence as a Disruptive Technology—A Systematic Literature Review [25]

The paper reports a RR conducted to learn more about artificial intelligence as a disruptive technology and its effects on different domains.

RR shortcuts. The authors do not explicitly explain shortcuts used. But only one search engine (Web of science) was used. They excluded papers from authors with only 1 publication on the topic, they considered that "they did not have a serious approach toward this topic."

They used tools to automate analysis of the text of the studies. There is no quality assessment of the studies.

Comments on the report.

- In the title it is indicated that it is an SR, but in the paper the authors stated that their objective was to make a RR.
- No information is included about the knowledge gap or the background of the topic investigated. Research is carried out on AI as a disruptive technology, only an introduction to disruptive technologies is included and not on AI. Furthermore, the introduction given is quite simple.
- The number of reviewers or their roles is not included.
- The list of primary studies considered is not included.
- The authors used MonkeyLearn (a natural language processing software with AI) and VOSViewer to automate the classification and analysis of papers (without indicating much detail about that). Nothing more about synthesis, it seems too poor to me. Although then they not only present results from those tools but also more content grouped into topics that it is not known where it came from.
- An analysis of the limitations of the process or the evidence is not included.

Additional comments.

- The RQs do not seem appropriate for the search conducted. The authors say they survey disruptive technologies (DT) and compare with AI (as another disruptive technology), for example RQ1 aims to understand how DTs evolved over time. However the search string includes the term AI.
- The authors consider that RR is a problem-oriented method unlike SRs. The authors stated: "Among the advantages of using a rapid review, as mentioned by (Cartaxo, 2018), we contend that it provides reliable content while fostering the discovery of new additional concepts (1); Thus, it is considered a problem-oriented method that supports a flexible knowledge transfer environment (2)."

H. A Rapid Review on the Use of Free and Open Source Technologies and Software Applied to Precision Agriculture Practices [30]

The paper presents an RR conducted with the goal of identifying free and open-source software capable of solving real-world problems in precision architecture.

RR shortcuts. Shortcuts used are not indicated nor do they justify why they use an RR. The authors searched using Google Scholar, GitHub, and GitLab. One reviewer conducted searches. There is no quality assessment of the studies.

Comments on the report.

- Although the abstract is not structured, it has appropriate information.
- The introduction to the background and rationale of RR is adequate.
- One reviewer carried out the searches and another reviewed his work, but it is not clear whether this also applied to other stages such as selection or extraction.
- The data items collected are explained quite well but the synthesis process is not explained. The results involve classifications but also a narrative description of each study.
- A discussion of the results is reported, including an analysis of the limitations of the study.

I. Challenges to Use Role Playing in Software Engineering Education: A Rapid Review [12]

The paper presents an RR undertaken with the objective of providing a comprehensive perspective on the challenges associated with the utilization of Role-Playing in Software Engineering Education.

RR shortcuts. No shortcuts used are indicated and the justification for using a RR is not convincing (see additional comment below). The authors use Scopus, Web of Science and IEEE Xplore. The results are complemented with suggested sources by the research team (without more information). The number or roles of the reviewers are not explained. There is no quality assessment of the studies.

Comments on the report.

- It has 13 pages (Springer format), not including those that contain references.
- Although the abstract is not structured, it includes adequate information.
- Although the background and rationale of the study are described very well, the knowledge gap is not entirely clear nor are there any previous studies.
- No number or roles of reviewers are indicated.
- The explanation of the synthesis process is very brief. The papers were classified and a descriptive synthesis of each one was made. But the results include challenges grouped into certain categories, this is not explained how it was elaborated.
- There are results (see, e.g., 5.1) that include themes and explanations without references to primary studies.
- The discussion includes a (brief) analysis of the limitations of the process. Interpretation of the evidence in relation to background or other evidence is not included.

Additional comments.

- The authors consider that RR is a problem-oriented method unlike SRs. The authors, when explaining

RRs, stated: "Rapid Reviews (RR) are practice-oriented secondary studies, and their main goal is to provide evidence to support decision-making towards the solution, or at least attenuation, of issues practitioners face in practice [1] (Cartaxo, 2020, Book Chapter)". They also pointed out that RRs deliver evidence in less time and justify their use of an RR with these both arguments.

J. Model-based security testing in IoT systems: A Rapid Review [16]

The paper reports an RR conducted to understand the use of Model-based security testing in Internet-of-Things systems.

RR shortcuts. The authors used a single search engine (Scopus) in two rounds: until 2021 and then adding until April 2022. They complemented with "backward and forward snowballing procedure on a balanced and randomly chosen subset of selected papers." They do not indicate how the extraction and synthesis were done so there could be other shortcuts. There is no quality assessment of the studies.

Comments on the report.

- The questions were answered mainly with classification of the selected studies and with a narrative synthesis of some particular information from the studies gropued in some categories. In some questions, e.g. RQ6 on challenges or future directions themes used seems to be elaborated through some type of synthesis of the primary studies, but is not indicated how it is done.
- The discussion is brief but seems complete, including limitations, keyfindings and future research. There are no reflections about previous works but it seems that there were no works that cover all the concepts of the RR, that is, the results are not comparable.

Additional comments.

- Although the authors correctly explain RRs stating that they are complementary to the SRs, they also report the following, which it seems to apply currently only in the health field: "However, it has been shown that Rapid Review complemented by a rigorous snowballing process, can achieve as good results as Full Systematic Reviews" (10.1016/j.jclinepi.2017.12.001).

K. Analyzing the Challenges for Future Smart and Sustainable Cities [21]

The paper reports an RR undertaken to analyze approaches of Smart Cities implementation and main challenges.

RR shortcuts. No shortcuts used are explicitly indicated. But the authors used only one search engine (Scopus). The number or roles of the reviewers are not explained. There is no quality assessment of the studies.

Comments on the report.

- It is not indicated in the title that it is an RR.
- The abstract is not structured and does not include information on results or conclusions.
- Although some background is included, the knowledge gap is not made explicit.
- The authors report a mixture of of process definition and and process execution in places. Thus it seems that they have used a generic process to guide their planning, but then report the specific processes that they used in the paper.
- The number or roles of reviewers are not indicated.
- The results are presented using some classifications, but also some dimensions and challenges are presented (the qualitative synthesis technique used is not indicated).
- No limitations of the process or evidence are included in the discussion. Although related works are reviewed, they are not discussed against the evidence of the RR.

L. A Rapid Review of Responsible AI frameworks: How to guide the development of ethical AI [5]

The paper reports a RR undertaken to learn about frameworks proposed to help and speed up the adoption of Responsible Artificial Intelligence practices. They searched for white literature and also grey literature.

RR shortcuts. No shortcuts made are explicit. To search for white literature the authors used Scopus and Google Scholar (of the latter they only kept the first 20 pages). Algorithm Watch, OECD database and Google were used for grey literature. The selection and extraction was done by two reviewers and a third author resolved conflicts. There is no quality assessment of the studies.

Comments on the report.

- Although the abstract is not structured, it includes the corresponding information.
- The selection and extraction stages are very well explained.
- They include supplementary material on extracted data including classification of studies.
- The authors mostly performed categorization of the studies. They indicate in considerable detail the categories used and how they were classified. No methods or statistics of agreement between reviewers when classifying are indicated.
- The supplementary material describes what was found in each study.
- We accessed a preprint but we assume that the following items are met but in the paper published at

the conference: Support & Competing interests of the review authors.

M. On Internet-of-things Devices in Ambient Assisted Living Solutions [3]

The paper reports a RR conducted to learn about Internet-of-Things devices that have been using in Ambient Assisted Living Solutions for elderly people.

RR shortcuts. Shortcuts made are not explicitly reported. The authors used only one search engine (Scopus). A single reviewer performed the selection. The extraction and synthesis processes and the number or roles of the reviewers in these stages are not reported. There is no quality assessment of the studies. To disseminate the results, the authors prepared an Evidence Briefing.

Comments on the report.

- Paper is only five pages, not including those that contain references.
- It does not indicate in the title that it is an RR.
- The abstract is not structured and only presents results and conclusions.
- It is not explained how the extraction and synthesis were done or the number or role of the reviewers.
- They authors reported using thematic analysis but do not indicate whether more information, e.g. whether it was deductive or inductive or how they identified themes.
- The paper do not include the list of selected papers.
- The results are only reported in an evidence briefing.
- There is no discussion as such but limitations of the process and a very brief reflection on the study and its differences from related work are included.

N. Tools for Refactoring to Microservices [10]

The paper presents a RR conducted to learn about tools for refactoring to microservices. Tools (i.e. software) were searched. The paper presents a RR conducted to learn about tools for refactoring to microservices. Tools (i.e. software) were searched. Subsequently, three master's students inspected and tested the tools.

RR shortcuts. Shortcuts made are not explicitly stated. The authors searched in ACM DL, IEEE Xplore, Springer Link, and Google Scholar. The number and roles of reviewers are not explained so there may be other shortcuts. There is no quality assessment of the studies.

Comments on the report.

- It has five pages, not including those that contain references.
- It does not indicate in the title that it is an RR.
- It is not a structured abstract but its information is adequate.
- Although brief, the introduction includes all it should cover.
- It does not say what date the searches were carried out nor adapted search strings or details of the searches beyond a generic search string.
- The number or roles of the reviewers are not stated (only that 3 master's students inspected the tools found).
- A description of each tool and a table with some minimal categorizations is included.
- The discussion is very brief but includes limitations of the process and some few reflections on the results. For each tool there is practical information such as whether they are maintained or not. The discussion seems acceptable according to the objective of the study.
- The authors include supplementary material with selected studies, extracted data and classification of what was found.

O. Adopting microservices and DevOps in the cyber-physical systems domain: A rapid review and case study [9]

The paper presents research on using microservices and DevOps in the Cyber-Physical Systems (CPS) domain, especially when migrating legacy systems. First the authors performed an RR to analyze the scientific state of art for microservices and DevOps in the context of CPS. Second, they conducted an interview-based case study at Siemens AG to compare the literature findings with industry experiences. With the interviews, the results of the RR were not validated with the practitioners, but rather questions were asked to obtain information that was later compared with the results of the RR.

RR shortcuts. Four reviewers participated in the RR and their roles in the process are indicated. Searched in Google Scholar, ACM DL, IEEE Xplore, Science Direct, and Springer Link. They only considered the first 50 results of each search. They also did backward and forward snowballing. The selection was made by pairs of reviewers. Although they did pilot tests of the extraction to agree on criteria, then the extraction of each paper was done by a single reviewer. Although the synthesis process is explained, it does not indicate number or roles of the reviewers. There is no quality assessment of the studies.

Comments on the report.

- Although the abstract is not structured, it has adequate information.
- The selection, extraction and synthesis processes are very well explained, including roles and activities taken by the reviewers (except for the synthesis).
- The authors include supplementary material with information of primary studies (assigned categories) and intermediate and final results of the thematic analysis.

- The discussion includes RR results and results of the interviews with Siemens AG personnel, including comparison with related work. The study also includes main takeaways and their implications. Limitations of the study are included, including the limitations of the RR process.

Additional comments.

- The authors consider that RR is a problem-oriented method unlike SRs. The authors say: "Structured approaches to literature surveys like systematic literature reviews or systematic mapping studies provide a rigorous and reproducible process, but also require considerable effort. Additionally, their results can be hard to integrate into industry practice. We therefore decided to conduct a Rapid Review [...] The protocol of a rapid review is still systematic, but may consciously sacrifice rigor and extensiveness for industry relevance and efficiency."

P. Current Risk Situation Training in Industry, and Games as a Strategy for Playful, Engaging and Motivating Training [29]

This study investigated Risk Situation Training in Industry and digital games as a strategy to address this training. It includes a survey conducted with training managers to understand training taught via traditional methods, and an RR conducted to examine the literature on the benefits of using games with a purpose to support risk situation training. The RR is done in collaboration with managers from industry with whom the initial problem and results are discussed.

RR shortcuts. A single search engine (Scopus) was used. The complete reading and extraction was done by a single reviewer. The number or roles of the reviewers in the selection are not stated. The results are reported in an evidence briefing (although it is two pages instead of one), although these results are expanded in the paper when answering the RQs. There is no quality assessment of the studies.

Comments on the report.

- It is not indicated in the title that it is an RR.
- The abstract is not structured and does not satisfactorily cover all the information that is necessary, it lacks of information about results and conclusions.
- Although a lot of related work on risk situation training and digital games for training is presented, the need for RR is not made explicit. It is indicated that there is an SR on risk situation training in a specific sector of the industry, although it was not validated with practitioners. The authors of the study do not specify whether this SR or its primary studies are used in any way.
- PICOC is used to define the search string.
- It is not reported number or roles of reviewers, it is only said that a single researcher did the complete reading and extraction.
- The RR's protocol (comprising planning and execution) and results were validated with the managers.
- The analysis process/method used is not indicated. It seems that categorization of the studies was done.
- The discussion and limitations include all parts of the study (also the RR results).
- Supplementary material is included with selection results and interview data. In addition, the evidence briefing is presented.

Q. Support for Accessible Software Coding: Results of a Rapid Literature Review [11]

The paper presents a study that sought to identify which models and tools support developers in the software coding phase in order to meet accessibility requirements. To validate the results of the RR, the authors conducted an opinion survey with professionals in the technological area.

RR shortcuts. The authors used two search engines (IEEE Xplore and Scopus). A single reviewer conducted the RR and other researchers evaluated what was done. The synthesis process is not explained. There is no quality assessment of the studies.

Comments on the report.

- According to the extraction form and the results, it seems that the papers were classified and in addition to that a summary of each study was included. However, the synthesis process was not explained.
- The link to the supplementary material does not work (Accessed 1/23/2024).
- The discussion included reflections of all stages of the study. It seems complete and also includes limitations of the RR process. Recommendations for practice are also included.

Additional comments.

