Experiences from conducting rapid reviews in collaboration with practitioners — Two industrial cases

Sergio Rico a,∗, Nauman Bin Ali b, Emelie Engström a, Martin Höst a

a Lund University, Department of Computer Science, SE-221 00, Lund, Sweden
b Blekinge Institute of Technology, Department of Software Engineering, SE-371 79, Karlskrona, Sweden

A B S T R A C T

Context: Evidence-based software engineering (EBSE) aims to improve research utilization in practice. It relies on systematic methods to identify, appraise, and synthesize existing research findings to answer questions of interest for practice. However, the lack of practitioners’ involvement in these studies’ design, execution, and reporting indicates a lack of appreciation for the need for knowledge exchange between researchers and practitioners. The resultant systematic literature studies often lack relevance for practice.

Objective: This paper explores the use of Rapid Reviews (RRs), in fostering knowledge exchange between academia and industry. Through the lens of two case studies, we delve into the practical application and experience of conducting RRs.

Methods: We analyzed the conduct of two rapid reviews by two different groups of researchers and practitioners. We collected data through interviews, and the documents produced during the review (like review protocols, search results, and presentations). The interviews were analyzed using thematic analysis.

Results: We report how the two groups of researchers and practitioners performed the rapid reviews. We observed some benefits, like promoting dialogue and paving the way for future collaborations. We also found that practitioners entrusted the researchers to develop and follow a rigorous approach and were more interested in the applicability of the findings in their context. The problems investigated in these two cases were relevant but not the most immediate ones. Therefore, rapidness was not a priority for the practitioners.

Conclusion: The study illustrates that rapid reviews can support researcher-practitioner communication and industry-academia collaboration. Furthermore, the recommendations based on the experiences from the two cases complement the detailed guidelines researchers and practitioners may follow to increase interaction and knowledge exchange.

1. Introduction

As an applied research area, software engineering research relies on a deep understanding of industrial software engineering practices to produce relevant and applicable knowledge. Without such understanding, researchers risk focusing on irrelevant aspects of existing problems [1,2], missing necessary information [3], providing solutions that do not apply nor are generalizable to other contexts, or presenting results in a complicated way that is difficult for practitioners to access, interpret and implement [4].

In many cases, a deep understanding of industrial practices requires close collaboration with industry. Garousi et al. [5] identified industry collaboration and the use of appropriate research approaches [6] as two of the most frequent improvement suggestions for increasing the relevance of software engineering research [2]. To motivate such collaborations, both academia and industry need to benefit from them.

Secondary studies could play a role in connecting research and practice. In an ideal scenario, researchers would start with a need from practice, convert it into an answerable research question, identify, critically appraise, and aggregate available evidence to help answer the question and document the approach and findings in research papers [7]. Despite the substantial quantity of secondary studies published in software engineering [8,9], their focus is not necessarily aligned with practical relevance [10]. Furthermore, researchers need to improve how they communicate research results to practitioners, including evidence-based recommendations [11]. A concern emerging from this scenario is the minimal involvement of practitioners in these secondary studies.

The lack of practitioners’ involvement can limit the contextual relevance of the findings. Such a pattern might suggest an under-appreciation of the contextual nature of software engineering [12,13].

https://doi.org/10.1016/j.infsof.2023.107364
Received 9 February 2023; Received in revised form 6 October 2023; Accepted 8 November 2023
Available online 20 November 2023
0950-5849/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
The underlying assumption is that knowledge can be transferred or communicated to practice at the end of the literature studies. One consequence of this approach is that the relevance of the findings of a literature review is perceived as low for practitioners [10].

We believe that software engineering knowledge is socially constructed [14] and context bound [15]. Given the challenges posed by limited practitioner participation and the insufficient integration of their insights and concerns in secondary studies, leading to decreased practical relevance, we advocate for Rapid Reviews (RR) [16, 17] as a tool to bridge the gap between research and practice. RR are systematic reviews designed to support decision-making under time constraints [16, 18]. They offer two characteristics that make them a strong candidate for connecting research and practice. First, they are conducted in a short period of time, which practitioners appreciate. Second, the studies are framed in the context of practitioners, making the results relevant for them. We have outlined practical actions for involving practitioners in the RR process [19], focusing on knowledge exchange between researchers and practitioners and identifying opportunities for industry-academia collaboration during the RR [19].

In this study, we investigate the practical application of RRs in two independent cases of industry-academia collaboration. We studied how the RRs were implemented and gathered information about expectations, the usefulness of RRs, the results achieved, and the experiences of the researchers and practitioners involved. Based on the two cases, this paper presents the following contributions:

- A description of how the teams conducted the RRs.
- An identification of the benefits and challenges of conducting RRs with practitioners.
- Further recommendations for researchers and practitioners when conducting RRs.

The remainder of this paper is structured as follows. Section 2 presents background and related work. Section 3 presents the steps of an RR and our suggestions for involving practitioners. Section 4 describes the method followed in this study. In Section 5, we present the study’s results. After that, we present some recommendations for conducting RRs in Section 6. Section 7 discusses the results, and Section 8 concludes the paper.

2. Background and related work

In this section, we briefly discuss the need for practical relevance of secondary studies in software engineering. We also point out examples of systematic reviews with practitioner involvement and recent rapid reviews.

2.1. Secondary studies in software engineering

Researchers in software engineering have widely adopted the use of secondary studies [20] as a means to synthesize software engineering knowledge. However, these studies have mainly been used in academic environments, and to identify gaps in research [8], and are criticized for the lack of industry-relevant results [21]. There is a need to connect secondary studies with practice. A few improvements have been suggested to make the presentation of the results more meaningful for teachers and practitioners [22].

Some voices in the software engineering research community have claimed that secondary studies need to connect more with practice [11]. There are a few examples of secondary studies that have involved practitioners (see e.g., [23]) however, none of the existing guidelines sufficiently incorporate interaction with practitioners. For this reason, we propose to conduct rapid reviews interactively with practitioners.

2.2. Rapid reviews in software engineering

Rapid Reviews are a well-known approach in medicine for synthesizing research findings under time constraints. In software engineering, the concept of rapid reviews and the guidelines to conduct these reviews is introduced by Cartaxo et al. [16]. They emphasized the potential of rapid reviews to provide decision-makers with relevant information quickly. Recent rapid reviews in software engineering, inspired by this emphasis, include studies on contemporary topics like migrating from monoliths to microservices architecture [24], security testing in IoT [25], and testing of context-aware software systems [26].

Building on Cartaxo et al.’s foundation, our proposal [19] describes how and when practitioners may be involved in the review process. Recognizing the context-dependent nature of software engineering, our goal is to foster knowledge exchange between researchers and practitioners by creating opportunities for meaningful interactions.

3. The steps of a rapid review (RR)

Cartaxo et al. outline three main phases for conducting an RR: planning, performing, and reporting [17]. Our approach supplements these foundational phases by specifying practical activities and guidelines for active practitioner involvement at each step [19]. In our approach, Steps 1 and 2 align with the planning phase, Steps 3 and 4 correspond to the performing phase, and Step 5 pertains to the reporting phase of Cartaxo et al.’s proposal. The steps for conducting an RR, as elaborated by our proposal for involving practitioners, are illustrated in Fig. 1 and briefly described below:

1. Prepare the review: The first step involves forming a team of researchers and practitioners participating in the review. The researcher leading an RR presents the general aim of an RR, and the typical process, timeline, and expected time commitments. Next, the team needs to agree on the expectations, the extent of involvement, and the responsibilities of researchers and practitioners. The RR topic emerges from the industry’s specific needs and relevant aspects of their context, e.g., current software engineering practices. At the end of this step, a team is formed, and they have identified preliminary information needs.

2. Identify research questions and develop the RR protocol: The second step involves more detailed planning, describing the initial research questions and a protocol. Researchers and practitioners refine the research questions as the understanding of the practitioners’ context improves. This is an iterative process during which the review team develops an understanding of the terminology and domain jargon. It may take time to develop a consensus on the research questions and the scope of the review. Once the research questions are sufficiently clear, the review team further articulates decisions like the search strategy, inclusion and exclusion criteria, and analysis approach. The protocol is a living document that is revised and updated throughout an RR.

3. Search and select papers: In the third step, the search and selection of papers are performed. Several decisions to ensure a rapid literature review are taken in this step at the cost of completeness of coverage. Furthermore, the team develops criteria and a shared understanding of what papers are considered relevant, in particular, taking the practitioner’s perspective and context into consideration.

4. Extract and synthesize data: Step four is about extracting data from the included papers and synthesizing the results and findings. Preparing the reports and templates to be filled with the results may help to save time and focus on the synthesis. These reports may include summaries, slides, infographics, etc. To make the results more accessible for practitioners, the review team can consider creating narrative synthesis [27] and provide summaries [28].
5. Disseminate RR results: The final step in the RR is to disseminate the RR results. The dissemination actions are designed to communicate the results to practitioners and researchers. When sharing the results with the practitioners, an active role of the practitioners involved in the team may add more context and increase the interest in the findings.

4. Research methodology

In this study, we aim to collect the experience of using the proposed RR approach with practitioner involvement [19] in practice and identify further improvements in the guidelines. For this purpose, we posed the following research questions:

1. How did the teams comprising researchers and practitioners conduct the RRs?
   To answer this question, we collect information about how two teams comprising both researchers and practitioners followed the steps proposed to conduct an RR [19]. The approach is briefly summarized in Section 3. We describe how each team applied the guidelines. The detailed results of their application and insights are presented in Section 5.1.

2. What are the benefits and challenges when conducting RRs?
   With this research question, our goal is to collect the benefits and challenges observed during the conduct of the reviews. The benefits are detailed in Section 5.2, and the challenges are presented in Section 5.3.

We attempted to answer the above research questions in a two-case exploratory case study. Our study adhered to the guidelines provided by Runeson and Höst [29]. In the two cases, the RRs were conducted by two independent groups of researchers and practitioners.

There were three main types of participants in the reviews. Practitioners (\(P_{i,j}\), i.e., \(j\)th practitioner in review \(i\)) participate from outside the university for the purpose of the review, researchers (\(R_{i,j}\)) participate from the university for the purpose of the review, and meta-level researchers (\(M_{i,j}\)) participate from the university to conduct the research presented in this paper.

4.1. Case-SoftSelection: Selection of software components

This RR was conducted as a collaboration between one researcher and one practitioner from case company 1. The company is a multinational company developing software and hardware in the area of networking and communications. The review initially focused on exploring the criteria for selecting open-source and closed-source software tools in software development. The goal of the RR was to identify important factors and challenges in selecting software tools and to provide recommendations for improving the selection process at the company.

The following participants were involved in the review:

- Practitioner \(P_{1,1}\): An experienced practitioner working with “technology studies”, which involves understanding current research
- Researcher \(R_{1,1}\): A senior researcher from Lund University, active in the area of Requirements Engineering, with experience from conducting secondary studies
4.2. Case-MLTest: Machine learning testing

This RR was conducted as a collaboration between researchers and practitioners from case company 2, a manufacturer of network cameras for physical security and video surveillance industries. The review’s objective was to understand more about ML testing, and identify research results of interest to the case company.

The following participants were involved in the review:

- Practitioner P2,1: A developer with a background in mechatronics and mathematics, who has worked at the company for about five years with machine learning applications.
- Practitioner P2,2: A researcher employed by the case company, with a Ph.D. in mathematics, who is currently continuing research in the same area as their Ph.D. topic and applying their research in product development.
- Practitioners P2,3 and P2,4 also participated in the review from the case company. However, they were not interviewed for this study since they only attended the meetings and were less actively involved in the steps.
- Researcher R1,1: An experienced researcher in the area of machine learning in software engineering.
- Researcher R1,2: A Ph.D. student in the area of software testing and machine learning. The review is relevant to the researcher’s thesis work.
- Researcher R3,1: Experienced researcher in the area of software testing. The third author of this paper.
- Meta-level researcher M1, with the same role as in Case-SoftSelection.

Researchers R2,1, …, R2,3 have been involved in traditional literature reviews mainly with academic participants prior to this RR. This means that the concept of a systematic literature review is not new to them. The conducted RR was presented in a conference publication [31].

4.3. Data collection

We employed semi-structured interviews as data collection method [32]. All interviews were conducted in English. While we identified central themes and prepared guiding questions, the interview format was flexible, adjusting to the direction of the conversation. Thus, they were not interrupted if the interviewee jumped forward to an interesting topic. This approach allowed us to tailor questions and delve deeper into spontaneously merging topics of interest. The complete list of questions can be found as additional material.1

For Case-SoftSelection, we conducted interviews before and after the review, while for Case-MLTest, interviews were held only after the review. Researcher M1 was present throughout both reviews, even directly observing Case-MLTest’s planning. This researcher helped guide the RR method in both cases.

After the reviews, researchers and practitioners from both review teams were interviewed to capture their experiences. These discussions covered how they managed the reviews, their collaboration dynamics, and their reflections on the results. They also shared insights on the expectations, outcomes, and future joint work. Furthermore, we delved into their perspectives on research evidence and what constitutes a valuable research contribution and a good research paper.

Additionally, we had access to the specific files associated with the RR where the researchers documented the steps and archived files pertinent to their RR activities. These documents offered supplementary insights, enabling us to cross-reference and validate the interview responses. For example, we looked at the details of the RR, such as the search string, the inclusion/exclusion criteria, the number of papers found, and the presentation slides.

4.4. Coding and analysis

The eight interviews were transcribed and coded using QSR International’s NVivo qualitative data analysis software for coding and analysis. M1 participation in the RRs enabled us to triangulate interview data with project documentation and refine our recommendations. For our coding and analysis, we employed the thematic analysis approach recommended by Cruzes and Dybå [33]. Below, we describe the codes used in the analysis and the steps followed to analyze the data.

4.4.1. Coding levels

We used three levels of coding to organize and analyze the data. The first level (level-1) acted as an index to map the chunks of the interview to one or more of the aspects of interest, such as benefits, challenges, and steps. The second-level (level-2) grouped findings, and the third-level (level-3) codes described or qualified the second-level codes. For example, suppose the level-1 code was about benefits. In that case, the level-2 code could be about a specific benefit (e.g., mutual understanding), and the level-3 codes could describe unique aspects of this benefit or capture relevant findings (e.g., views alignment, define common terminology, other’s perspective). The initial level-1 codes were:

- Case Description: We used this code to index chunks of the interview that we could use to describe the case. Some level-2 codes we were interested in include background, profiles of researchers and practitioners, initial views of each other, experiences working with industry/academia, and participation in secondary studies.
- Expectations: We captured with this code the interviewee’s expectations. With level-2 codes, we classified the expectations according to who had the expectations (researchers or practitioners) and the type of expectation (e.g., empirical validation, exchange, new knowledge).
- Steps (code for research question 1, i.e., about the conduct of an RR): With this code, we indexed the chunks of the interview where the interviewees described how they conducted the RR. The level-2 codes are the steps of the RR (see Section 3).
- Benefits and Challenges (related to research question 2, i.e., benefits and challenges of RRs): With this code, we indexed the chunks of the interview where we identified positive aspects brought by the RR (i.e., benefits), or challenges in conducting an RR. Each of the benefits was coded under a level-2 code. The initial set of level-2 codes were the expected benefits and challenges based on the experiences from rapid reviews in medicine [34] and our experience conducting secondary studies in software engineering.