- The authors stated that it is necessary to validate the results of an RR with practitioners (although this is not strictly necessary) and that is the rationale why they conducted opinion survey to validate the results. The authors specifically say: "As the purpose of the RRs is to provide information and evidence to professionals in a timely manner and thus contribute to solving practical problems faced by them, the evaluation of data extracted from the literature with professionals in the technology area is necessary."

R. Exploring ML testing in practice – Lessons learned from an interactive rapid review with Axis Communications [31]

The paper presents a RR conducted to establish the state of the art of machine learning (ML) testing. The RR is conducted in collaboration with practitioners from the company Axis Communications and was intended to be the beginning of collaboration on ML testing between the company and the university. The practitioners collaborated throughout the RR and gave feedback on the results.

RR shortcuts. For the search, the authors used three recently published secondary studies on ML testing. The researchers selected from the set of 180 primary studies included in those three secondary studies. They focused on answering a subset of the questions defined together with the company. Studies strongly connected to the company context were selected. Technological rules were extracted as a way to structure the synthesis and facilitate the process and reception of the evidence by the practitioners. As the RR was a starting point for collaboration, no formal dissemination activities were planned. There is no quality assessment of the studies.

Comments on the report.

- The abstract is not structured but contains all the necessary information.
- The search, selection and extraction stages are quite detailed although it is not indicated who the reviewers were. Only for selection the roles are reported.
- The synthesis involved extracting technological rules from the studies. A brief review of each study is also included. It is not reported who made this synthesis or how it was validated. Selected studies were categorized according to the answers to RQs and it is reported which questions could not be answered (they had many).
- Supplementary material is included with the primary studies and their classification according to the RQs they respond to.

S. Assessing attitudes towards evidence-based software engineering in a government agency [22]

In this paper the authors studied the attitudes towards EBSE of stakeholders working in a government agency and to assess whether knowledge of EBSE would impact their working practices. As a starting point, they conducted a RR of secondary studies on evidence-based practice in other disciplines. The RR results were used to complement the discussion of the study results with the government agency.

RR shortcuts. Only one search engine is used (SCOPUS). It is reported that a single author conducted the RR. An RR is conducted on secondary studies, but the authors do not justify their decision or whether this is due to shortening the process. There is no quality assessment of the studies.

Comments on the report.

- It does not indicate in the title that it is an RR, but the RR is conducted and reported as part of the related work.
- It is not explained how the extraction was done.
- They authors reported using thematic analysis but do not indicate whether more information, e.g. whether it was deductive or inductive or how they identified themes.
- Although the results of the empirical study are discussed in the discussion and compared with the results of the RR, there is no fair discussion of the limitations of the RR.
- The authors include supplementary material with search string, selected studies and some of its characteristics.

T. A Catalog of Object-Relational Mapping Code Smells for Java [15]

The paper presents a catalog of Object-Relational Mapping Code Smells for Java extracted from the state of research and practice, through a combination of an RR and a grey literature review. To evaluate the catalog the authors conducted an opinion survey with software developers.

RR shortcuts. A single search engine (Scopus) was used. The Grey literature review (GLR) implied searching with Google, limiting it to the first 50 results according to the page ranking. The other stages of the process were done jointly for both reviews. Sources that did not have a minimum quality standard (sources without sufficient detail or with unclear explanations) were excluded. The authors say that this standard could not be established a priori for the GLR given the nature of the material to be surveyed (no information is given about this for the RR). No number or roles of reviewers are indicated. The results are presented in an evidence briefing, although it seems that the opinion survey included more detail of the catalog for validation (including code examples). There is no quality assessment of the studies.

Comments on the report.

- The title does not include RR but the title seems appropriate since it includes a catalog constructed from the results of the RR.
- The abstract is not structured but contains adequate information.
- The authors conduct the stages of the RR and the GLR from the selection stage together.
- Provided as supplementary material: intermediate and final results of searches and selection. Evidence briefing with the results.
- No number or roles of reviewers are indicated.

- Although there is a subsection titled "Synthesis procedure", few information is reported about that. It is only reported that some studies explicitly mention problems and in other cases they inferred them.Not including roles or activities conducted.
- The results include a catalog of code smells and a list of primary studies from which they were inferred. In addition to the description of each code smell, notes and additional information are included that seem to be taken from primary studies.
- A validation of the catalog using a survey of practitioners is included. There is no general discussion, but the results are briefly positioned with respect to related work and the limitations of the study are reported.

Additional comments.

- The authors consider that RR is a problem-oriented method unlike SRs. The authors stated: "The use of RR has the main objective of providing evidence to assist decision making regarding problems that professionals usually have. Thus, RR is the best fit in our context, since this research was initially motivated by a problem in a real-world project."

U. Software selection in large-scale software engineering: A model and criteria based on interactive rapid reviews [6]

The paper presents a model to support software development organizations in selecting and evaluating software components and tools. The model was developed in collaboration with Ericsson AB practitioners using as a basis, in addition to the practitioners' experiences, the results of three RRs. Two of the RRs investigated software tools (one related to CASE tools and the other related to continuous deployment/Devops) and one investigated methods of assessing the quality of tools. The model has been validated through a focus group with practitioners from Ericsson AB and by practical use also in that company.

RR shortcuts. The first RR was conducted by a single reviewer, the search was done using a single search engine (Scopus). Another author included studies resulting from complementary searches (no further detail on this is added). It appears that the extraction and synthesis involved developing a set of criteria to use in the model (but no further information is provided). Scopus was also used in the second and third RR, but the roles or number of reviewers who participated are not detailed. The results of the three RRs were presented to the practitioners (it is not said how or by what means) to then develop the model. Apparently no quality assessment was carried out on the studies considered.

Comments on the report.

- In the first two RRs, some of the reviewers add papers from "complementary searches" without specifying more information.
- The first RR is described more completely. The others do not indicate who the reviewers were or their roles, or methods of synthesis.
- The results of the RRs were incorporated into the model that was being developed. Results of the RRs as such are not reported.
- The results of the RRs are not discussed either. Nor are limitations of the RRs and their process included.

V. An evidence-based roadmap for IoT software systems engineering [19]

The paper presents an evidence-based roadmap for Internet-of-Things (IoT) development to support developers in specifying, designing, and implementing IoT systems. The roadmap has seven facets and to prepare it, a RR was conducted for each of the facets. To validate the roadmap, an experimental study was carried out that involved a project to create a software system for Oximeter IoT for healthcare domain carried out by seven undergraduate students as part of an assignment in the Software Development course (2021).

RR shortcuts. A metaprocol was developed for all RRs. A single search engine (Scopus) was used, complemented by snowballing. The selection and extraction stages were carried out by a single reviewer (different for each RR), the coding of the results and the synthesis were carried out jointly between two reviewers in meetings. Upon completing the RR of a facet of the roadmap, the results were reviewed and the results of the previous RRs were adjusted. No quality assessment of the studies was conducted.

Comments on the report.

- The title does not include the term RR but they create a roadmap based on the results of the RRs.
- An extensive technical report (147 pages) with the RR report is included as supplementary material. It has a meta protocol and then the RR of each facet of the roadmap is explained. For each of the RRs, their results were reported using an evidence briefing. These reports are quite detailed.
- They also provide much more supplementary material on building the roadmap and studying observations.
- PICOC was used to build the search string.
- In the supplementary material there are conclusions (as a discussion of the results) of each RR. But limitations of the process or the evidence are not included. The paper includes analysis of threats to the validity of the roadmap and the observational study. Limitations of the RR process and evidence are included.

W. Experiences from conducting rapid reviews in collaboration with practitioners — Two industrial cases [28]

In this paper the authors analyzed the artifacts of the RR reported in [31] and one of the RRs reported in [6]. They also conducted interviews with the reviewers, aiming to gain a better understanding of how RRs were conducted. Specifically, they sought to evaluate the application of recent guidelines for conducting RRs in collaboration with practitioners, as well as to comprehend the benefits and challenges associated with RRs.

RR shortcuts. They are those described in the studies considered already reported before.

Comments on the report.

- Some information from the analyzed RRs is added: e.g. background of reviewers or other issues that may not be required by SEGRESS. For example, how the RR questions were selected together with the practitioners, since the study seeks to analyze the collaborative process used to conduct the RRs.

IV. STUDIES EXCLUDED IN LATE STAGES OF THE ANALYSIS

Here we discuss studies excluded in late stages of the analysis and the rationale for the decision.

In [13], the authors present a catalog and taxonomy of code smells prepared from a search in the literature. They do not reported the study as an RR and declare "The literature review was, to a large extent, inspired by the methodology behind rapid reviews". They did 56 searches (including both white and grey literature) for different code smells and then classify and analyze what they find to create a catalog and taxonomy. After searching they read the results to consider which existing taxonomy is the most referenced, to find out if there is a source that aggregates all the code smells and to investigate how the code smells are discussed. These are some of their research questions and although some findings are discussed, it is not clear that the searches they conducted should be considered as RRs. The authors only report information on the search stage, they do not declare having considered protocols or conducted other stages of RR process.

The studies presented in [20] and [2] (the latter is a technical report) are preliminary intermediate works of [19] and [18], respectively. Therefore we considered in our analysis only these last two.

V. LIMITATIONS

Our review is subject to certain limitations. The selection, extraction, and classification of publications were conducted by a single researcher (Pizard), potentially introducing bias into the results of these activities. To address this concern, we implemented two measures. Firstly, Pizard validated the selection, extraction, and classification results some days after their completion. Additionally, after the elaboration of this report, he conducted a further verification of the results against the identified publications. Secondly, alongside this report, we have made the spreadsheet containing information on the selection, extraction, and classification of the articles publicly available.

It is important to note that our results exclusively encompass publications that reference the seminal works on RRs in SE authored by Cartaxo et al. We have not considered potential papers that address RRs but directly cite works or guidelines from the health field, where RRs as a method originated.

[1] Abdelfattah, A.S., Cerny, T., 2023. Roadmap to Reasoning in Microservice Systems: A Rapid Review. Applied Sciences 13.

[2] Amalfitano, D., Matalonga, S., Doreste, A., Fasolino, A.R., Travassos, G.H., 2019. A Rapid Review on Testing of Context-Aware Contemporary Software System. URL: `https://www.cos.ufrj.br/uploadfile/publicacao/2910.pdf`.

[3] Baldassarre, M.T., Caivano, D., Dimauro, G., Romano, S., Scanniello, G., 2021. On Internet-of-things Devices in Ambient Assisted Living Solutions. International Conference on Information Systems Development .

[4] Baldassarre, M.T., Caivano, D., Fernandez Nieto, B., Gigante, D., Ragone, A., 2023. The Social Impact of Generative AI: An Analysis on ChatGPT, p. 363–373.

[5] Barletta, V.S., Caivano, D., Gigante, D., Ragone, A., 2023. A Rapid Review of Responsible AI Frameworks: How to Guide the Development of Ethical AI, in: International Conference on Evaluation and Assessment in Software Engineering, p. 358–367.

[6] Bjarnason, E., Åberg, P., bin Ali, N., 2023. Software selection in large-scale software engineering: A model and criteria based on interactive rapid reviews. Empirical Software Engineering 28, 51.

[7] 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 (EASE), pp. 24–34.

[8] Cartaxo, B., Pinto, G., Soares, S., 2020. Contemporary Empirical Methods in Software Engineering. chapter Rapid Reviews in Software Engineering. pp. 357–384.

[9] Fritzsch, J., Bogner, J., Haug, M., Franco da Silva, A.C., Rubner, C., Saft, M., Sauer, H., Wagner, S., 2022. Adopting microservices and DevOps in the cyber-physical systems domain: A rapid review and case study.

Software: Practice and Experience .

[10] Fritzsch, J., Correia, F., Bogner, J., Wagner, S., 2023. Tools for Refactoring to Microservices: A Preliminary Usability Report, in: International Conference on Microservices, pp. 1–6.

[11] Furukawa, C.A., Soares, S., Cagnin, M.I., Paiva, B., 2022. Support for Accessible Software Coding : Results of a Rapid Literature Review. CLEI electronic journal, 25, 1–13.

[12] Hidalgo, M., Astudillo, H., Castro, L.M., 2024. Challenges to use Role Playing in Software Engineering Education: A Rapid Review, in: Applied Informatics, pp. 245–260.

[13] Jerzyk, M., Madeyski, L., 2023. Code Smells: A Comprehensive Online Catalog and Taxonomy. pp. 543–576.

[14] Kitchenham, B., Madeyski, L., Budgen, D., 2023. SEGRESS: Software Engineering Guidelines for REporting Secondary Studies. IEEE Transactions on Software Engineering 49, 1273–1298.

[15] Loli, S., Teixeira, L., Cartaxo, B., 2020. A Catalog of Object-Relational Mapping Code Smells for Java, p. 82–91.

[16] Lonetti, F., Bertolino, A., Di Giandomenico, F., 2023. Model-based security testing in IoT systems: A Rapid Review. Information and Software Technology 164, 107326.

[17] Marchetto, A., 2023. A rapid review on fuzz security testing for software protocol implementations, in: Bonfanti, S., Gargantini, A., Salvaneschi, P. (Eds.), Testing Software and Systems, Springer Nature Switzerland, Cham. pp. 3–20.

[18] Matalonga, S., Amalfitano, D., Doreste, A., Fasolino, A.R., Travassos, G.H., 2022. Alternatives for testing of context-aware software systems in non-academic settings: results from a Rapid Review. Information and Software Technology 149, 106937.

[19] Motta, R.C., de Oliveira, K.M., Travassos, G.H., 2023. An evidence-based roadmap for IoT software systems engineering. Journal of Systems and Software 201, 111680.

[20] Motta, R.C., de Oliveira, K.M., Travassos, G.H., 2020. Towards a Roadmap for the Internet of Things Software Systems Engineering, in: International Conference on Management of Digital EcoSystems, p. 111–114.