The initial set of codes evolved during the coding and analysis process [33]. For instance, we added one more level-1 code related to the outcomes of the RR. Similarly, level-2 codes were added when we identified new benefits or challenges. The final set of codes is available as additional material.2

---

1 Link to additional material.
2 Link to additional material.
4.4.2. Coding steps and analysis

We have followed the steps described below to code the interviews:

1. In NVIVO, the first author coded the chunks of the interview (i.e., questions-answers) with the level-1 codes.

2. To facilitate the exchange within the research team, the chunks of the interview were exported to a spreadsheet document (Fig. 2) with the following columns:
   - Check validity: Used to indicate whether other researchers reviewed the chunk.
   - Document: An identifier that represents the RR and the interview. For example, RR1P1 Before means the first interview with the practitioner in Case-SoftSelection.
   - Interview fragment: The verbatim text of the interview. Each cell contained a question answer pair.
   - First-level code: The level-1 code was used to code the interview fragment. One interview fragment can be coded under multiple level-1 codes.
   - Second level code: As described above, the level-2 code groups the findings. One first-level code can be coded under multiple second-level codes.
   - Third level Code: The description/qualification of the 2nd level code. One second-level code can be coded under multiple 3rd level codes.
   - One column for recommendations
   - One column for comments from each author

3. The first author coded the interviews.

4. The third and fourth authors reviewed the coding in the spreadsheet. Disagreements were discussed and resolved to reach a consensus. This review was facilitated by a shared repository where all authors could access the raw data and analysis files.

5. The second author also accessed the shared repository and provided reviews on the coding, incorporating feedback and suggestions.

6. The first author addressed the comments and suggestions from the reviews.

While the coding was being done, the research team met weekly to discuss the coding process and the evolving codes. Once the interviews were coded, the codes related to the steps were used to develop the narrative about how the RRs were conducted (as reported in Section 5.1). Similarly, the codes related to the benefits and challenges (2nd-level) were used as a basis for synthesizing and summarizing the findings reported in Section 5.2.

The set of recommendations was developed incrementally. The first set of recommendations was derived from explicit coding (i.e., recommendations expressed by the interviewees). Furthermore, some recommendations are responses to identified challenges. Finally, all authors reviewed and discussed the final set of recommendations in a meeting and revised them while writing the manuscript.

5. Results and analysis

In this section, we present the results of the coding and analysis. The section has three subsections outlined in accordance with the research questions. Section 5.1 describes how the RRs were conducted. Section 5.2 includes the benefits of conducting RRs as a tool for researchers-practitioners knowledge exchange. Section 5.3 describes aspects found challenging for the review teams when conducting RRs.

5.1. Conducting RRs

Here, we describe step by step how the two review teams conducted the RRs based on the guidelines. Table 1 shows a summary of how the RRs were conducted. We used two styles to format the quoted text in the subsections. A sidebar marks text from the guidelines and takes an entire paragraph, whereas text from the interviews is embedded within quotation marks in the text.

5.1.1. Preparation

“The first step involves forming a team of researchers and practitioners participating in the review. The researchers who lead the RR need to present the aims, process, timeline, and expected commitment to the group.”

Since the level of prior knowledge about RR differed in the two cases, the need for an introduction to the method varied. In Case-SoftSelection, the researcher leading the review was new to the concept of RRs, while the researcher leading Case-MLTest had been involved in developing our supplement to the RR guidelines. Thus, we provided the RR guidelines to the researcher R1 to lead Case-SoftSelection. Moreover, we provided material to support the presentation of the ideas, initial planning, and a document to develop the protocol. In Case-MLTest, the third author of this paper R3 was part of the team conducting the RR.

In both cases, the review teams were formed after the initial discussions, when the topic was agreed on. In Case-SoftSelection, there were no changes on the practitioners’ side, while the research team was formed based on the emerging topic. Initially, with only one researcher, R1, one more researcher (second author in this paper) joined the team after one iteration of search and selection. In Case-MLTest, three more practitioners with different but relevant roles were added to the team, while the researchers’ team remained the same.

“Next, the team needs to agree on the expectations, the degree of involvement, and the responsibilities of researchers and practitioners. Ideally, the RR topic emerges from the industry’s specific needs, and relevant aspects of the context and current practices are introduced to the team.”

In both cases, the motivation for conducting the RRs was to explore possibilities for industry-academia collaborations on new topics. Thus the industry’s specific needs were not of the highest priority.

The idea to conduct Case-SoftSelection came when a practitioner R1 reached out to a group of researchers wanting to explore ways to work together. The third author of this paper took part in these initial discussions and suggested conducting an RR as a first step. At this point, the main interest of conducting the review was to explore potential...
The two review teams that were formed were of different sizes. Case-SoftSelection initially involved one practitioner and one researcher, while Case-MLTest involved four researchers and four practitioners. In both cases, the information need was described as a very high level of abstraction, “How to test ML applications” and “How to select tools and software components”. Even though the information needs were general at this point, and one was more specific than the other, they described the topics of interest at that time. These needs were the starting point for starting the search and specifying the research questions. The particular elements of the context that influence the review were not identified at this stage but evolved through interaction between researchers and practitioners during the reviews in both cases.

5.1.2. Identify research questions and develop the RR protocol

“The second step consists of more formal planning where the RR research questions are defined and an initial protocol to conduct the RR is started.”

After two meetings in Case-SoftSelection, the researcher and practitioner agreed on a more precise idea about the topic to explore and the next steps in the RR. Then, the review team comprised the researcher $R_{1,1}$ and practitioner $P_{1,1}$. The review’s main topic was the selection of software components.

In Case-MLTest, during a meeting with the practitioners, the researchers shared the overview and the list of topics. Then, the practitioners presented their work and products supported by machine learning and challenges. After the meeting, the two groups agreed on the broad topic of the RR, testing of ML applications.

“Defining research questions with practitioners is an iterative process that requires understanding the practitioners’ context, practices, challenges, and terminology.”

Since, in none of the cases, any specific research questions were decided upfront, some initial effort was spent identifying questions of high relevance for everyone involved. In Case-SoftSelection, after a couple of initial conversations between $P_{1,1}$ and $R_{1,1}$, they came up with a preliminary idea about exploring the selection of tools and software components. The researcher reviewed papers related to the first research question about the criteria for selecting software components. Meanwhile, the practitioner also identified criteria, not from research papers but by reflecting on their own experience and consulting their colleagues. In Case-MLTest, an initial brainstorming meeting was held. Before this meeting, the researchers developed a preliminary taxonomy of state-of-the-art ML-testing using the SERP-taxonomy architecture [36]. The taxonomy served two purposes, to present a general overview of the published literature to the practitioners and guide the discussions about the practitioners’ context and needs. This meeting resulted in a list of potential research questions, which were then ranked independently by everyone involved in the meeting (four researchers and four practitioners) to select the most relevant questions. Based on the ranking, the first research question was formulated about data and input testing, i.e., how to test the data. In ML testing, assessing a dataset validates its quality, consistency, and relevance for a given machine learning task. It is not about testing in the traditional software sense but ensuring the data’s suitability for modeling [37].

“Then, it may take time to develop agreements about the research questions and related terminology.”

### Table 1

<table>
<thead>
<tr>
<th>Case</th>
<th>Case-SoftSelection</th>
<th>Case-MLTest</th>
</tr>
</thead>
<tbody>
<tr>
<td>General topic</td>
<td>Software component selection</td>
<td>Testing machine-learning systems</td>
</tr>
<tr>
<td>Practitioners' expectations</td>
<td>Pilot industry-academia communication, explore research findings on criteria for software selection</td>
<td>Explore research on machine learning testing</td>
</tr>
<tr>
<td>Researchers' expectations</td>
<td>Pilot industry-academia communication, develop a collaborative network</td>
<td>Explore machine-learning testing in practice, networking</td>
</tr>
<tr>
<td>Team</td>
<td>1 Researcher ($R_{1,1}$), 1 Practitioner ($P_{1,1}$)</td>
<td>4 Practitioners ($P_{2,1}$, $P_{2,2}$, $P_{2,3}$, $P_{2,4}$), 4 Researchers ($R_{1,1}$, $R_{1,2}$, $R_{1,3}$, $M_{1}$)</td>
</tr>
<tr>
<td>Research questions</td>
<td>What criteria are relevant for the company to consider when selecting a SE tool or component?</td>
<td>General question: How to test the dataset? In ML Systems</td>
</tr>
<tr>
<td>Search</td>
<td>Key-words search in Scopus</td>
<td>Based on 3 systematic literature reviews</td>
</tr>
<tr>
<td>Papers found</td>
<td>147 primary studies, 27 papers coded</td>
<td>180 primary studies mapped to 10 challenges. 5 of the papers mapped to the general question</td>
</tr>
<tr>
<td>Analysis</td>
<td>Extracting criteria from papers, and exchanging with practitioners</td>
<td>Developing a taxonomy of ML-testing based on literature and one case context</td>
</tr>
<tr>
<td>RRs outcomes</td>
<td>Preliminary model for component selection</td>
<td>Examples of problem–solution matches (non-extensive)</td>
</tr>
<tr>
<td>Dissemination</td>
<td>Share preliminary results with practitioners, working sessions to build a model, research paper</td>
<td>Share papers with practitioners, share findings with company representatives</td>
</tr>
<tr>
<td>Post-RRs</td>
<td>The second iteration to extend the search and validate the model</td>
<td>Master thesis proposal, new projects</td>
</tr>
</tbody>
</table>
As stated above, in both cases, it required several interactions, in terms of meetings, workshops, and offline communication, to define the research questions for the review. R_{1.1} pointed out that provided presentations, templates, and checklists were a help to communicate expectations within the team:

"Since we all know what a literature review is. Having the presentation slides and the template for having different things to fill in made it very clear. These are the things we need to agree, and it was good to present them to the practitioner to get them to understand what the method was about. And to try and get the scope nailed down. I think that was our biggest challenge at the beginning, to have something that was a reasonable scope that was clear and sort of not all over the place. So I would say that having these templates helped, but then, of course, the discussion still has to be there and you have to get the practitioner into the little box that is easier for us to handle" - [R_{1.1}]. Furthermore, the background section in the template (part of the material initially to R_{1.1}) showed to be helpful in validating the problem understanding with the case company. R_{1.1} filled it out during the initial discussions to develop a problem understanding. This was then sent to P_{1.1} to confirm the view and fill in the gaps. In both cases, the final set of review questions was exploratory. Furthermore, time limitations prevented the exploration of the initial broad topics, but questions were refined during the reviews.

"Once the research questions are defined, the review team may develop an initial version of the review protocol. The protocol contains information about the search strategy, inclusion and exclusion criteria, approach to conducting the analysis, and the decisions made along with the review. Besides, the protocol is updated with the progress and result of the following steps."

Researcher R_{1.1} did not use the template we provided for the RR protocol. In their view, the document did not help accomplish the RR faster because it included many details to fill in. Instead, they kept track of the steps, search strings, inclusion criteria, decision, and search details in auxiliary files according to their own preferences. These files were stored in a project repository where the researcher kept track of review steps, meetings, advances in the review, search results, bibliography files, document drafts, and reflections and suggestions about the process. One of the reasons for researcher R_{1.1} to keep track of the RR in detail was to be prepared if the results would be used for an academic publication. The review team agreed that they would work iteratively, and the initial goal was to build a model that could summarize and synthesize their findings. In Case-MLTest, the researchers started developing the research protocol based on the template we provided. The protocol specified the research questions, search strategy, and inclusion criteria. An additional document kept track of the work plan, activities, roles, and responsibilities.

5.1.3. Search and select papers

"In the third step, the search and selection of papers are performed. Several decisions to ensure a rapid literature review are taken in this step at the cost of completeness of coverage. Furthermore, the team develops criteria and a shared understanding of what papers are considered relevant for the RR."

In both cases, the researchers mainly searched and selected papers. They had previous experience conducting systematic literature reviews and followed similar principles. They kept track of the process and documented the decisions that were made. The practitioners did not have any opinions about the process for finding the papers and were more interested in finding something applicable in their context than ensuring extensiveness in the search: "I don’t really know how much time they spent looking for it, so it’s hard to know if it would be possible to look more or not. I’m quite confident that they spent more than we could do from our side. So I still think it is very valuable, and I trust their opinion enough not to spend more time myself on it if they come to some conclusion. I would say." - [P_{2.1}].

In Case-SoftSelection, Scopus³ was used to search for literature. Search and selection were made in three iterations while defining the final scope within the review team. In Case-MLTest, the search step was skipped since the researchers were aware of three recent literature reviews on the topic and used them as a starting point. Although they were confident in the rigor of the searches in those secondary studies, they were also aware that the field is active and it is possible to miss something. However, completeness was not the main priority . Q. How important is being systematic vs. finding something applicable for them? A. Being extensive wasn’t our priority. So we wanted to really have something applicable. Once you have something applicable, it is easier to start from that. Q. And for them? Do you think it is the same? A. Yes, they do not care about completeness." - [R_{2.1}].

In Case-SoftSelection, the review topic was not the main topic of the researcher’s expertise. However, they were confident in the systematic review process.

In both cases, selecting the relevant set of papers was iterative and involved feedback from the practitioners. In Case-SoftSelection, the review team had meetings discussing their findings. In Case-MLTest, practitioners were involved in reading and commenting on papers of potential interest, which helped refine the inclusion/exclusion criteria. "then they started doing the review, finding a bunch of articles. And then they sent us a few of them, and we looked at them. They were not very relevant. They talked about the training data and stuff that we’ve talked about. But not really for image training data. It was more general for other kinds of media. And I found after reading those first papers I couldn’t really see how to apply those general techniques to image data. So then there was a second round where they added this criterion that we wanted to work with images or videos. And then, I found more relevant work." - [P_{2.1}]. After the first iteration in Case-SoftSelection, the review team realized that many of the papers were quite old, so they adapted the search strategy to find more recent research.

In both instances, there were an equal number of research publications that were screened (hundreds) and reviewed (30–40) studies. However, since the goal of Case-MLTest was to find an applicable technique rather than develop a more general theory, as in Case-SoftSelection, the procedure of excluding papers continued until only a handful of papers remained. The practitioners then evaluated the papers and provided feedback to the researchers about their relevance to their current problems. As a result, the inclusion and exclusion criteria were updated.

5.1.4. Extract and synthesize data

"Step four is about synthesizing the results and findings from the included papers. An idea to better communicate the findings is to design the reports and documents that will be used to share the results in advance. It is vital to ensure that the findings will be easy to follow for the practitioners. For that reason, it is suggested to use narrative synthesis and practitioner-friendly summaries. A recommended practice is to hold reaction meetings where the RR team presents preliminary results to the team or an extended group of practitioners. The reaction meetings give feedback to the team and may inspire them on how to communicate the results."

The data analysis approaches differed between the two cases due to the somewhat different goals. Case-SoftSelection had a higher ambition

³ Scopus is one of the largest abstract and citation databases of peer-reviewed literature: scientific journals, books and conference proceedings. (Retrieved from https://www.scopus.com, [38]).
of synthesizing results of a larger share of included papers. On the other hand, in Case-MLTest more effort was spent selecting relevant and applicable approaches for their case company’s context.

Case-SoftSelection applied thematic coding to find answers to their research questions. The researcher derived an initial set of codes that evolved along with the coding based on the research question. The papers were coded using Nvivo. After clustering the codes, the outcome of the RR was a list of criteria for selecting software components. On the other hand, the main focus of the coding in Case-MLTest was to improve the inclusion and exclusion criteria. As a result, the initial SERP-taxonomy [36] was extended with information retrieved from the included papers and the practitioners’ feedback regarding their context. Based on the common taxonomy, researchers were forced to be explicit about the relevant details of the proposals.

5.1.5. Disseminate results

"The final step in the RR is disseminating the RR results. The dissemination actions are designed to communicate the results to practitioners and researchers. When sharing the results with the practitioners, the practitioners involved in the review team have an active role, e.g., when presenting or discussing the results, they may add more context and thus increasing the interest in the findings. In addition, even though it is not the primary goal, the researchers involved may be interested in communicating the findings to academic audiences through research papers."

In neither case did the researchers conduct specific actions to disseminate the results beyond the review team, e.g., within the organizations. However, in Case-MLTest, preliminary results were shared with practitioners from another business unit where the results were relevant. "And I read maybe three of those articles. And one of them I passed on to another team that actually [] evaluated [] and compared it with a few other techniques that they were aware of." - [P22]. After the reviews, both review teams also reported their results and experiences in scientific publications (see [30,31] for Case-SoftSelection and Case-MLTest respectively).