[21] Paes, V.d.C., Pessoa, C.H.M., Pagliusi, R.P., Barbosa, C.E., Argôlo, M., de Lima, Y.O., Salazar, H., Lyra, A., de Souza, J.M., 2023. Analyzing the Challenges for Future Smart and Sustainable Cities. Sustainability 15.

[22] Pizard, S., Acerenza, F., Vallespir, D., Kitchenham, B., 2023. Assessing attitudes towards evidence-based software engineering in a government agency. Information and Software Technology , 107101.

[23] Ponce, F., Márquez, G., Astudillo, H., 2019. Migrating from monolithic architecture to microservices: A Rapid Review, in: 2019 38th International Conference of the Chilean Computer Science Society (SCCC), pp. 1–7.

[24] Păvăloaia, V.D., Necula, S.C., 2023a. Artificial Intelligence as a Disruptive Technology - A Systematic Literature Review. Electronics 12.

[25] Păvăloaia, V.D., Necula, S.C., 2023b. Artificial Intelligence as a Disruptive Technology: A Systematic Literature Review. Electronics 12.

[26] Radu, L.D., 2020. Disruptive Technologies in Smart Cities: A Survey on Current Trends and Challenges. Smart Cities 3, 1022–1038.

[27] Rico, S., Ali, N.B., Engström, E., Host, M., 2020. Guidelines for conducting interactive rapid reviews in software engineering – from a focus on technology transfer to knowledge exchange. URL: `https://doi.org/10.5281/zenodo.4327725`.

[28] Rico, S., Ali, N.B., Engström, E., Höst, M., 2024. Experiences from conducting rapid reviews in collaboration with practitioners — Two industrial cases. Information and Software Technology 167, 107364.

[29] Rufino Júnior, R., Classe, T.M.d., Santos, R.P.d., Siqueira, S.W.M., 2023. Current Risk Situation Training in Industry, and Games as a Strategy for Playful, Engaging and Motivating Training. Journal on Interactive Systems 14, 138–156.

[30] dos Santos, R.P., Fachada, N., Beko, M., Leithardt, V.R.Q., 2023. A Rapid Review on the Use of Free and Open Source Technologies and Software Applied to Precision Agriculture Practices. Journal of Sensor and Actuator Networks 12.

[31] Song, Q., Borg, M., Engström, E., Ardö, H., Rico, S., 2022. Exploring ML Testing in Practice: Lessons Learned from an Interactive Rapid Review with Axis Communications, in: Proceedings of the 1st International Conference on AI Engineering: Software Engineering for AI, p. 10–21.

Evidence-based Software Engineering Revisited: Evaluation of a Practice-driven Application

Sebastián Pizard, Diego Vallespir, Barbara Kitchenham, *Member, IEEE*

Abstract*—*Context: **Systematic reviews (SRs) are the main method for supporting evidence-based software engineering (EBSE). However, although SRs have been widely adopted by SE researchers, so far, there have been no studies assessing the value of the broader EBSE framework. *Objective:* To evaluate an EBSE application in an industry environment. *Method:* Using the participant observation method, we conducted an EBSE-based project to address an industry problem. This includes collaborating with practitioners to diagnose the problem, collect evidence through a rapid review (RR), and transfer the results to the company. *Results:* The practitioners utilized some recommendations that addressed their problems. The biggest barriers we encountered were the difficulty in finding relevant evidence, the complexity of applying evidence, and the lack of guidelines or examples of EBSE use. Factors that supported our project were: close collaboration with the company and its commitment to process improvement, appropriate dissemination of the results, using an RR, and participation of external researchers. *Conclusions:* Currently, the use of EBSE is challenging and requires both professional and research skills. While it is a valuable instrument for researchers to enhance collaboration with industry, it may not be as suitable for general use by practitioners, as initially hoped.**

***Index Terms*—Evidence-based software engineering, Evidence-based practice, Rapid review, Participant observation, Industry-academia collaboration, Knowledge transfer.**

I. 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 [1]. Systematic reviews (SRs[1]) are a key component of EBSE as they provide a rigorous and transparent approach to searching and synthesizing the existing research evidence on a particular topic or research question. Since EBSE introduction in 2004, SRs have been widely adopted by SE researchers, allowing them to synthesize research on many software engineering (SE) topics. To illustrate, Kamei et al. identified 446 SRs published only in the top SE journals and conferences pre-2019 [2].

Although SRs support EBSE, they are not synonymous [3]. EBSE goes beyond the search and synthesis of evidence to also include (1) converting practical problems into questions that can be answered with evidence, (2) applying evidence obtained from SRs by considering the context, preferences, and expertise of stakeholders, and (3) evaluating the use of evidence and the performance of the process.

Several authors acknowledged the lack of EBSE adoption by industry [4]–[8], with some of them highlighting that it may be because SRs do not address practice-relevant problems [4], [9] or lack of useful recommendations for practitioners [8]. In addition, very few SRs involve non-academic stakeholders, without whom, additional EBSE activities are unlikely to be needed. For example, among the 169 SRs published in 2011, 2014, and 2018 identified by Kamei et al. [2], we found only three SRs that were produced jointly with industry [10]–[12] (refer to Section III). In addition, some practice-driven SRs studied practical problems, presented evidence to practitioners, and evaluated the benefits of its use (e.g., [13]–[15]). However, none of them evaluated the overall application of EBSE and studied what challenges arise when using the approach. As researchers and educators, this lack of practical examples of EBSE use and evaluation raises questions about the purpose of EBSE, and whether we should simply remain focused on academic use of SRs.

In this study, we present an evaluation of an EBSE application in an industrial environment conducted to investigate the question:

- ***RQ:*** *What issues, barriers, and facilitators arise when using EBSE in an industry setting?*

Our study focused on investigating challenges or issues that may arise when implementing EBSE in an industry setting, and any factors that facilitated its application. Specifically, we conducted an EBSE-based project to address an industry problem in a software company. This included collaborating with the practitioners to diagnose the problems, collecting evidence through a Rapid Review[2] (RR), and transferring knowledge to the company. We studied in depth the process of applying EBSE using the participant observation method and by collecting and analyzing a large qualitative data set comprising meeting audio recordings, correspondence, and personal notes. Finally, we compared our results with the early concerns identified by the researchers who proposed EBSE twenty years ago.

The remainder of this paper is structured as follows. In Section II we outline the original EBSE proposal and early

Sebastián Pizard and Diego Vallespir are with the School of Engineering, Universidad de la República, Julio Herrera y Reissig 565, Montevideo, Uruguay E-mail: spizard@fing.edu.uy; dvallesp@fing.edu.uy.

Diego Vallespir is also with Pyxis Research, Pyxis, Blvr. España 2323, Montevideo, Uruguay E-mail: diego.vallespir@pyxis.tech

Barbara Kitchenham is with the School of Computing and Mathematics, Keele University, ST5 5BG Keele, Staffordshire, U.K. E-mail: b.a.kitchenham @keele.ac.uk.

[1] In this study, we use the term SR to refer to any form of systematic review, this includes, e.g., mappings studies (MS) and rapid reviews (RR).

[2] *"A rapid review is a form of knowledge synthesis that accelerates the process of conducting a traditional systematic review through streamlining or omitting various methods to produce evidence for stakeholders in a resource-efficient manner"* [16].

concerns. Section III includes a brief analysis of the extent to which SRs are being used to support EBSE applications. Section IV presents the design and context of our research. In Section V, we present details of our EBSE application and the results (i.e., issues, barriers, and facilitators). A discussion of the results and their significance and the limitations of the study is included in Section VI. Finally, we present our final remarks in Section VII.

II. EBSE Proposal and Early Concerns

In 2004, Kitchenham et al. proposed five steps that are needed to practice EBSE [1]. SRs, the core tool of the evidence-based approach, usually support steps 1-4 steps of the process.

1) Convert a relevant problem into an answerable question.
2) Find the best evidence[3] with which to answer the question.
3) Critically appraising the evidence for its validity (closeness to the truth), impact (size of the effect), and applicability (how useful it is likely to be).
4) Integrate the critical appraisal with SE expertise and stakeholders' values and circumstances.
5) Evaluate the effectiveness and efficiency of the previous steps and seek ways to improve them.

The proposal was based on the steps used successfully in Evidence-based Medicine (EBM). Despite the apparent similarity between the high-level process steps of EBM and EBSE, the authors reflected, this would not be a guarantee that the scientific, technological, and organizational mechanisms that support EBM will apply to EBSE. Thus, the authors subsequently examined in the rest of their paper the problems or challenges that researchers would face when applying EBSE. In 2005, when EBSE was introduced to practitioners by Dybå et al [17], the authors also reflected on the challenges that practitioners might face when using EBSE. Below we present a summary of the concerns identified in both studies[4].

A. Step 1: Ask an answerable question

Specificity of questions & Small body of evidence. The challenge in this step is to translate the practical problem into a question that is specific enough for its answer to contribute to the solution of the problem and broad enough to obtain answers from the available scientific literature. More specific questions are clearer and are a means to achieve more relevant results for the problem to be addressed in its specific context. In SE, less stringent questioning may be necessary [17] due to the smaller and more diverse body of empirical research compared to other disciplines such as healthcare. In healthcare, studies often feature controlled experiments with clear treatments and control groups, which are less common in SE. This diversity and lack of structure mean that SE practitioners may struggle to find answers to highly specific questions.

B. Step 2: Find the best evidence

Difficulty finding evidence. The body of evidence is fragmented with little attempt either to summarize topics or to integrate evidence. Both articles highlighted the lack of infrastructure to support the search for evidence. At the time, EBM already had specialist digital indexing systems, such as Medine, and the Cochrane Collaboration (www.cochrane.org) which published and updated SRs from all major areas of healthcare online. In SE, there were some online databases (e.g. IEEExplore and ACM-DL) that cataloged scientific articles, but none dedicated to the search for evidence.

Lack or inadequacy of evidence. Kitchenham et al. [1] criticized SE empirical research as immature. SE research-based studies usually did not acknowledge the problem of individual skill differences when comparing SE techniques. They also suggested that the lack of standards for empirical studies and the lack of replications were reasons that existing evidence was often unreliable. Dybå et al. also pointed out the need to report research results in a manner more accessible to practitioners [17]. Finally, both studies called for encouraging the gathering of evidence from studies of industry projects (e.g., using field experiments).

C. Step 3: Critically appraise the evidence

Difficulty assessing study quality. In the context of lack of standards for empirical research, the quality of published papers is likely to be poor. However, the same context means that critical appraisal is really difficult for practitioners (and often also for researchers) [17].

Need for contextual information. Dybå et al. highlighted the need to have detailed information on the context in which empirical studies are carried out, in order to be able to better evaluate them and decide how to integrate the evidence they present. (Contextual information is also required for EBSE Step 4.)

D. Step 4: Apply the evidence

Process Improvement Commitment. Both studies suggested that EBSE would work well in an organization with a strong commitment to process improvement (something strongly promoted at that time) [1].

Complexity of applying evidence & Collaboration between practitioners and researchers. Although there are certain decisions that can be made by individual practitioners, the decision-making process in SE usually considers organizational aspects, the experience, and skills of developers, customer requirements, and project constraints, among others. For this reason, the authors reflected that the process of applying evidence would be demanding, especially for practitioners, for which Dybå et al. recommend approaching experts or collaborating directly with researchers [17].

[3]There is a certain ambiguity in using "best evidence" in this context. Because if we already found the best evidence in Step 2, there is no point in appraising it critically in Step 3. A sensible meaning of "best" in Step 2 would be evidence that best matches the EBSE question defined in Step 1.

[4]This summary is the result of analyzing the original texts using thematic analysis [18] to identify the main issues, challenges, and facilitators presented by the authors. The analysis and the results were validated by Kitchenham, one of the authors of both papers.

E. Step 5: Evaluate performance

This step seeks both to reflect on the use of EBSE as well as to confirm that the changes introduced had the expected results [17].

Difficulty isolating effects. According to Kitchenham et al. [1], it is generally difficult to evaluate the performance of a particular SE technique since our concern is not usually the specific task to which the technique is applied but the final outcome of the project of which the task is a part. The authors argued that it is difficult to isolate the impact of a technique because (1) the techniques used interact with many others during the software development process and (2) the immediate outputs of a technique will not necessarily have a strong relationship with the outputs observed at the end of the project.

Early evaluations & Postmortem analysis. From the software process improvement viewpoint, Dybå et al. also stated that given the need to adapt and learn from the rapid changes in the software development process, we should not conduct evaluations only at the end of the project [17]. So they proposed that in addition to holding postmortem meetings, after-action reviews should also be held (i.e., brief meetings to evaluate a change while it is being carried out).

F. General aspects

A general barrier to adopting EBSE is the attitude of stakeholders to SE evidence, **Research evidence ignored.** Kitchenham et al. suggested that research results were depreciated by practitioners and other stakeholders because researchers address issues not relevant to industry and they present their results in a manner inappropriate and unsuitable for decision-makers[5].

III. Existing practice-driven use of EBSE

Despite the several potential problems suggested by its proponents, we are not aware of studies that directly assess practice-driven applications of EBSE.

Ten years ago, in the only related study that we know of, Santos and da Silva surveyed the authors of 44 SRs published between 2004 and 2010 [9] to investigate the motivations for their SRs. They did not ask if the secondary studies were commissioned by non-academic stakeholders or carried out in collaboration with industry, so we cannot confirm how many SRs correspond to the practice-driven use of EBSE. However, their results suggest a lack of connection between SRs and practice. Most researchers wanted to learn more about a topic or to support their own research. A quarter of the authors indicated that their SRs sought practical solutions that could be used in industry, although it was unclear whether they had confirmed the value of their results with practitioners.