In both cases, preliminary results were presented by the researchers at different stages during the review. The manner in which these were presented was also influenced by the expectations of the practitioners. In Case-SoftSelection, the topic was more general, and the identified papers were more divergent. There the researcher extracted and synthesized contributions and presented a preliminary model of component selection (a taxonomy of criteria) to the practitioner. Researcher R1 shared the list of criteria with practitioner P1. They also shared some of the papers (actual pdf files) that were the most relevant to the topic. Then, jointly throughout a series of meetings and discussions (at least three meetings including a working session with a whiteboard), they integrated the criteria found in the research papers with the ones collected by the practitioner to produce a model for selecting software components. The model and research findings were not communicated to a larger group of practitioners within the company. Instead, the model was an input for a research study where the review team planned to complete and evaluate the model. In Case-MLTest, the practitioners were up to date with current research and were used to reading research papers. The preliminary results were presented in terms of selected papers and the researcher’s reflections on potential inclusion and exclusion criteria.

The final results of Case-SoftSelection were reported as a short Powerpoint presentation showing the taxonomy and explanations of the identified criteria. In Case-MLTest visual abstracts [39] were created, summarizing the contributions of the five best problem-solution matches, and presented at the concluding review meeting.

5.2. Benefits

By benefits, we mean the positive impact of the RRs on the researchers, practitioners, their relationships, and their organizations. A benefit could be experienced in different ways by different stakeholders. For example, obtaining results from literature in a structured way can be seen as a benefit for the practitioners if we assume that they usually do not conduct systematic reviews. On the other hand, for the researchers who typically perform systematic reviews, the benefit is the possibility of involving practitioners in the process and increasing the industry relevance of the reviews.

Before analyzing the benefits of conducting RRs, let us examine their context and preconditions. When analyzing the advantages of Case-SoftSelection for developing networks for collaboration, it is fair to say that from the beginning, an important motivation for P1 was to find ways to collaborate with academia. The RR approach seemed to be a way to start working on something concrete to find common topics of interest and build a relationship. According to R1, P2 was more interested in the meta-level, i.e., finding ways to collaborate. So the overarching goal in Case-SoftSelection was to explore ways to collaborate. The researcher, R1, was also interested in collaboration with industry, although not explicitly as a research topic, but rather in finding practical means to work smoothly with industry. Thus, the Case-SoftSelection may be seen as a way to pilot and assess the feasibility of collaboration with academia.

Case-MLTest was motivated by researchers P1, P3 since they were working on a project about testing machine learning. They were interested in understanding industry practices and networking with practitioners in the field. According to P2, the RR allowed the participants to work on a specific problem and look at the horizon for future collaborative work. The topic interested them, and they allocated resources and got involved in the review. In the long term, the Case-MLTest contributed to identifying common interesting topics, meeting potential new collaborators, and determining how they can complement each other to work together. As one of the results, the review team got an overview of the field that facilitates identifying opportunities for new studies.

From the interview material, several benefits were identified and can be organized under four headings: (1) terminology alignment, (2) pilot future collaborations, (3) overview of the field, and (4) usage of research.

5.2.1. Terminology alignment

One of the advantages of using RR in collaborative research is the alignment of terminology between researchers and practitioners. This alignment is more challenging to achieve in a researchers-only review, but it is more likely to happen in an RR due to the continuous interaction between researchers and practitioners.

In the Case-SoftSelection, the researcher and practitioner started collaborating in this review and had different views of the topic. While the practitioner seemed more interested in the novelty, precisely the publication date, the researchers focused on the criteria alone, regardless of the publication date. On the other hand, the practitioner perceived that his involvement helped the researcher focus on relevant findings. By initially defining the topic and research questions, they reconciled these different views. "So we spent a lot of time on that. I think that helps, and I think that’s good because when you start to talk about actual issues, it’s much easier." - [P1]. After the initial discussions, the researcher and practitioner agreed on a more precise idea about the topic to explore and the next steps.

The initial findings from the RR were synthesized into a preliminary model for component selection following the idea of sharing preliminary results with practitioners. This step was also provided an opportunity to align the terminology and concepts used by both parties and to develop a common understanding of the topic. "Now we have our model and I sort of introduced the thought that we could sort of redo the
literature review and then he was a bit more interested in yes that might be a good idea. And he had actually found an article that he was referring to and I read that was relevant.” - [R_{1,1}]

Subsequently, we observed indications that interaction when conducting the steps of the RR contributed to understanding each other's perspectives and advancing on developing common outcomes that synthesize the practitioner's and researcher's perspectives and findings.

5.2.2. Piloting future collaborations

The formation of new networks emerged as a benefit for practitioners and researchers in our cases. It is seen as positive that they have learned how to collaborate and have found a process with meetings to manage the work. The researchers also emphasized the positive impact of gaining an understanding of the industrial context and problems, which facilitates defining future research questions and projects. Another advantage of RRs is their limited time commitment, making them suitable as a pilot for future collaborations. Unlike larger project requiring long-term collaboration, an RR serve as a feasibility test for collaborative work. After the RR, both parties can decide whether to extend their collaboration into larger projects. In our case studies, Case-SoftSelection and Case-MLTest, there was mutual interest in continuing collaborations. For example, in Case-SoftSelection, a master's thesis project was developed, while in Case-MLTest, further iterations of the RR were planned along with empirical evaluations within the company.

5.2.3. Overview of the field

The researchers thought that they obtained not only knowledge from the literature but also from the company. For them, one contribution is that they were able to understand the industrial perspective of the problem that they investigated: "Well, I think all the way through the study, we have learned a lot about the topic that we are looking at. But I think the most significant for me from this rapid review with [Company] was that we actually knew what was interesting from the industry perspective" - [R_{1,2}]

Even if the researchers did not know the review's topic in detail before the review starts, conducting the RR was an opportunity to gain knowledge on the industrial perspective in context.

5.2.4. Usage of the results of the review

When it comes to the usage of the results of the review, it can be seen in different ways. Industry practitioners find it positive to get a general understanding of the research front. Even if they typically do not engage in literature reviews, they show interest in the research landscape. "...maybe you can get confirmation on your own ideas that it is like basic stuff, right? or do we totally diverge" - [P_{1,1}]

By contrasting their actual practices with the outcomes from research papers, practitioners get indications if their practices are roughly in line with other companies or if they are behind other companies in a specific topic.

However, [P_{1,1}] pointed out that this contrast is not easy to do. Besides, it depends on the area. Overall, the participants found value in becoming aware of the research state of the art and having some clues to what scientists and other industries are working on. It should also be noted that the participants identified research literature before, but this way of identifying literature was good compared to other approaches (e.g., Twitter and ad-hoc search were mentioned). The industry representatives did not express high requirements on empirical evidence. Instead, they said, for example, that finding one single relevant paper with a relevant solution would be valuable. On the other hand, they would be more cautious if the paper only presented theoretical results and see that as a risk. They stated that it is positive if the paper includes evaluations, but if it does not require too much effort, they may try it out themselves in their context. It is also mentioned that they want to compare new solutions quantitatively with metrics when they find new solutions.

Moreover, RR results offer insights into research gaps and opportunities for future research. Actually, Case-SoftSelection spawned M.Sc. collaboration projects and ideas for future applications. One practitioner highlighted the usefulness of gaining insights into the interests of researchers from reading papers, as it made it easier to collaborate with researchers in the future. "I think so because it gives me much more insight into what they are interested in. So if I’m sitting with a problem in the future, then I feel that I have more knowledge of when would this actually be interesting to research, and then I could reach out to them and ask [if] there is something that you want to cooperate with.” - [P_{2,1}]

5.3. Challenges

In the two cases analyzed in this study, we identified some challenges when conducting RRs. By challenge, we refer to something that poses difficulties when performing the review. The challenges and limitations are grouped in the following main sections:

- challenges related to roles i.e., researchers and practitioners not being aware of their responsibilities
- challenges related to the lack of results matching the needs and expectations of the review team
- challenges related to the timeliness of the reviews

5.3.1. Involvement and expectations

Conducting a literature review is demanding, and having researchers and practitioners collaborating in the loop poses even more challenges. Practitioners have their objectives related to the current challenges and directions in their companies, and researchers are interested in finding more generalized results, which is also reflected in the way primary studies are written.

We noticed in the two cases that the practitioners had little awareness of their role in the RR. Even though the term and steps were introduced at the beginning of the RR, and they were actively involved in the activities, they were unaware that they were participating in a different type of literature review and the steps to conduct it. This has a positive side, in that it does not burden the practitioners with research aspects they are unfamiliar with. However, it also has a negative side, since if the practitioners were more aware of the steps, they could be more involved in the process and relate the results to their own context. Additionally, if their experiences were positive, they could be more motivated to conduct RRs in the future.

Another challenge observed was the disparity in expectations and goals, which affected how the team conducted the review. For researchers, conducting literature reviews is a regular research task, but practitioners can have other expectations. In Case-SoftSelection, there was a mismatch between the researchers' and the practitioners' expectations and goals. The researchers perceived that the practitioners were not as interested in the generalized knowledge. Instead, practitioners seemed more interested in specific, individual studies or findings. To address this challenge, the researcher implemented an action plan guided by the review protocol focused on pinpointing specific research questions. This approach piqued the practitioners' interest and heightened the review's relevance. "I think it was very good to have something concrete to do to produce some output to talk about. Otherwise, I think my collaborator [...] has a lot of things to say for [their]self, so it has been good to have [the process] to get some information and some knowledge from the review.” - [R_{1,1}]

5.3.2. Lack of results matching needs and expectations

In both cases, we noted that the papers found during the review did not precisely match the practitioners’ needs and expectations. This challenge can also be seen as a result of different expectations between researchers and practitioners. The identified articles did not genuinely meet the expectations of the practitioners. There were three main types of mismatch between the papers found during the review and the
practitioners' needs and expectations. One type of mismatch was that the practitioners had specific questions from their specialized field, and the papers found were not directly related to these questions. This can be challenging for the review team, as they may not be able to find the information they are looking for and may be disappointed with the results of the review.

The second type of mismatch was that the identified papers were older than expected by the practitioners. The practitioners may be looking for the most up-to-date information on a topic, and older papers may not be perceived as relevant or useful. However, it is important to recognize that research and practice have different paces, and what may be considered an up-to-date problem in practice could be considered a problem solved in research.

Finally, the papers can also be considered to be too “long-term” or too theoretical for application in the short term. Practitioners may be looking for more practical or applied information that they can use in their work immediately. Theoretical or long-term papers may not be considered useful in these cases. One of the practitioners commented on this challenge: “I think the things I found there [were] probably a lot more long-term than what I’m looking for. For example, one of the papers was interesting and perhaps we can use this at some point, but it’s not something that I’m going to spend more time now because I think it will take [1] too long time to get payback for it.”  

That is, the identified papers were older, more long-term, and not in the practitioners’ specific fields than they would have wanted.

5.3.3. Timeliness

We aimed to get participants’ perceptions of the time required to perform the RRs and how timely the results were produced. One researcher ($R_{1}$) stated that the review was relatively rapid compared to previous reviews they had conducted. However, the researcher also had experiences from other secondary studies conducted in a shorter time frame. The practitioners (Case-MLTest) did not discuss the lead time as much as they discussed the effort involved them. They noticed they had many other tasks in parallel, and they did not have to spend a lot of effort on this review. Interestingly, while the RRs spanned a relatively long period, they were still seen as rapid primarily because of the limited dedicated time participants had to allocate. Regular meetings were seen as a good way to keep the work going.

In both cases, the researchers pointed out the need to use tools that could support activities e.g., search, selection, managing references, and analysis. However, these tools were not used in any of the reviews.

6. Recommendations for conducting RR

Based on the results obtained, we propose the following recommendations for researchers and practitioners conducting RRs. These recommendations aim to enhance the value of the RR for researchers and practitioners, maximize benefits, and address previously described challenges. Derived from our findings and discussions among co-authors during analysis and manuscript preparation, these recommendations are enriched by the valuable insights of co-authors who directly participated in the RRs. Table 2 lists the recommendations for each step proposed to conduct RRs. Below, we describe the recommendations in more detail following the steps of the RR process.

6.1. Prepare

The team leading the RR can incorporate external experts to the review team if additional expertise is considered necessary. Although the RRs guidelines suggest forming a team of researchers and practitioners in the first step, the teams do not necessarily maintain the same composition throughout the process. For example, the Case-SoftSelection, the team was formed by one researcher and one practitioner. However, the team was expanded for a second iteration to include another researcher when they realized that they would like to consider the views of a researcher with a different background and more specific experience.

Specific attributes of the practitioners or their organization’s culture can streamline research collaboration. Companies with strong research and innovation foundations tend to be more collaborative partners. For instance, in Case-SoftSelection, the company was a large telecom company, and the practitioners were from a part of the organization that was in charge of the frontier of cutting-edge technologies. Consequently, as part of their work assignments, the practitioners focused on both advances in academic research and conducting applied research. In Case-MLTest, the company started as a startup incubated in a university environment and later was acquired by the company. These facts indicate, to some extent, an openness and willingness to work with researchers. Thus, when planning to conduct RRs with industry partners, we suggest identify cultural aspects that may influence a positive environment for the RR. Identifying these aspects is not straightforward. However, researchers can be aware of some signals that suggest a willingness to collaborate, e.g., attitude during the meetings, openness to discuss current problems, and dedication of time. In summary, if the researchers understand the practitioners’ cultural aspects and context, they can promote actions to develop a positive environment for the RR, which can increase the chances of success and the potential benefits.

Involve practitioners with a research background or appreciation for research. In Case-SoftSelection and Case-MLTest, the practitioners involved had research experience and, therefore, some appreciation for research work and working with researchers. For instance, in Case-MLTest, the researchers signaled that the communication was much more straightforward since one of the practitioners had a Ph.D. and his work included contact with research. However, research background or appreciation for research includes not only Ph.D. holders. Practitioners who have co-supervised master’s theses or participated in research studies can also be suitable candidates. For instance, as seen in Case-SoftSelection, the practitioner had a positive attitude toward working with researchers making it easier to start talking.

In our two cases, the researchers were in charge of planning the RRs. Based on their experiences, we highlight some aspects to consider when planning the RRs. At the very beginning of the RR, the researcher should identify the expectations and motivations of the practitioners to conduct the RR. These motivations differ slightly from the information needs explicitly related to the RR topic. By the expectations and motivations, we mean the implicit reasons that encourage practitioners to work with researchers. These reasons may vary. Some examples are: getting feedback from different perspectives, hiring people, building a brand and reputation, getting help with a particular problem, or fulfilling a requirement from managers and staff. If the researchers know what motivates the practitioners to participate, then they can develop consensus on a goal that brings value to both sides.

RRs are envisioned as a joint effort between researchers and practitioners. However, once the project starts, there is an inherent risk of losing commitment and willingness to work. Reflecting on the two cases, we noted instances where the engagement could have been lost. Previous studies on industry-academia collaboration have pointed out the topic’s relevance as key to maintaining engagement and commitment to the collaboration [40]. In the context of our RRs, the selection of mutual interest helped to maintain engagement and commitment to the collaboration. Therefore, ensuring that the selected topics have to be interesting for both researchers and practitioners is vital. This helps to maintain engagement and commitment to the collaboration. Furthermore, one way to maintain the interaction, particularly evident in Case-SoftSelection, was to plan for incremental and tangible deliverables that could be discussed in the meetings. These small concrete outcomes, including summaries of papers, short lists of papers, or overviews of main findings, facilitated sustained engagement and reinforced the practitioners’ perception of value.
6.2. Define RQ

Getting information about the practitioners’ context is critical when scoping the problem and formulating research questions. For this purpose, the team should plan for a lot of initial interaction to get a deep and common understanding of the review questions. In our two cases, it took several iterations to define the questions. The researchers’ efforts in this stage should focus on the practitioners’ context. In Case-MLTest, holding the input meeting where the practitioners presented the problem and contexts to the researchers was an opportunity to discuss with the practitioners. These types of meetings allow the researchers to ask specific questions about the context and identify other variables that may be relevant to the problem. Furthermore, in our two cases, the teams used tools to support this step, like developing SERP-taxonomy [36] and ranking the topics according to the participant’s interests.