It is fair to say that the SE community has tried to promote the use of SRs results. The Voice of Evidence (VoE) column appeared for a decade in the IEEE Software magazine (2007-2017) intending to extract practical lessons from SE articles (most of them SRs) to share with practitioners. According to its editors, the main challenge was to effectively translate the articles (most of them SRs) into takeaways that connect with practitioners' concerns [19]. Another currently active initiative is the EDOS Center newsletter[6] which outlines research results for Norwegian agencies. There are no impact assessments of either of the two initiatives (only a citation analysis of the VoE column [19]).

In summary, very little is known about the impact of SE SRs on practice, and the last study on SR author motivations was conducted more than eleven years ago. Motivated by this we decided to do a brief investigation to assess the extent to which SR are being used to support EBSE.

Before our investigation, we knew of only five studies whose authors indicate practice-driven uses of EBSE, all of which were supported by some form of SR [13]–[15], [20], [21]. To understand the extent of practice-driven use of EBSE better, we analyzed a sample of SRs in SE to identify whether they report any industry involvement (as a proxy for all practice-driven use of EBSE). We took as a basis the tertiary review reported by Kamei et al. (the most recent tertiary review sharing its SRs list) [2]. In this study, the authors analyzed 446 SRs published up to 2018 in venues with a minimum h5-index (20 for conferences and 25 for journals). We selected and analyzed the motivation and use of the results of all 169 listed SRs published in 2011, 2014, and 2018 (38% of the total). These SRs provide information about more recent SRs than those investigated by Santos and da Silva and allow us to investigate whether there are any clear changes in the motivation for conducting SRs over the period.

We classified the studies using the following categories:

- **Search for knowledge.** These SRs seek to gain knowledge of a particular field, although some indicate it explicitly and others do not (but neither do they report any other use or motivation). An example is an SR carried out to understand the state of the art of SE in startups [22].
- **SRs are complemented by other evidence.** SRs in this category are also conducted to gain knowledge, but the results are complemented or validated by conducting another study (e.g., surveys). In one such study, the authors conducted an SR to learn more about the terminology used in global SE (GSE) [23]. Subsequently, the results were complemented by a study with experts to create a GSE taxonomy.
- **Investigations of the EBSE process.** In these studies, the SR process is investigated and improved. In one of them, the authors investigated the repeatability of the SRs [24]. For that, they trained novice researchers to conduct an SR. Subsequently, their results were compared with those obtained in a same-purpose SR previously published.
- **Practice-driven SRs.** In these studies, the authors engage with companies or other non-academic stakeholders who defined the SRs' purposes or intended to use their results. In one such study, researchers were asked, during a joint industry-academia project, to build a tool to measure

[5]Kitchenham et al. originally discussed this issue in the context of Step 5.

[6]Effective Digitalization of the Public Sector (EDOS) is part of Simula Metropolitan Center for Digital Engineering. Its newsletter is available at https://enedos.substack.com/

TABLE I
SUMMARY OF THE CHARACTERISTICS OF THE PRACTICE-DRIVEN SRS FOUND.

Study	[10]	[20]	[13]	[11]	[12]	[21]	[14]	[15]
Type*	SR	SR	RR	MS	MS	SR	RR	RR
Publication year	2011	2013	2018	2018	2018	2022	2022	2023
Nature of Stakeholders Engagement								
Close collaboration reported			x				x	x
Part of broader collaborative projects	x			x	x	x	x	
Use of Results								
Results presented to company	x	x	x				x	x
Results used by company			x					x
Used in subsequent collaboration				x	x	x		
Feedback from Stakeholders								
Stakeholders feedback given	x	x	x				x	x
Positive use perception of evidence	x	x	x	x	x	x	x	x
Considered role of SR in decision making			x					
Not directly relevant to users	x						x	
Feedback used to improve SR	x						x	x

* SR refers to systematic reviews, MS to mapping studies, and RR to rapid reviews. As defined by the authors.

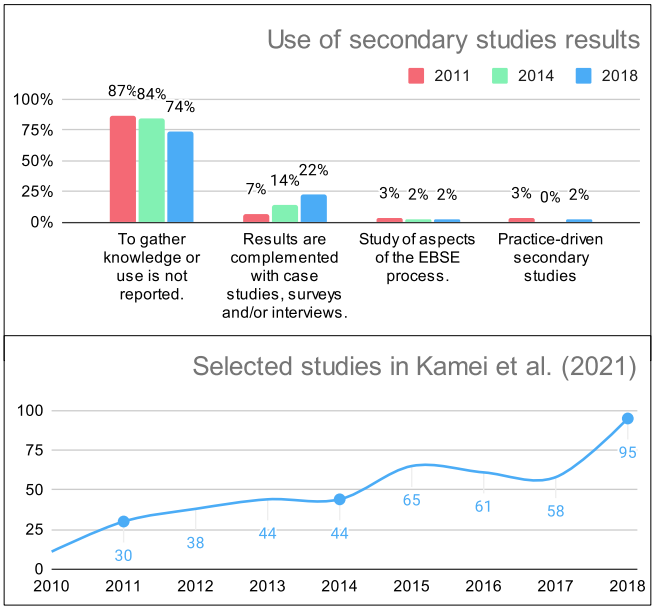


Fig. 1. Reported use of SRs selected from Kamei et al. [2].

socio-technical congruence [12]. They began their work by conducting an SR to learn more about the topic.

The results of this survey are shown in Figure 1. They confirm the lack of studies on the practice-driven use of EBSE. Only 3 of the 169 studies published in 2011, 2014 & 2018 reported SRs with practice-driven use (or motivation), all of which involved active collaborations with the industry.

In summary, we are aware of eight studies of practice-driven use of EBSE (five we knew beforehand and three that we obtained in our survey). The characteristics of those eight studies are shown in Table I.

It is also worth mentioning that: in [20] the stakeholders' requirements were surveyed through a case study, [13] includes the report of practitioners' perception of using RRs as a decision-making support method, [15] reports three RRs, all related to the development of a software selection model. Two investigated the selection of software tools (i.e., CASE tools and tools for continuous deployment/Devops, respectively), and one investigated methods of assessing the quality of tools.

Table I reveals four points of interest.

- All the studies reported collaborations with practitioners from individual companies (which helps SRs to be well-focused on stakeholders' requirements). None were commissioned by other stakeholders such as government agencies, professional bodies, or industry associations.
- Five of the eight studies reported that the companies made use of the SR results, although, in three of those cases, the results were used as part of a wider collaboration.
- RRs seem to be positioned as a method of special interest for the practice-driven use of EBSE.
- In all cases the feedback from the practitioners was positive. However, in two cases the results of the SR were not directly applicable.

This analysis has several limitations. It considers only published SRs (we have not deliberately searched for industry white papers). The SRs were classified only by the first author, and it shares the limitations of the study by Kamei et al. concerning the assembly of the list of SRs [2].

Our brief research confirms that there are few published practice-driven SRs. Although seven of the eight studies we found reported making use of the results of the SR, only one reported the practitioners' perceptions of using RRs as a decision-making support method. If we want EBSE to help bridge the gap between industry and academia, we should foster its use to improve practice. Still, we cannot recommend the adoption of EBSE if we have little evidence of its efficacy. Therefore, this indicates that our current study addresses a research gap by evaluating a practice-driven application of EBSE, identifying the naturally emerging issues, barriers, and facilitators.

IV. Study Design

Our trial of EBSE focused on assisting a small software company to address difficulties with its Knowledge Management (KM) processes.

The project team consisted of two students and one researcher (the first author) and the project was part of the students' capstone project. Students were trained on EBSE and worked to retrieve and synthesize evidence to help the company tackle its KM problems. The students carried out the project with close supervision, suggestions, and validations from the first author.

Our research method was participant observation. Participant observation is the process in which the observer remains in a social situation for the purpose of scientific research [25]. It has the *"unique strength of describing complex aspects of cognition, social interaction, and culture over time"* [26], which we believe is essential to study a complex issue such as the practice-driven use of EBSE.

Although we expected both to provide information the company could use to improve its SE process, and to observe the introduction and results of the process changes, our goal was not to introduce process changes, but to study the EBSE process (including the actions of the EBSE project team and the company staff). Our research goal meant we needed to maintain a neutral stance concerning the project's outcome. Crucial in this regard was acknowledging that participant observation studies demand researchers to be aware of, and control for, the risk of experimenter bias during when planning and conducting their study.

In our study, the first author immersed himself in an EBSE application setting to conduct the research. His role had two purposes: he led the EBSE-based project team, aiming to support a software development company while simultaneously observing the activities, people, artifacts, and interactions during the project. This approach enabled him to experience the practice-oriented use of EBSE firsthand, gaining profound insights into its complexities and nuances.

To minimize experimenter bias, the following steps were taken: (1) Data related to different activities were collected throughout the project from multiple sources. (2) Analysis of the data was validated at different stages by different researchers. Additionally, it should be noted that one of the co-authors, who held a positive bias towards EBSE, contributed solely to the post-trial discussions and presentation of results.

We used a rigorous pre-planned analysis process and systematically tracked all data to improve reliability (e.g., to avoid misinterpretation of the data) [27]. We used the O'Brien et al. checklist for reporting qualitative research to enhance reporting clarity and completeness [28]. We also considered the eight criteria for the quality of qualitative research proposed by Tracy [29], i.e., worthy topic, rich rigor, sincerity, credibility, resonance, significant contribution, ethics, and meaningful coherence. Our study addressed these criteria by detailing the context and procedures, combining diverse sources and reflections, being transparent and self-reflective, and aiming to make a valuable and coherent contribution to the field while adhering to ethical guidelines.

A. Research Program

The EBSE-based project that we conducted in collaboration with the software company served as the foundation for two interconnected yet distinct empirical studies. The first study, i.e., [30], entailed an external replication of the initial study proposing the use of RR in SE [13], *[Note for reviewers: The paper reporting this study is currently under review, and we include it as supplementary material. It includes detailed information on the RR conducted and feedback from practitioners.]* The second study is the one reported in this paper.

TABLE II
Comparison of steps of EBSE & Action Research

EBSE Steps	Action Research
1. Converting the need for information into an answerable question.	1. Diagnosis
2. Finding the best evidence with which to answer that question.	2. Planning
3. Appraising evidence validity, impact, and applicability.	3. Intervention
4. Integrating the appraised evidence with expertise and stakeholders' values and circumstances.	
5. Evaluating effectiveness and efficiency in executing previous steps and seeking ways to improve.	4. Evaluation
	5. Reflection

In the context of first study, we employed Action Research, replicating the same research method as in the original study. The action taken was the conduct of an RR to provide evidence to practitioners aimed at addressing their problems. However, from the company's perspective, their involvement was in an EBSE project (i.e., an application of all EBSE steps). We can define EBSE as a form of action research, where the goal is to adopt evidence-based actions to address a problem (refer to Table II for comparison of the steps of both). Given this comparison, it is reasonable to consider EBSE as a specific type of Action Research. This perspective enables us to utilize the Participant Observation method to evaluate the project as an example of EBSE.

It is noteworthy to mention that it was necessary to reconcile the application of both methods. According to its definitions, Participant Observation seeks to understand the phenomenon being studied, while Action Research focuses on creating positive change. In our case, we aimed to assist the company while also conducting a fair evaluation of the EBSE application. To reconcile these objectives, we strived to maintain a neutral perspective and focused our project on applying EBSE steps as they are defined. This involved explaining in advance to practitioners that our objective, in addition to trying to assist them, was to evaluate EBSE, and that we would base our suggestions for process improvements on scientific evidence collected through an RR. We also clarified to participants several times that both their positive and negative feedback was valuable. Finally, we analyzed and recorded any need for deviation from these objectives, e.g., including clarifications to the evidence obtained.

B. Research Context

Here we describe certain aspects of the research context.

Project team: The team consisted of the first author and two undergraduate students about to finish their computer science degrees. The first author has ten years of industry experience as a technical lead and software quality manager and twelve years as a member of the university. This study is part of his doctoral research that focuses on investigating EBSE adoption. Participating in the project was part of the students' capstone project. Also, both of them had full-time jobs related to software development. In particular, one of them was also part of the company's development team, and was so during the first half of the project. Both of them were trained in the planning and conduct of SRs. The training was led by the first author and based on an EBSE and SRs course he teaches [31], [32]. Finally, although Vallespir was not directly part of the team, Pizard consulted him regarding the team's decisions throughout the process.

Selection of Research Topic & Company: The EBSE team was the one that initiated the contact with the company and proposed the research project. Previously, we discussed which companies and organizations had a strong relationship with us and subsequently, we identified a research topic that was relevant to them. We chose our research topic to be about common problems present in the industry but not critical issues. In this way, practitioners could work together with our EBSE team without revealing sensitive information and with no expectation of dealing with emergencies or serious project problems. In addition, we sought a problem that did not have an *a priori* well-known solution in the software industry, but a solution to which would have a positive impact on the organization with which we worked. Finally, we verified that there were studies on the topic, carried out in non-academic contexts, that reported observations, lessons learned, or recommendations.

Given the characteristics of EBSE and the inexperience of the project members, we believe that the pre-selection of a topic area mitigated many risks related to knowledge of the domain, the stakeholders' expectations, and the existence of evidence on the topic.

In particular, we defined knowledge management (KM) as the research topic. It is a topic for which there are non-critical but important problems in the industry, especially in software development using agile methodologies (see, e.g., [33], [34]).

The company: The company was a UK company specializing in digital out-of-home (DOOH) advertising[7]. The company's IT department, located in Uruguay, was responsible for developing and maintaining a platform to manage advertising campaigns.

The requesters: The company's technical product leader and the project manager accompanied all the stages of the EBSE project, from the diagnosis of the problems to the dissemination of evidence. They answered questions, carried out intermediate validations, and received the collected evidence. For the purposes of this study, we have considered them as EBSE requesters. Their educational level was Intermediate[8] (one with *upper secondary education* and the other one with *post-secondary non-tertiary education*). As sources of information for supporting practice they usually talked to colleagues, read technology forums or blog articles, and watched technology videos (e.g., from the Microsoft Youtube channel). Neither of them consulted scientific literature.