6.3. Search and select papers

One of the critical aspects of our proposal is to be in continuous contact with the practitioners. During the search and selection of papers, the practitioners’ feedback is key to ensuring that the papers are relevant to the practitioners’ context. The practitioners can be involved in refining inclusion/exclusion criteria by providing them with preliminary results as in Case-SoftSelection or by sharing papers to read and react to as in Case-MLTest. In the cases studied in this paper, we saw how defining small concrete outcomes was beneficial to promoting interaction in RRs includes one specific step named dissemination of results for the practitioners’ context. The reason for this lack of results was that the topic was recent both in research and academia. Therefore there were no techniques to apply directly to the practitioner’s context, i.e., computer vision. However, some papers brought ideas and insights that could be useful for the practitioner’s context. Additionally, a very positive result of this Case-MLTest is that the researchers and practitioners formulated a master thesis directly on the topic, a Ph.D. project related, and plan to keep doing joint work around the specific topic. Thus, the results were not negative, but the researchers need to be prepared to handle similar situations and find ways to communicate results in a way that is useful and relevant to the practitioners.

6.4. Extract and synthesize data

Once the review team has selected a set of papers, the next step is to extract the information from the papers and synthesize it. As the review is supposed to be conducted in a short time, the guidelines suggest thematic analysis as a suitable method for this step. Thematic analysis may help to overcome terminology and context gaps. We have seen in Case-SoftSelection that the researchers adapted the extraction and synthesis to their expectations and available literature i.e., building a model for software component selection. On the other hand, in Case-MLTest, synthesis was unnecessary since the practitioners were interested in finding specific papers that matched their problem. Then, the synthesis was not their top need but papers they could implement in their context. Thus, the review team needs to adapt analysis to expectations and available literature to ensure that the results of the RR are relevant and useful to the practitioners.

6.5. Disseminate results

The overall message when disseminating results is to be aware of different terminologies and contexts. It is also important to care about sharing valuable results for the practitioners’ context. Our proposal to increase interaction in RRs includes one specific step named dissemination, although researchers often disseminate results to practitioners throughout the RR process, e.g., when sharing preliminary results. In Case-SoftSelection, some terms only used inside the company were unfamiliar to the researchers. While in Case-MLTest, the topic of testing

---

<table>
<thead>
<tr>
<th>Step</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prepare</td>
<td>Incorporate external experts to the review team when needed✓</td>
</tr>
<tr>
<td></td>
<td>Identify cultural aspects that can result in a positive environment for the RR✓</td>
</tr>
<tr>
<td></td>
<td>Involve practitioners with a research background or appreciation for research✓</td>
</tr>
<tr>
<td></td>
<td>Identify the expectations and motivations of the practitioners to conduct the RR✓</td>
</tr>
<tr>
<td></td>
<td>Develop consensus on a goal that brings value to both sides✓</td>
</tr>
<tr>
<td></td>
<td>Select topics interesting for both researchers and practitioners✓</td>
</tr>
<tr>
<td>Define RQ</td>
<td>Plan for a lot of initial interaction✓</td>
</tr>
<tr>
<td></td>
<td>Focus on the practitioners’ context✓</td>
</tr>
<tr>
<td></td>
<td>Hold the input meeting where practitioners present the problem and context✓</td>
</tr>
<tr>
<td>Search and selection</td>
<td>Get feedback on the preliminary results✓</td>
</tr>
<tr>
<td></td>
<td>Define small concrete outcomes✓</td>
</tr>
<tr>
<td></td>
<td>Be prepared to handle results that could be considered, in principle, negative or incomplete✓</td>
</tr>
<tr>
<td>Extract and synthesize data</td>
<td>Thematic analysis may help to overcome terminology and context gaps✓</td>
</tr>
<tr>
<td></td>
<td>Adapt the analysis to expectations and available literature✓</td>
</tr>
<tr>
<td>Disseminate results</td>
<td>Be aware of different terminologies✓</td>
</tr>
<tr>
<td></td>
<td>Find means and ways to have practitioners’ friendly communication✓</td>
</tr>
<tr>
<td></td>
<td>Translate the results✓</td>
</tr>
<tr>
<td></td>
<td>Remember that the guidelines suggest a flexible approach that can be adapted to the needs of each RR✓</td>
</tr>
<tr>
<td></td>
<td>Keep the RRs focused, rapid, and interactive✓</td>
</tr>
<tr>
<td></td>
<td>Take the opportunity to learn to work together✓</td>
</tr>
<tr>
<td></td>
<td>Meet, talk, and develop joint work sessions✓</td>
</tr>
</tbody>
</table>

Table 2

Recommendations for conducting RRs. ① Case-SoftSelection, ② Case-MLTest, and ✓ (derived from analysis).
machine learning was relatively recent in industry and academia and lacked standard terminology. To disseminate results under these scenarios, the review team needs to be aware of different terminologies and use understandable and relevant language to both parties.

Another aspect to highlight is the need to translate results. By this, we mean to make the results understandable in connection with the practitioners’ context. For example, in Case-SoftSelection, the key information was extracted, shared in joint work sessions, and summarized in presentations. The translation of the results is essential to ensure that the results are valuable and useful to the practitioners.

6.6. RR management

The following recommendations are based on what we observed worked well in Case-SoftSelection and Case-MLTest related to the RR process and working together. First, regarding the steps and activities, it is essential to remember that the guidelines suggest a flexible approach that can be adapted to the needs of each RR. Therefore, as we saw in Case-SoftSelection and Case-MLTest, the review team can adapt to keep a flexible approach and adapt the steps to the needs of the RR. Second, the questions that best fit the RRs are narrow, specific, and related to current problems faced by the practitioners. Selecting narrow questions relevant to both parties is essential to keep the RRs focused, rapid, and interactive. Third, we observed the advantages of having a shared repository and keeping track of the decisions made while conducting the review. The shared repository facilitated communication within the review team, and the memories supported the researchers when sharing the results and writing the academic papers. Fourth, researchers and practitioners should take the opportunity to learn to work together. This requires mutual understanding and respect and may not happen immediately. Finally, the researchers and practitioners in Case-SoftSelection and Case-MLTest recognized the importance of meeting, talking, and developing joint working sessions to foster knowledge exchange.

7. Discussion

This section discusses the study’s results and the implications of the findings for future research and practice. The discussion is organized as follows. First, we discuss the study’s results, specifically regarding conducting RRs with practitioners (7.1) and the benefits and challenges observed in our cases (7.2). Second, we discuss the implications of the findings for future research and practice (7.3). Third, we discuss the RRs proposal in terms of interaction, rigor, relevance, and flexibility, which are the key aspects of the RRs (7.4). Finally, we discuss the study’s limitations and the threats to validity (7.5).

7.1. Conducting RRs with practitioners

Our suggestions emphasizing practitioner involvement were designed to enhance the relevance of results, benefiting both researchers and practitioners. Our findings concerning practitioners’ positive perspectives and experiences align with prior studies about practitioners’ attitudes on evidence from RRs and secondary studies [16,41]. Practitioners value insights derived from research findings, especially when research questions address tangible, real-world challenges they face daily. In our cases, strong networking laid the foundation for future collaborative efforts and provided an avenue to evaluate the viability of extended collaborations. These opportunities are likely attributed to the close collaboration between researchers and practitioners, with the latter being actively involved throughout the process.

This study underscores the significant advantages of integrating practitioners into the RR process. Although there has been a marked increase in RRs published in software engineering conferences and journals, many employ the RR methodology primarily to streamline systematic reviews. Conversely, we advocate for RRs as a vehicle to foster collaboration between researchers and practitioners. Concerning logistical considerations, considering our suggestions for involvement and recommendations in this paper can facilitate more streamlined and efficient interactions between the two groups.