C. Research Activities

The purpose of the project was to provide support to the company to improve its KM practices. We agreed with the company that our work would involve using EBSE, so we went through the five steps of the process. Figure 2 shows the main activities carried out (which will later be detailed in Section V).

The goal of our study to was to investigate the use of the EBSE framework. As a means of assessing the value of EBSE framework, we intended to identify potential barriers and facilitators to EBSE use that reflected the viewpoints of both the EBSE team members and the company staff members. We wanted the factors we identified both to be verifiable and to arise naturally as part of the EBSE project. To achieve our goals, we collected and recorded data throughout the EBSE project including final feedback. Data was obtained from a variety of sources: Audio recordings of all team meetings (including the final retrospective meeting) and meetings with the company, all emails and messaging app communications, a researcher's personal diary kept by the first author), and all EBSE project documents (including students' capstone project report, reports for requesters, and other artifacts).

We used a qualitative data analysis process with the following stages. First, the first author performed the initial data analysis using a method that was strongly based on thematic analysis with a realistic approach [18]. The second and third authors reviewed the results and suggested issues to review, reflect on, or expand on. Subsequently, the first author did a second analysis interpreting comparatively the results with the early concerns identified by the researchers who proposed EBSE twenty years ago. Finally, this analysis was revised and expanded by all the authors.

D. Ethical Issues

Although our university did not require our study to be approved by an ethics committee, we took care to consider possible negative impacts of the study on the participants. As recommended [35], we ensured that participation in the project was voluntary and the research process was transparent to all participants. Company members and students were informed of the characteristics of the research prior to giving their consent to participate.

There were two other major considerations. Firstly, ensuring that the students' education experience was not adversely

[7]Digital out-of-home advertising (DOOH) is advertising designed to reach consumers when they are not at home and that is also dynamically and digitally displayed. This includes digital transit, digital billboards, and digital place-based displays.

[8]According to UNESCO's ISCED 2011 classification. https://ilostat.ilo.org/resources/concepts-and-definitions/classification-education/

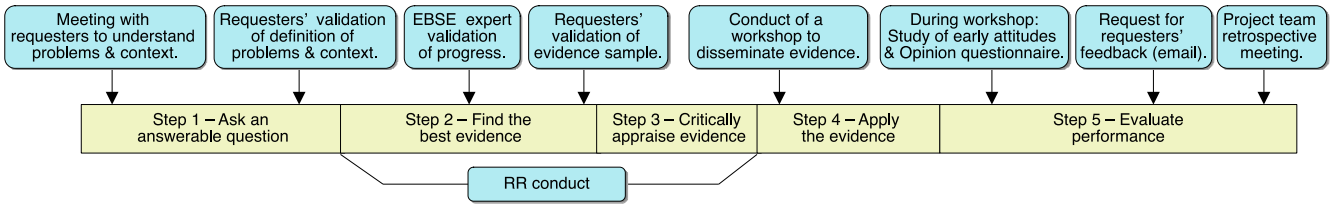


Fig. 2. Project stages.

impacted by the study: (1) Students should not be required to undertake tasks beyond their capabilities. This concern was addressed by appropriate training and supervision. (2) Students should not feel obligated to express support for the EBSE framework. The students were assured that the outcome of the study in terms of whether or not it was favorable to the use of EBSE would not impact their capstone project marks. Secondly, ensuring that the interests of the company are not adversely affected by the study: (1) The company would receive the best scientific information to help them address their process issues. This was assured by the personal experience and supervisory role of the first author. (2) Commercially or personally sensitive information would be kept confidential or anonymous as appropriate. Specifically, only the company and roles are identified, and specific comments are not attributed to specific individuals.

V. EBSE Application and Results

Below we present for each EBSE step the central characteristics of our application and the response to RQ, that is, the main issues (indicated with [·]), barriers (indicated with [–]), and facilitators (indicated with [+]) that arose during the process. As a complement to this section, the paper reporting the RR conduct includes information regarding the RR conduct and feedback from practitioners on the results of the RR, including representative quotes [30].

A. Step 1. Ask an answerable question

In this step, the team met with members of the company to understand their context and the problems they had in the KM area. We defined the following research question: *What are some empirically validated recommendations for knowledge management for software development companies?* As a validation activity, we prepared a summary of the context of the company and its KM problems which we shared with the requesters (refer to Table III).

[·] Considering requesters' needs. The requesters had no knowledge of SE research or EBSE, so we started the kickoff meeting by introducing both topics. We talked and agreed on the expectations about the results of the project. In addition, we found that the company had a continuous improvement process that involved identifying small improvements that were introduced in future sprints. Specifically, they told us that the results of our project were going to be incorporated into that improvement process. Thus, we understood that they preferred a set of small self-standing recommendations rather than a single major process change. This restriction had consequences in the subsequent stages (i.e., in the selection of primary studies, and in how we elaborated recommendations based on the evidence).

[·/-] Specificity of questions & Small body of evidence. Given the difficulty in finding evidence when testing specific questions in preliminary searches, we decided to use a rather broad question. Additional aspects (i.e., limiting the context of primary studies to small or midsize companies using agile methods, or accepting specific recommendations and not proposed frameworks or models) were considered in subsequent steps of EBSE, e.g., using them as criteria for prioritizing process changes.

[+] Strong link with company. The fact that one of the students was part of the company facilitated their willingness to work with us and also helped us better understand their problems.

TABLE III
Summary of Context and Problems related to KM.

The company is a spin-off of a UK advertising agency specializing in digital out-of-home (DOOH) and is responsible for managing the entire life cycle of a platform consisting of four products. The staff are geographically distributed, with the CEO and four account executives located outside of Uruguay, and the software development team based in Montevideo.
The team follows Scrum methodology with two-week sprints, daily stand-up meetings, and regular demos and retrospectives with stakeholders. They use various online tools for KM, including GitHub for code storage, Lucidcharts for architecture diagrams and planning, Visual Studio Online (VSO) wiki for test cases, VSO board for backlog management, Trello boards for tracking tasks, and Google Docs for architecture records and spreadsheets.
However, they face several challenges with KM. They struggle with finding the right documents, as there are duplicates or similar documents with different media and dates. Keeping the documentation up-to-date and eliminating unnecessary or outdated documents is also a challenge. The lack of standard definitions for document types creates confusion, with each person generally deciding what type of documentation to create.
The company is concerned about knowledge centralization in specific roles, particularly in the QA and DevOps manager positions, where there is only one team member for each role. This poses a risk in terms of knowledge sharing and continuity if these individuals are unavailable or leave the company.

B. Step 2. Find the best evidence

The question defined above was answered using a rapid review (RR). Its main characteristics are summarized in Table IV, following the proposal of Cartaxo et al. [6]. Before starting, we checked that there were no SRs on KM in agile methodologies that were able to answer the research question.

TABLE IV
MOST RELEVANT CHARACTERISTICS OF THE RAPID REVIEW CONDUCTED AS PART OF THE EBSE APPLICATION.

Problem	To find practical and applicable recommendations on KM for the company. The context of the studies should be similar to that of the company and the recommendations must have been empirically validated.
Research Question	What are some empirically validated recommendations for KM for software development companies?
Protocol	It was written during the first weeks of the RR. We did initial searches to validate the existence of evidence.
Stakeholders' roles	The requesters met with us during the diagnosis and validated the document that summarized its context and the problems to be addressed. They met with us to validate the evidence we found in the early stages of RR. Together with other software development team members, they participated in the results dissemination workshop.
Time Frame	The RR was done in three and a half months and the total time spent by team members was ∼150hs, including team meetings and knowledge dissemination transfer activities with practitioners.
Search Strategy	Keyword-based search in Scopus.
Selection Procedure	We considered available studies in English, with practical recommendations, addressing issues of KM in software development companies, and not presenting models or theoretical frameworks. Each reviewer assessed half of the candidate primary studies. To validate an adequate level of agreement, they reviewed the first 30 candidate primary studies together, and the kappa statistic was calculated (the value obtained was 0.618, which indicates a good level of agreement). Subsequently, the reviewers used two rounds: (1) checking titles and abstracts and (2) reading the full text. From the 425 studies returned by Scopus, we identified 21 primary studies [36]–[56].
Evidence Appraisal	We assessed the evidence's relevance to practice by studying the context and methods used for its generation.
Extraction Procedure	Each reviewer extracted data from half of the papers. Data extracted: Context of the study (agile, year of study, company), research method (experimental validation), validation result, and recommendations.
Synthesis Procedure	We used content analysis with an inductive approach (adapted from [57]). The stages carried out were: Translation into Spanish of the original texts, labeling of the different types of recommendations using open coding, grouping, and categorization of the fragments according to their codes (e.g., grouping similar or complementary codes), and, finally, creation of descriptions. As a tool to facilitate the coding, we used the freeware tool Saturate[9]. Each reviewer synthesized half of the papers. In two meetings with the first author, the results were reviewed and adjusted to obtain 21 recommendations for practice.
Report/Diffusion	We prepared an evidence briefing with the results and conducted a workshop with the development team.

[-] Difficulty finding evidence & Lack or inadequacy of evidence. In preliminary searches, we found that very few studies had adequate evidence to address our problem. The studies were difficult to find (so we had to test several questions and adjust their specificity, as discussed in the previous step). Furthermore, they did not include recommendations for practice suitable for our needs. To mitigate the latter issue, and knowing that very few SE studies usually include recommendations for practice [8], we decided to also consider those that included lessons learned, reflections, or certain observed behaviors. We analyzed those observations, as well as the context, to develop practice-oriented recommendations.

[+] Using a rapid review. We opted to conduct a RR because we had a low-resource setting. Students had a limited time period to complete their capstone project and could only allocate a limited effort each week. We also believed that the results of a RR would be adequate to address the issues identified. The RR method allowed us to work in an agile way while maintaining scientific rigor.

[+] Validations with requesters. We also conducted two validations with requesters to assess the evidence. They approved the sample evidence we presented to them. However, they wanted to obtain recommendations from contexts similar to their own. They also commented that some recommendations of the sample seemed useful to them but they did not know how to put them into practice. This activity was useful to validate the adequacy of the evidence found early in the process.

C. Step 3. Critically appraise the evidence

We evaluated the evidence's relevance to practice (i.e., applicability). This included understanding the context in which the evidence was generated and the research methods used. For example, for organizations that participated in a primary study, we identified in which country they carried out their activity and their size. In this way, we sought to provide more information to requesters so that they could better evaluate the evidence and select the one that best fits their context.

[–] Need for contextual information. Given the company context was that of a small company using agile methods, we did not need much detail about the research context of the primary studies. Even so, not all studies presented the same kind of information and some of them did not include enough context details.

D. Step 4. Applying the evidence

Since the RR was conducted by the EBSE project team, we needed a method of informing the company of the results of our RR. In addition, unlike evidence-based medicine, we were reporting a number of very different recommendations, not reporting evidence that compared two well-defined process options, so we also needed to help the company staff select the most suitable recommendations.

Firstly, following Cartaxo et al.'s reporting suggestions [4], we developed an evidence briefing to summarize the results of our RR. We, then, conducted a workshop to discuss our recommendations and to identify the specific process changes that the company could adopt. In the workshop, we presented the recommendations we derived from the evidence. Then, we assisted the company staff to prioritize the recommendations in terms of *ease of implementation* and *potential benefits*, in a hands-on exercise. We also discussed which of the recommendations the company could implement and how the process changes could be introduced. The project manager (one of the requesters) and three other members of the software development team participated in the meeting.

[–] Complexity of applying evidence. The recommendations obtained in the RR were general principles not specific

process changes. So, during the intermediate validations, the requesters expressed concerns regarding the implementation of these recommendations. Thus, although our EBSE application sought to have a strong emphasis on scientific evidence, it was necessary to include examples, suggestions, or clarifications that emerged from our professional practice and knowledge of the company. We did not add or remove any recommendations from the original set.

[+] Process improvement commitment & Collaboration between practitioners and researchers. Two factors facilitated the reception of the evidence. First, as the development team was used to improving their process, it was relatively easy for them to evaluate the evidence and consider how the recommendations could be applied to their processes. Second, the close collaboration between requesters and researchers throughout the EBSE project helped to make the results interesting to the development team, and the discussion served to better examine the recommendations and consider strategies to implement them.

[+] Appropriate dissemination. Based on the recommendation of Cartaxo et al. to incorporate discussions of the results as dissemination activities [13], we not only prepared an evidence briefing with the RR findings, but also held a workshop in which the attendees began to discuss ways of implementing the recommendations, the required effort and the potential benefits.

E. Step 5. Evaluate performance

In this step, we sought to address two issues: a) assessing whether the process change had a successful outcome and b) analyzing whether the team could have done a better job applying EBSE. We investigated both issues as follows.

We collected requesters' perceptions of the EBSE project results and the challenges faced during its conduct in three instances. First, we recorded (with prior approval) the workshop in audio to analyze the attendees' initial attitudes. At the end of the workshop, we circulated a questionnaire for participants to express their opinions about the project and its results. Finally, eight months later, we requested feedback on the process changes from the requesters via email. In addition, the EBSE team held a retrospective meeting to discuss and reflect on the EBSE application and its results.

Evaluation of results. None of the workshop participants suggested the recommendations were inappropriate. It is noteworthy that all attendees demonstrated a clear understanding of the recommendations and their potential implementation and actively participated in the prioritization process. Following the workshop, the company underwent significant changes, including an acquisition, role restructuring, shifts in organizational hierarchy, and eventual closure. However, even amidst these transformations, the requesters, in response to our email requesting feedback, confirmed that they had utilized some of the recommendations before the closure occurred. In the workshop and the subsequent feedback requests, the only negative comment regarding the EBSE process was related to the time it took to complete the RR (more detail in two paragraphs below).

Evaluation of the process. The main challenges we faced were the lack of guidelines and examples for conducting RRs for industry and the difficulty in finding appropriate evidence. At that time, there was only one previous study by Cartaxo et al. that applied RR in supporting SE practices [13]. Due to the EBSE team's lack of experience in conducting this type of secondary study, there was a substantial risk of not being able to provide evidence-based recommendations to address KM issues effectively. To mitigate this, we worked cautiously at each stage, conducting preliminary literature searches and verifying the adequacy of retrieved studies for the RR. Intermediate validations with the requesters were also carried out to ensure the relevance of the evidence found. The second challenge arose from the limited availability of studies with appropriate evidence. Given the scarcity of SE studies that offer practice recommendations, we decided to include primary studies with lessons learned or empirically validated observations.

[·] Considering requesters' needs (Step 1). One of the requesters would have preferred a shorter timeline. The limited weekly availability of the students led to the research taking three months to complete, which would have otherwise only required about two weeks of full-time work. In addition, the requesters had not emphasized EBSE project timescales, so we underestimated the need for prompt results. This highlights two key points: practitioners value quicker processes for results, and all requesters' needs are significant, not just those regarding the specific problem at hand.

[+] Early evaluations & Postmortem analysis. Both the retrospective meeting and the written feedback from requesters were useful to better understand the performance of the EBSE process and the company's perceptions of the results.

F. General aspects

Three factors affected all aspects of the EBSE project.

[–] Inexperience & Lack of guidelines and examples of using EBSE in the software industry. During our study, Cartaxo et al's research was the sole instance of employing a RR to support SE practice [13]. Also, we could not find detailed guidelines on how to apply each step of EBSE. Given our limited experience in conducting RRs and applying EBSE, we identified a substantial risk of being unable to develop scientifically-backed recommendations that effectively assist requesters in addressing their KM issues. To mitigate this, we carefully validated each EBSE activity. However, this caution resulted in extending the EBSE project timescales.

[–] Complexity when working with an industry partner. Collaborations between industry and academia usually face challenges, and our study was no exception. The most notable challenge was that we were unable to finish conducting EBSE final step, because of the changes that the company underwent. In addition, one of the requesters, the technical leader, was unable to attend the workshop (subsequently, the results were sent to him), and on a few occasions, the requesters' responses to our inquiries had a noticeable delay.

[+] Participation of external researcher. During primary study selection, we validated the progress in the use of EBSE

with an external researcher with some EBSE experience. This activity was especially useful to validate the rigor of our EBSE application and to have an objective view of the risks that were still present in the project.

VI. Discussion

This section revisits our main findings on the barriers and facilitators that we identified when applying EBSE. We also compare our results with the early concerns of EBSE proponents. Subsequently, we reflect on the implications our results have on the fit-to-purpose of EBSE. Finally, we present our assessment of the limitations of our work.

A. Barriers to EBSE application

The barriers that arose in our study could be grouped into: Difficulty in obtaining relevant evidence, Complexity of applying evidence, and Lack of guidelines and examples of using EBSE in industry.

Difficulty in obtaining relevant evidence. One of the challenges we faced was finding evidence able to address the company's problems. There were no evidence-based answers to the company's specific KM problems, and evidence related to more general issues was hard to find. This meant we had to use a rather broad, high-level research question. Lack of relevant evidence has been reported in previous studies [14], [58]. The major disadvantage of this approach is that the answers may be less specific than required.

The limitations of SE evidence affect research reliability: (1) Lack of empirical studies forces reliance on less reliable sources, (2) Absence of clear recommendations may require restructuring study results, and (3) Lack of standardized approaches may complicate aggregating results. Point 1 weakens the strength of evidence, while points 2 and 3 risk biases because other researchers, with less detailed knowledge of the primary study details, may misunderstand or misrepresent the conclusions that can be validly drawn from the evidence.

Researchers have proposed a variety of methods to address the limitations of SE evidence, i.e. using evidence considering its nature and quality (e.g., for descriptive and correlational uses and not to support claims of causality) [59], researching issues relevant to the industry [8], [60], collaborating closely with practitioners [60], including recommendations for practice [58], conducting more interdisciplinary research [61], including grey literature [62]–[65], and appropriately considering research context [66], [67].

Complexity of applying evidence. Another challenge was the complexity of applying the evidence we found. This was mainly because the evidence was in the form of high-level recommendations not directly implementable process guidelines. The nature of the evidence was a particular problem because our requesters did not themselves have any SE research experience. Thus, we needed to summarize the collected research in a way that it could be understood by practitioners.

This issue has not been raised by any of the other collaborative EBSE-related studies that we are aware of (i.e., [11]–[13], [15], [21]), because, in all those cases, the requesters either had SE research experience or specialist knowledge of the topic of interest.

Lack of guidelines and examples of using EBSE in industry. In contrast to SRs, there are no detailed guidelines for the five EBSE steps, and there are only a very limited number of examples of using secondary studies in an industry context.

First, the EBSE team had no previous experience working with EBSE and RRs in industry. Our professional experience and our strong link with the company helped us to establish a good working relationship with our requesters. However, our EBSE activities might have been both more focused and more efficient if there had been some practical SE guidelines to support our EBSE process.

As it was, since the EBSE team chose to use an RR (rather than a full SR), we followed, as far as possible, the process reported by Cartaxo et al. [13]. In particular, we prepared an evidence briefing and discussed the RR results with the requesters. However, in general, we acted in response to the project circumstances, rather than following prepared plans. For example,

- In order to define a research question we had to do several preliminary searches.
- We discussed our initial evidence with the requests, adjusting our search, selection criteria, and aggregation process to address their concerns.
- We developed a method to assist the requesters to rank the RR recommendations which took place when we presented the RR results.

Steps 3 and 5 of EBSE posed particular problems. Step 3 requires a critical assessment of the available evidence, but following Cartaxo et al.'s guidelines for RRs, we omitted any formal assessment of the identified primary studies. In practice, since we needed both to include weak forms of evidence from the primary studies (such as lessons learned), and to derive recommendations ourselves, the most important issues for the requesters were information about the type of companies generating the evidence and of the methods used to do so. It is also important to recognize that failure to assess evidence in terms of its validity, impact, and usefulness is a major deviation from the original specification of evidence-based practice defined in medical practice

Step 5 calls for reflection on the EBSE activity, but we found no examples of studies discussing the use of an SR as part of an industry-academia collaboration that reported any evaluation of the EBSE process itself. In practice, we reflected on how we performed each step, in terms of the issues that arose, how we addressed those issues, any ways in which we might have improved our performance, and the requesters' views of the EBSE project. Also, due to the closure of the company, we were unable to confirm that the recommendations were adopted and delivered the expected benefits.

Taken together, these issues seem to imply that the analogy between evidence-based health care and evidence-based software engineering is somewhat problematic. Perhaps in the case of healthcare, the steps of evidence-based practice are more straightforward, so a more detailed process is irrelevant. In SE it appears that we need more guidance to cope with broad

questions, reporting SE evidence to practitioners, developing practical context-appropriate recommendations, and evaluating EBSE activities.

Practical problems when working with an industry partner. As other researchers have pointed out, academic interactions with industry partners are difficult (see, e.g., [68]). In our case, the company placed constraints on the collaboration (i.e., defining a broad problem with several different aspects, and requesting evidence that would map to small incremental changes) that interfered with academic goals and was further complicated by company circumstances (e.g., the company changes and closing).

B. Facilitators for EBSE application

Several factors supported our EBSE project.

Close collaboration with industry partner & Considering their needs. The close collaboration with the requesters was a decisive factor in achieving our goals. This included activities in which their participation was explicit, e.g., meeting to understand their problems, validating our understanding with them later, also validating the evidence early, and jointly participating in the workshop to disseminate the results. But it also required an explicit effort on our part to consider their needs, e.g., in considering how they would apply the evidence. Even so, we did not detect that one of the requesters wanted results in less time.

Appropriate dissemination. In accordance with Cartaxo et al.'s recommendation to incorporate result discussions as part of dissemination activities [13], we not only created an evidence briefing report summarizing the findings of the RR, but we also conducted a workshop where participants engaged in discussions about the evidence and its application. This combination proved effective in disseminating our results. It seems that a one-page evidence briefing report alone is insufficient to ensure the practical use of the results. Based on our experience, it may be helpful to include the following in future reports: (1) practical implementation guidance for each recommendation, (2) dependencies between recommendations, and (3) indications of the strength of evidence supporting each recommendation.

Process improvement commitment. The company's commitment to process improvement was of great help in its willingness to reflect on its processes and to receive recommendations in a positive way. It is important to point out that, EBSE is not an alternative to process improvement but a framework that can be used to support process improvement. In particular, as suggested by Dybå et al. [17] and confirmed by our study), EBSE seems a useful method to find candidate solutions to support process improvement.

Conduct of an RR & Participation of external researcher. Two aspects that helped us achieve adequate methodological rigor were using an RR and the validation of experts. Using an RR allowed us to obtain interesting results for the company within the available low-resource setting (i.e., limited effort and use by non-experts) and without losing scientific rigor. Validation by experts helped two stages of the study, during the early stages of the RR, expert validation gave us confidence in our RR process, and during the analysis and reporting of results, the participation of an EBSE expert helped substantially improved the quality of this study.

C. Revisting EBSE after 20 years

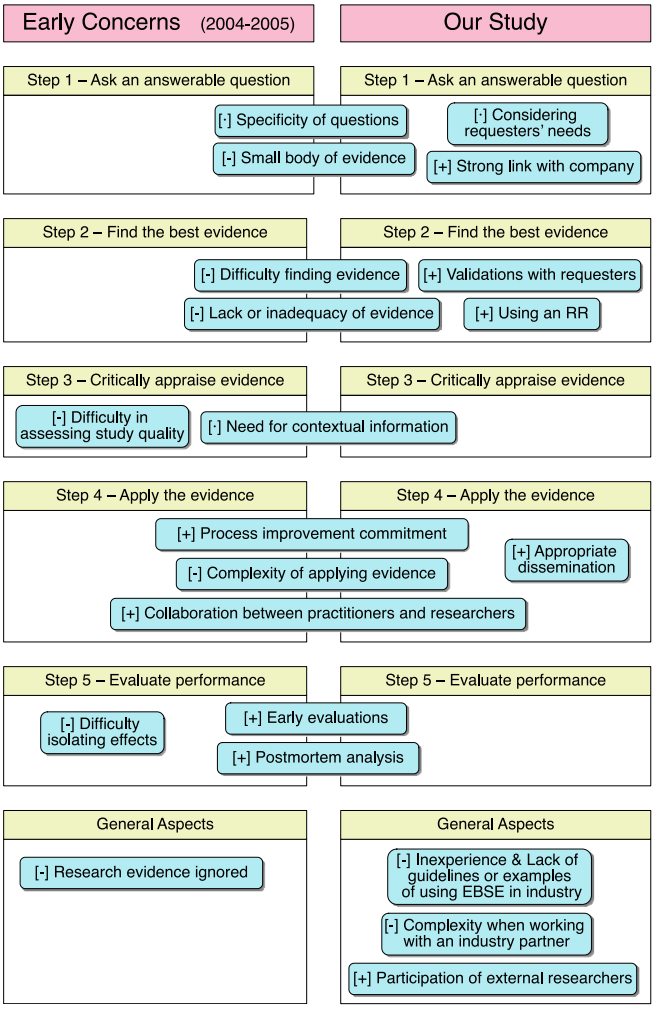


Fig. 3. Comparison of the early concerns identified by EBSE proponents and the results of our study. Notation: [·] indicates main issues, [–] represents barriers, and [+] denotes facilitators.

More than half of the barriers we identified in our use of EBSE had already been identified 20 years ago (see Figure 3). While certain aspects may have improved, there remain persistent barriers that continue to impact EBSE applications, despite their long-standing recognition. However, the major challenge, which has been identified in other studies as well, remains the lack or inadequacy of evidence. This represents a significant challenge to face when adopting EBSE. In addition, collaboration between academia and industry remains a challenge for the entire community. Thus, it is clear the EBSE has not achieved its goal of bridging the gap between industry and academia.

Our study has not exhibited all of the concerns raised by proponents of EBSE, but it is important to note that we present a specific case, that has some aspects that were not

anticipated by the EBSE proponents (i.e., the use of an RR and the lack of critical appraisal of evidence[10]). In addition, we were unable to fully evaluate the effect of implementing the recommendations, so we cannot make conclusions about the difficulty of isolating the effects of specific process changes. However, it is noteworthy that although the practitioners were not used to employing scientific research results, this did not cause a barrier to our collaboration.

Certainly, good progress has been made on some of the challenges identified 20 years ago. SRs are highly appreciated by the academic community (see, e.g., the tertiary study conducted by Kamei et al. [2]). There are standards or guidelines for several types of empirical methods (e.g., [27], [69]), including SRs. Much work has been done to improve evidence aggregation methods and their reports (see, e.g., [4], [13], [70]–[73]). We do not yet have a central repository of published SRs. However, for some time now it has been possible to pre-register reports in several venues [74]. Finally, there are also initiatives to generate more appropriate evidence for non-academic stakeholders (e.g. the ESEM's Industry, Government, and Community track[11]).

Both current research and our results indicate that EBSE has yet to fulfill its objective of effectively supporting practitioners in utilizing research evidence. It seems that proponents of EBSE underestimated the disparities between healthcare and computing, leading to challenges in its implementation:

- Healthcare options are generally more easily understood and adopted by practitioners without the need for extensive explanation, unlike SE methods. For instance, doctors prescribing a new drug, or nurses adjusting patient care practices such as coma patient turning frequency or hand-washing protocols.
- Healthcare options undergo rigorous evaluation before deployment. This is mainly due to national and international regulation, and ensures that a body of empirical studies are available as a basis for evidence-informed recommendations. However, in the computing industry, companies often deploy new methods without waiting for independent validation, or any proper understanding of their risks. There are many examples where the adoption of new computing methods to address specific problems has introduced new problems. For example, object-oriented design was supposed to ensure more reliable systems because keeping data and code together would make testing more effective. However, the emphasis on developing self-standing, independent objects led to extensive use of code clones which caused more maintenance problems and, in turn, necessitated the development of aspect-oriented design [75].

Furthermore, for certain practitioners, distinguishing between empirical evidence and their personal experience or the claims of thought leaders may not be inherently clear. For example, the practitioners who collaborated in our study learned about the concept of empirical evidence from our collaboration. Decision-making processes frequently rely on guidance from these thought leaders or consulting firms, even when disseminating questionable claims (as evidenced by the recent McKinsey case on developer productivity, refer to [76], [77]). Thus, the sole publication of a paper does not guarantee effective communication of SE evidence to practitioners. This certainly could have been another influential factor in EBSE adoption.

D. Is EBSE framework fit-for-purpose?

EBSE can be challenging for practitioners without research experience, but our study confirms that it can be effectively utilized in industry-academia collaboration. EBSE application raises several challenges and requires different skills that include academic knowledge, and professional experience. Successfully using EBSE requires the ability to convert academic recommendations into actionable process changes and the proficiency to design and utilize effective engagement mechanisms for the dissemination of knowledge.

Evidence is still somewhat limited, and our study highlights that practitioners value evidence that aligns with their practical concerns. This implies that academics should consider the needs of the industry when pursuing their research agendas. Likewise, practitioners should actively engage with academics to identify topics they consider essential and to facilitate evaluations of industry practices and trials of new technologies.

Step 1. The lack of evidence remains a significant concern, indicating that SE should adopt relatively high-level research questions. In particular, keyword searches specifying practitioner context can restrict access to potentially useful information.

Steps 2, 3 & 4. We consider these steps need to be better understood in the SE context. First, it is unclear what is meant by "finding best evidence" in Step 2, given that Step 3 of EBSE promotes the assessment of evidence in terms of validity, impact, and applicability. We suggest interpreting this as evidence that best fits the research question. Second, although EBSE separates Steps 3 and 4, in our study, we addressed Step 3 and the initial planning for Step 4 together when prioritizing recommendations. This integration of the two steps provided a means to link the selection of evidence-based process change recommendations with the process improvement capability of the company. In other words, evidence assessment can help define the required process improvement activities. Third, if we accept the need for broad questions, we should expect Step 4 to deal with prioritizing available process change options and defining how to implement and monitor the selected process changes. Our study confirms that existing procedures for managing process change mean practitioners may be prepared for change but may place restrictions on the type of change they can manage. Fourth, in terms of Step 3, better standards for empirical methods would improve the likelihood of obtaining good-quality evidence, but it would require research expertise to assess specific evidence effectively.

Steps 4 & 5. The process changes planned in Step 4 should include procedures for monitoring the process change. Our

[10]We evaluated evidence in terms of its relevance to practice (i.e., applicability).

[11]https://conf.researchr.org/track/esem-2024/esem-2024-industry-government-community

experience suggests that monitoring the process change, and therefore evaluating EBSE, can also be a complex task, and about which we have the following reflections.

Comparing EBSE. When evaluating the EBSE process, it is important to consider what to compare it against. Two possible alternatives are: (1) comparing with expert opinion via consultancy or social media posts, (2) comparing with software process improvement models such as the Capability Maturity Model. In addition, in a specific evaluation study, it is useful to determine if the organization has defined procedures to manage process change. In our study, the EBSE team students had no experience in EBSE, process improvement methods, or evaluation. As the company was taken over, this was not a problem. However, any future EBSE evaluation should consider the final EBSE step in more detail. Researchers doing this should have some experience in undertaking field evaluations of process change. Also, such an evaluation might require monitoring the process over several weeks or months to decide if the recommended changes were fully adopted and addressed the diagnosed problems.

Evaluating the use of EBSE or the value of the collected evidence. Is it possible to evaluate EBSE by applying it to a SE problem, or are we evaluating the value of the collected evidence? From practitioners' viewpoint, it does not seem easy to isolate the usefulness of EBSE from the benefit of the collected evidence. In this study, we tried to make practitioners aware of EBSE and its methods throughout the EBSE project. We also had a final meeting in which we asked questions not only about the KM recommendations, but also about the evidence-based approach, in order to encourage their reflection on the two topics separately. However, we need more EBSE evaluation studies to be able to identify cross-case results or observations that are needed for a more objective meta-evaluation of EBSE.

E. Limitations

An important limitation of our study was the fact that the company closed down, so we were unable to confirm whether or not the process changes that the company tried to introduce were successful. We can only confirm that the company identified recommendations that they planned to adopt at our dissemination workshop and the requesters reported that some changes had been adopted before the company closed. We must note that in any case, our approach to monitoring the changes would still have been indirect, since at that point, none of the members of the EBSE team worked for the company, and we had not planned any direct monitoring of the change process.

We also identified two issues related to conflicts of interest. Firstly, the affiliation of one of the students with the company and the knowledge among requesters and company members that the research was part of their capstone project raised concerns about potential positive bias in their feedback. To mitigate this, we consistently emphasized to participants the equal importance of positive and negative results. We gathered their perceptions of the research results through multiple methods and on different occasions. Secondly, the first author sought to complete his research on the adoption of EBSE (that is part of his PhD), which could be thought of as more valuable with positive results. To minimize the risk of bias in favor of positive results, the first author diligently recorded decisions and actions in a detailed journal. They also reported and sought input from the second and third authors at various stages of the study, including planning, conduct, analysis, and reporting. These measures were implemented to ensure transparency and reduce potential bias.

Although students had prior training and collaboration with Pizard, this marked their initial involvement in conducting a secondary study. To mitigate possible deviations from the RR methodology, we strictly adhered to Cartaxo et al. recommendations [13], including protocol development beforehand, while seeking guidance from Vallespir. Moreover, we validated the RR protocol and the conduct of initial stages with Fernando Acerenza, a researcher knowledgeable in EBSE.

The selection of the topic to work on with the company introduced some limitations. The primary limitation is our focus on non-critical issues to avoid disclosing sensitive information and not expecting to address serious or urgent problems. This approach may restrict the breadth and potential impact of the research findings. Additionally, it may seem that the topic addressed (i.e., KM) and the company's problems were more aligned with management than with SE. However, there are diverse and recognized studies on KM in SE indicating the importance of this topic in our field (e.g., [78]–[80]). In our case, out of the 20 selected studies, 13 were from IT and SE venues, also showing this topic's relevance to SE. Even so, dealing with a topic close to management and not a classic one within SE can be considered a limitation.

VII. Final Remarks

Currently, EBSE seems more appropriate as an approach for researchers to enhance collaboration with practitioners, rather than a freely accessible mechanism for all stakeholders as initially expected two decades ago. However, research on practice-driven use of EBSE is not so well explored, presenting certain challenges. Our study's findings contribute to addressing this gap, and although it is a single application in a specific context, we venture to present the recommendations outlined in Table V.

We are moderately optimistic that EBSE will be more widely adopted in the future. Successful experiences of applying EBSE will be the greatest attraction to motivate its use where success is closely linked to overcoming the identified barriers, especially the lack or inadequacy of evidence. However, if we want to confirm if the EBSE framework is valuable, or whether it needs to be revised, reports on practice-driven applications of EBSE are essential.

VIII. Acknowledgments

We thank Forrest Shull for telling us about the Voice of Evidence column from IEEE Software magazine and Magne Jørgensen for mentioning the EDOS newsletter. We thank all the members of the company with whom we worked and the middle and senior management who allowed the

TABLE V
RECOMMENDATIONS FOR CONDUCTING PRACTICE-DRIVEN APPLICATIONS OF EBSE.

1. Consider all EBSE steps and not just conduct an SR. Taking a broader view of evidence utilization enables consideration of various aspects, such as leveraging previously published SRs on the topic or more accurately evaluating the application of the approach. Do not dismiss the assessment of studies (Step 3) before determining whether such information is necessary.
2. Initiate early and sustained engagement with practitioners to grasp their needs and expectations about evidence, fostering collaborative decision-making throughout the process. Intermediate validations, such as those of evidence discovered, appear to be crucial for achieving successful final results.
3. Consider practitioners' values (e.g., commitment to process improvement) and constraints on required evidence (e.g., openness to recommendations for large or small process changes, or timeframes required to obtain results).
4. Anticipate the limitations of evidence by employing broad questions, and be prepared to restructure recommendations found in primary studies. For instance, converting these recommendations into actionable process changes can enhance their practical utility.
5. Facilitate the reception of evidence by considering practitioners' expectations regarding knowledge transfer and incorporate face-to-face activities where they can discuss recommendations and implementation strategies.
6. Effectively evaluate evidence utilization by monitoring process changes and preemptively considering alternatives for comparison, such as expert opinions or performance metrics before implementing changes.
7. Strive to adhere to the EBSE process and uphold scientific rigor, including conducting validations with expert researchers, to mitigate potential researcher bias introduced by current evidence limitations.
8. Anticipate challenges that may arise in academia-industry collaborations. Specifically, researchers should maintain flexibility and consider recommendations from the literature (e.g.,[81]).

conduct of our study. We also thank Fernando Acerenza for his validation of our progress during the RR conduct. We thank Joaquín Lezama and Rodrigo García for their participation in the EBSE-based project team, whose contribution is better reported in the paper about the RR conduct [30]. Finally, we thank the ICSE 2024 anonymous reviewers whose comments and suggestions improved this report.

REFERENCES

[1] B. Kitchenham, T. Dybå, and M. Jørgensen, "Evidence-based software engineering," in *International Conference on Software Engineering (ICSE)*, 2004, pp. 273–281.

[2] F. Kamei, I. Wiese, C. Lima, I. Polato, V. Nepomuceno, W. Ferreira, M. Ribeiro, C. Pena, B. Cartaxo, G. Pinto, and S. Soares, "Grey Literature in Software Engineering: A critical review," *Information and Software Technology*, 2021.

[3] S. Beecham, D. Bowes, and K.-J. Stol, "Introduction to the ease 2016 special section: Evidence-based software engineering: Past, present, and future," *Information and Software Technology*, vol. 89, pp. 14–18, 2017.

[4] B. Cartaxo, G. Pinto, E. Vieira, and S. Soares, "Evidence briefings: Towards a medium to transfer knowledge from systematic reviews to practitioners," in *International Symposium on Empirical Software Engineering and Measurement*. New York, NY, USA: Association for Computing Machinery, 2016, pp. 57:1–57:10.

[5] B. Kitchenham, D. Budgen, and P. Brereton, *Evidence-based software engineering and systematic reviews*, ser. Chapman & Hall/CRC Innovations in Software Engineering and Software Development Series. London: CRC Press, 2015.

[6] B. Cartaxo, G. Pinto, and S. Soares, *Rapid Reviews in Software Engineering*. Cham: Springer International Publishing, 2020, pp. 357–384.

[7] E. Hassler, J. Carver, N. Kraft, and D. Hale, "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*. New York, NY, USA: Association for Computing Machinery, 2014, pp. 31:1–31:10.

[8] F. Da Silva, A. Santos, S. Soares, A. Frana, C. Monteiro, and F. Maciel, "Six years of systematic literature reviews in software engineering: An updated tertiary study," *Information and Software Technology*, vol. 53, no. 9, pp. 899–913, 2011.

[9] R. Santos and F. Da Silva, "Motivation to perform systematic reviews and their impact on software engineering practice," in *International Symposium on Empirical Software Engineering and Measurement*, 2013, pp. 292–295.

[10] T. B. Callo Arias, P. van der Spek, and P. Avgeriou, "A practice-driven systematic review of dependency analysis solutions," *Empirical Software Engineering*, 2011.

[11] M. Kalenda, P. Hyna, and B. Rossi, "Scaling agile in large organizations: Practices, challenges, and success factors," *Journal of Software: Evolution and Process*, 2018.

[12] J. M. Sierra, A. Vizcaíno, M. Genero, and M. Piattini, "A systematic mapping study about socio-technical congruence," *Information and Software Technology*, 2018.

[13] B. Cartaxo, G. Pinto, and S. Soares, "The role of rapid reviews in supporting decision-making in software engineering practice," in *International Conference on Evaluation and Assessment in Software Engineering*, 2018.

[14] Q. Song, M. Borg, E. Engström, H. Ardö, and S. Rico, "Exploring ml testing in practice: Lessons learned from an interactive rapid review with axis communications," in *International Conference on AI Engineering*, 2022.

[15] E. Bjarnason, P. Åberg, and N. bin Ali, "Software selection in large-scale software engineering: A model and criteria based on interactive rapid reviews," *Empirical Software Engineering*, 2023.

[16] C. Hamel, A. Michaud, M. Thuku, B. Skidmore, A. Stevens, B. Nussbaumer-Streit, and C. Garritty, "Defining rapid reviews: a systematic scoping review and thematic analysis of definitions and defining characteristics of rapid reviews," *Journal of Clinical Epidemiology*, vol. 129, pp. 74–85, 2021.

[17] T. Dyba, B. Kitchenham, and M. Jorgensen, "Evidence-based software engineering for practitioners," *IEEE Software*, vol. 22, no. 1, pp. 58–65, 2005.

[18] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[19] F. Shull, T. Dyba, H. Sharp, and R. Prikladnicki, "Voice of Evidence: A Look Back," *IEEE Software*, vol. 34, no. 4, pp. 23–25, 2017.

[20] A. Kasoju, K. Petersen, and M. Mäntylä, "Analyzing an automotive testing process with evidence-based software engineering," *Information and Software Technology*, 2013.

[21] T. Lewowski and L. Madeyski, *Code Smells Detection Using Artificial Intelligence Techniques: A Business-Driven Systematic Review*. Springer International Publishing, 2022.

[22] N. Paternoster, C. Giardino, M. Unterkalmsteiner, T. Gorschek, and P. Abrahamsson, "Software development in startup companies: A systematic mapping study," *Information and Software Technology*, vol. 56, no. 10, pp. 1200–1218, 2014.

[23] D. Šmite, C. Wohlin, Z. Galviņa, and R. Prikladnicki, "An empirically based terminology and taxonomy for global software engineering," *Empirical Software Engineering*, vol. 19, no. 1, pp. 105–153, 2014.

[24] B. Kitchenham, P. Brereton, Z. Li, D. Budgen, and A. Burn, "Repeatability of systematic literature reviews," in *15th Annual Conference on Evaluation & Assessment in Software Engineering (EASE 2011)*, 2011, pp. 46–55.

[25] M. S. Schwartz and C. G. Schwartz, "Problems in Participant Observation," *American Journal of Sociology*, vol. 60, no. 4, pp. 343–353, 1955.

[26] A. J. Ko, "A three-year participant observation of software startup software evolution," in *International Conference on Software Engineering: Software Engineering in Practice Track (ICSE-SEIP)*, 2017, pp. 3–12.

[27] P. Runeson, M. Host, A. Rainer, and B. Regnell, *Case Study Research in Software Engineering: Guidelines and Examples*, 1st ed. Wiley Publishing, 2012.

[28] B. C. O'Brien, I. B. Harris, T. J. Beckman, D. A. Reed, and D. A. Cook, "Standards for reporting qualitative research: a synthesis of recommendations," *Academic Medicine*, vol. 89, no. 9, pp. 1245–1251, 2014.

[29] S. J. Tracy, "Qualitative quality: Eight "big-tent" criteria for excellent qualitative research," *Qualitative Inquiry*, vol. 16, no. 10, pp. 837–851, 2010.

[30] S. Pizard, J. Acerenza, Lezama, R. García, D. Vallespir, and B. Kitchenham, "Using RR to support SE practice: A citation analysis and a replication study," *Submitted to Empirical Software Engineering*, 2024.

[31] S. Pizard, F. Acerenza, X. Otegui, S. Moreno, D. Vallespir, and B. Kitchenham, "Training students in evidence-based software engineering and systematic reviews: a systematic review and empirical study," *Empirical Software Engineering*, vol. 26, no. 3, p. 50, 2021.

[32] S. Pizard, D. Vallespir, and B. Kitchenham, "A longitudinal case study on the effects of an evidence-based software engineering training," in *International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET)*, 2022, pp. 1–13.

[33] S. Saito, Y. Iimura, A. K. Massey, and A. I. Antón, "Discovering undocumented knowledge through visualization of agile software development activities," *Requirements Engineering*, vol. 23, no. 3, pp. 381–399, Sep. 2018.

[34] T. S. Mendes, M. A. de F. Farias, M. Mendonça, H. F. Soares, M. Kalinowski, and R. O. Spínola, "Impacts of agile requirements documentation debt on software projects: A retrospective study," in *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, ser. SAC '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 1290–1295. [Online]. Available: https://doi.org/10.1145/2851613.2851761

[35] E. T. Stringer, *Action research third edition*, 3rd ed. Sage Publications, Inc., 2007.

[36] A. Birk and T. Dingsøyr, "Trends in learning software organizations: Current needs and future solutions," in *Third Biennial Conference, WM 2005 - Professional Knowledge Management*, K.-D. Althoff, A. Dengel, R. Bergmann, M. Nick, and T. Roth-Berghofer, Eds., 2005, pp. 70–75.

[37] D. Viana, J. Rabelo, T. Conte, A. Vieira, E. Barroso, and M. Dib, "A qualitative study about the life cycle of lessons learned," in *International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE)*, 2013, pp. 73–76.

[38] K. Yglesias, "Ibm's reuse programs: knowledge management and software reuse," in *International Conference on Software Reuse*, 1998, pp. 156–165.

[39] G. Matturro and A. Silva, "Rebec: A method for capturing experience during software development projects," in *International Conference on Knowledge Engineering and Knowledge Management - Knowledge Engineering and Management by the Masses*, P. Cimiano and H. S. Pinto, Eds., 2010, pp. 524–533.

[40] D. Viana, T. Conte, S. Marczak, R. Ferreira, and C. de Souza, "Knowledge creation and loss within a software organization: An exploratory case study," in *Hawaii International Conference on System Sciences*, 2015, pp. 3980–3989.

[41] S. Vasanthapriyan, J. Xiang, J. Tian, and S. Xiong, "Knowledge synthesis in software industries: a survey in sri lanka," *Knowledge Management Research & Practice*, vol. 15, no. 3, pp. 413–430, 2017.

[42] M. L. I. Gervigny and S. D. Nagowah, "Knowledge sharing for agile distributed teams: A case study of mauritius," in *International Conference on Infocom Technologies and Unmanned Systems - Trends and Future Directions (ICTUS)*, 2017, pp. 413–419.

[43] J. Soini, T. Makinen, and V. Tenhunen, "Managing and processing knowledge sharing between software organizations: A case study," in *Portland International Conference on Management of Engineering & Technology (PICMET)*, 2007, pp. 1108–1113.

[44] V. Santos, A. Goldman, H. Filho, D. Martins, and M. Cortés, "The influence of organizational factors on inter-team knowledge sharing effectiveness in agile environments," in *Hawaii International Conference on System Sciences*, 2014, pp. 4729–4738.

[45] D. Smite and T. Dingsøyr, "Fostering cross-site coordination through awareness: An investigation of state-of-the-practice through a focus group study," in *Euromicro Conference on Software Engineering and Advanced Applications*, 2012, pp. 337–344.

[46] H. Khan, A. Ahmad, and M. Alnuem, "Knowledge management: A solution to requirements understanding in global software engineering," *Research Journal of Applied Sciences, Engineering, and Technology*, vol. 4, pp. 2087–2099, 07 2012.

[47] A. Kammani, S. Aljahdali, and H. Date, "Km capability for software development: A case study of the indian software firms," *International Journal of Business Information Systems*, vol. 12, no. 1, p. 44–67, dec 2013.

[48] M. Pulkkinen., O. Mazhelis., P. Marttiin., and J. Meriluoto., "Support for knowledge and innovations in software development: Community within company: Inner source environment," in *International Conference on Web Information Systems and Technologies*, 2007, pp. 141–150.

[49] J. Soini, "Managing information and distributing knowledge in a knowledge-intensive business environment," in *Portland International Conference on Management of Engineering & Technology (PICMET)*, 2008, pp. 889–894.

[50] A. Heredia, R. Colomo-Palacios, and P. Soto-Acosta, "Tool-supported continuous business process innovation: a case study in globally distributed software teams," *European Journal of International Management*, vol. 11, no. 4, pp. 388–406, 2017.

[51] D. Šmite, N. B. Moe, A. Šāblis, and C. Wohlin, "Software teams and their knowledge networks in large-scale software development," *Information and Software Technology*, vol. 86, pp. 71–86, 2017.

[52] A. Heredia, J. Garcia-Guzman, A. Amescua-Seco, and A. Serrano, "Study of factors influencing the adoption of agile processes when using wikis," *International Journal of Software Engineering and Knowledge Engineering*, vol. 24, no. 06, pp. 859–885, 2014.

[53] M. Humayun, C. Gang, and I. Masood, "An empirical study on investigating the role of kms in promoting trust within gsd teams," in *International Conference on Evaluation and Assessment in Software Engineering (EASE)*, 2013, p. 207–211.

[54] J. L. Jurado, A. Fernandez, and C. A. Collazos, "Applying gamification in the context of knowledge management," in *International Conference on Knowledge Technologies and Data-Driven Business*, 2015.

[55] M. Milovanović, M. Minović, V. Štavljanin, M. Savković, and D. Starčević, "Wiki as a corporate learning tool: case study for software development company," *Behaviour & Information Technology*, vol. 31, no. 8, pp. 767–777, 2012.

[56] H. Chen, M. Baptista Nunes, G. Ragsdell, and X. An, "Extrinsic and intrinsic motivation for experience grounded tacit knowledge sharing in Chinese software organisations," *Journal of Knowledge Management*, vol. 22, no. 2, pp. 478–498, 2018.

[57] J. DeFranco and P. Laplante, "A content analysis process for qualitative software engineering research," *Innovations in Systems and Software Engineering*, vol. 13, no. 2-3, pp. 129–141, 2017.

[58] D. Budgen, P. Brereton, N. Williams, and S. Drummond, "What support do systematic reviews provide for evidence-informed teaching about software engineering practice?" *E-Informatica Software Engineering Journal*, vol. 14, no. 1, pp. 7–60, 2020.

[59] S. Pizard, F. Acerenza, D. Vallespir, and B. Kitchenham, "Assessing attitudes towards evidence-based software engineering in a government agency," *Information and Software Technology*, p. 107101, 2022.

[60] V. Garousi, M. Borg, and M. Oivo, "Practical relevance of software engineering research: synthesizing the community's voice," *Empirical Software Engineering*, vol. 25, no. 3, pp. 1687–1754, 2020. [Online]. Available: https://doi.org/10.1007/s10664-020-09803-0

[61] D. Méndez Fernández and J. H. Passoth, "Empirical software engineering: From discipline to interdiscipline," *Journal of Systems and Software*, vol. 148, pp. 170–179, 2019.

[62] B. A. Kitchenham, L. Madeyski, and D. Budgen, "How should software engineering secondary studies include grey material," *IEEE Transactions on Software Engineering*, pp. 1–1, 2022.

[63] V. Garousi, M. Felderer, and M. V. Mäntylä, "Guidelines for including grey literature and conducting multivocal literature reviews in software engineering," *Information and Software Technology*, vol. 106, pp. 101–121, 2019.

[64] A. Rainer and A. Williams, "Using blog-like documents to investigate software practice: Benefits, challenges, and research directions," *Journal of Software: Evolution and Process*, vol. 31, no. 11, p. e2197, 2019, e2197 smr.2197.

[65] A. Williams and J. Buchan, "Using the case survey methodology for finding high-quality grey literature in software engineering," in *The International Conference on Evaluation and Assessment in Software Engineering 2022*, ser. EASE 2022. New York, NY, USA: Association for Computing Machinery, 2022, p. 1–9.

[66] L. Briand, D. Bianculli, S. Nejati, F. Pastore, and M. Sabetzadeh, "The case for context-driven software engineering research: Generalizability is overrated," *IEEE Software*, vol. 34, no. 5, pp. 72–75, 2017.

[67] C. Wohlin and A. Rainer, "Challenges and recommendations to publishing and using credible evidence in software engineering," *Information and Software Technology*, p. 106555, feb 2021.

[68] V. Garousi, D. C. Shepherd, and K. Herkiloglu, "Successful Engagement of Practitioners and Software Engineering Researchers: Evidence

From 26 International Industry-Academia Collaborative Projects," *IEEE Software*, vol. 7459, no. c, pp. 1–13, 2019.

[69] J. Kontio, J. Bragge, and L. Lehtola, "The focus group method as an empirical tool in software engineering," *Guide to Advanced Empirical Software Engineering*, no. 2004, pp. 93–116, 2008.

[70] V. Garousi and M. Felderer, "Experience-based guidelines for effective and efficient data extraction in systematic reviews in software engineering," in *Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering*, ser. EASE '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 170–179. [Online]. Available: https://doi.org/10.1145/3084226.3084238

[71] B. Kitchenham and P. Brereton, "A systematic review of systematic review process research in software engineering," *Information and Software Technology*, vol. 55, no. 12, pp. 2049–2075, 2013.

[72] P. Avgeriou, M. Verbeek, S. Bibi, A. Chatzigeorgiou, and A. Ampatzoglou, "Identifying, categorizing and mitigating threats to validity in software engineering secondary studies," *Information and Software Technology*, vol. 106, no. February 2018, pp. 201–230, 2018.

[73] B. A. Kitchenham, L. Madeyski, and D. Budgen, "Segress: Software engineering guidelines for reporting secondary studies," *IEEE Transactions on Software Engineering*, pp. 1–1, 2022.

[74] N. A. Ernst and M. T. Baldassarre, "Registered reports in software engineering," *Empirical Software Engineering*, vol. 28, no. 2, p. 55, 2023. [Online]. Available: https://doi.org/10.1007/s10664-022-10277-5

[75] N. Saini, S. Singh, and Suman, "Code clones: Detection and management," *Procedia Computer Science*, vol. 132, pp. 718–727, 2018, international Conference on Computational Intelligence and Data Science.

[76] C. Gnanasambandam, M. Harrysson, A. Hussin, J. Keovichit, and S. Srivastava, "Yes, you can measure software developer productivity," *McKinsey's Digital and Technology, Media & Telecommunications Practices*, no. August, 2023. [Online]. Available: https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/yes-you-can-measure-software-developer-productivity#/

[77] D. North, "Mckinsey developer productivity review," Oct 2023. [Online]. Available: https://dannorth.net/mckinsey-review/

[78] F. O. Bjørnson and T. Dingsøyr, "Knowledge management in software engineering: A systematic review of studied concepts, findings and research methods used," *Information and Software Technology*, vol. 50, no. 11, pp. 1055–1068, 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950584908000487

[79] I. Rus and M. Lindvall, "Guest editors' introduction: Knowledge management in software engineering," *IEEE Software*, vol. 19, no. 03, pp. 26–38, may 2002.

[80] R. Ouriques, K. Wnuk, T. Gorschek, and R. B. Svensson, "The role of knowledge-based resources in agile software development contexts," *Journal of Systems and Software*, vol. 197, p. 111572, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0164121222002485

[81] V. Garousi, K. Petersen, and B. Ozkan, "Challenges and best practices in industry-academia collaborations in software engineering: A systematic literature review," *Information and Software Technology*, vol. 79, no. September 2014, pp. 106–127, 2016.