7.2. Benefits and challenges of RRs

We found that the RRs provided benefits individually for researchers and practitioners. Additionally, it contributed to building a relationship between them. On the individual level, researchers got chances to learn about the practitioners’ context and problems, besides getting an overview of the field from the practitioners’ perspective. This emphasis on practical relevance is identified the most in collaborative efforts [5]. On the other hand, practitioners get chances to take advantage of the research results and the researchers’ expertise, allowing them to get a broader view of the research area and develop an awareness of the state of the research.

Building long-term relationships is key to successful collaborations [42]. Therefore previous studies have highlighted the importance of relationship-building [43]. In this sense, the RRs provided opportunities for researchers and practitioners to meet, discuss, and exchange knowledge. These interactions allowed the researchers and practitioners to develop a shared understanding of the problem and the research area (Case-SoftSelection) and to build a shared vision of the problem-solution match (Case-MLTest).

The main challenge in the RRs was related to the availability of results, and some other minor challenges were how to organize the roles and the time and effort required. The availability of results was challenging in both cases, but with different representations. In Case-SoftSelection, the results got the initial impression of being outdated, and in Case-MLTest, the results were not directly applicable in the practitioners’ context. In both cases, the researchers and practitioners overcame the challenge by adapting the search and selection strategies and discussing the results. The early feedback from the practitioners helped the researchers to adapt the results to the practitioners’ context and therefore increase the relevance of the results. Since lack of relevance for practice is a critique of traditional literature reviews [22,41], researchers conducting systematic literature reviews could incorporate similar strategies to overcome this challenge.

Challenges related to the organization and roles could be overcome by clearly understanding each participant’s roles and responsibilities [5]. The RRs consumed little time in our two cases, but the time frame was distributed over a relatively long period. The RRs were not part of the main priority among the researchers’ and practitioners’ responsibilities. Besides, the RRs were not formally attached to a funded research project or a specific product/service in the industry. This lack of priority had a double effect. On the one hand, the RRs participants in the RRs were free to follow an exploratory approach. On the other hand, the RRs took longer than expected.

7.3. Implications for research and practice

As we mentioned in the introduction, our study implies a hypothesis that the relevance of literature studies will improve by involving practitioners in the process. However, only involvement is not sufficient. The essence is in giving practitioners an active voice, maintaining communication, allowing them to articulate their needs, and ensuring their insights shape the research focus. We found that conducting the RRs provided several occasions for the participants to meet, talk, and discuss. By focusing on topics directly related to the practitioners’ context and by actively integrating their voices into each step, we believe that the researcher’s efforts were more aligned with the practitioners’ needs compared to when the researchers worked on their own.

Our proposed approach for RRs suggests how and when, in the process, researchers and practitioners can interact. By identifying explicit roles and tasks for practitioners, we expected that RRs would offer a
higher degree of engagement from practitioners. This approach aligns with previous research in software engineering, which emphasizes the importance of industry collaboration and the use of appropriate research approaches to increase the relevance of research [40] and the importance of the context in the research process [12].

Given their smaller scale and lower level of required commitment, RRs are an initial step to estimate the feasibility of extended collaboration between researchers and practitioners. By undertaking these focused projects, both parties can familiarize themselves with each other’s working dynamics, methodologies, and interests. Additionally, conducting an RR early in the collaboration process can help determine whether the partnership is viable and whether both parties are willing to invest the necessary time and effort. For RRs to be successful, there should be a practical problem that practitioners are facing that researchers can help solve. Furthermore, the practitioners should be willing and interested in participating in the review process.

When initiating a research project, such as a Ph.D. project, conducting an extensive literature review is typical to gain an overview of the state of the art in the research area. Drawing insights from our cases, we see a way to complement this practice in the RRs by involving practitioners to capture their perspectives and insights. It is important to note that RRs should not be viewed as a substitute for systematic literature reviews but rather as a complementary approach. As described in our proposal [19], RRs prioritize knowledge exchange and context-awareness over covering the entire research area.

7.4. Interaction, rigor, relevance, and flexibility

We have determined that interaction, rigor, relevance, and flexibility are key aspects of the RRs that should be considered when conducting RRs. Our proposal of including practitioners in the RR goes beyond merely involving them in the process and doing tasks together. By assigning explicit roles and tasks for practitioners, we aimed for RRs to offer a higher engagement with practitioners. This serves dual purposes: to enrich dialogues and ensure that reviewed topics are relevant to practitioners. As a result, this approach facilitated mutual learning and set the stage for potential further projects.

Rigor embodies the scientific aspect of RRs. RRs aim to be systematic, meaning the search is not a random, ad-hoc search but follows a protocol to comprehensively cover the area’s literature. Balancing this in the context of RRs poses a challenge. Our observations from both cases highlighted this tension. While researchers appreciated the systematic nature of the RR, practitioners’ concerns were primarily about the direct applicability or relevance of the results. Including the practitioners’ voices made literature reviews more reflective of real-world scenarios, enhancing their practicality. The practitioners’ insights added depth and ensured the research questions and results were contextually relevant.

Flexibility in the RR pertains to the ability to adjust the process to the needs of the participants and the specific context. Such adaptability enabled RRs to meet specific case needs without losing focus on the main guidelines, proving their versatility in different research situations. In this paper, we presented two cases where the RR was tailored to the needs of the participants and the context. Although the foundational steps of the RR remained consistent, the actual implementation varied, from the search methodologies to how the results were presented.

7.5. Limitations of this study

In this study, we are observing and reflecting on the application of RRs in two cases of industry-academia collaboration. Conclusions are drawn based on interviews, observations, and the experiences of two researchers (also co-authors of this paper) as part of the review teams. We present no quantitative results and do not propose any causal models. Therefore, to reason about the validity of our conclusions, we apply the framework by Maxwell [44] comprising descriptive validity, interpretive validity, theoretical validity, generalizability, and evaluative validity.

Descriptive validity refers to the factual accuracy of the collected data. To achieve as accurate and complete data from interviews as possible, all researchers were involved in designing the interview protocol, two researchers conducted each interview, and interviews were recorded and automatically transcribed. In addition to data from interviews, observations were made by the first author in the initiation of the studies as well as by the second and third authors as participants in the studies. These observations may be biased by our different roles and pre-understanding of the RRs guidelines.

Interpretive validity refers to the researcher’s interpretation of the situation. In our case, it regards the interview situations. Not every nuance is captured in the interviews. To avoid misinterpretations, we let the interviewees read this manuscript. Regarding participatory observation, this threat is mitigated by the actual involvement in the cases.

Theoretical validity relates to interpretation or theorizing at a higher abstraction level. Our theoretical conclusions evolved through thematic coding, analysis, and writing this manuscript. All five authors were involved in both these activities, ensuring agreement among the researchers. To help the reader assess the theoretical validity, all steps of coding and interpretation have been transparently reported in this manuscript.

Generalizability A threat to our conclusions’ general validity is that we had RR experts (or at least access to them) in both cases. This means that we still do not know how feasible it is to implement the approach guided by the protocols alone. However, we provide examples to follow by describing how the RRs were conducted in those two cases. Furthermore, the relationships between industry and academia vary from place to place and between domains. It also depends on individual relationships between researchers and practitioners. Thus, the application of our findings may require adaptation in other situations. We still contend that the report’s general conclusions and recommendations can support other industry-academia collaborations, especially in the initiating stages.

Evaluative validity relates to our underlying values. Our recommendations are not neutral but based on assumptions about any envisioned stakeholder’s preferences. Although subjective, these assumptions are non-controversial (e.g., effective communication is good, producing relevant knowledge is desirable, and meeting the expectations of involved participants is good).

8. Conclusions

This paper provides a description of how two teams conducted rapid reviews. Both teams initiated the RRs with a broad topic, fostering a conducive interaction and knowledge exchange environment. During the planning phase, they interacted extensively to formulate research questions, identify inclusion/exclusion criteria, and establish expected outcomes. This involved aligning academic and industrial problem formulations through meetings and iterative discussions.

In the performing phase, practitioners trusted researchers to navigate research results and identify suitable articles. While practitioners did not participate actively in the detailed development of search and selection protocols, they played a crucial role in reviewing and commenting on the outputs, aiding the protocol development process, and ensuring the relevance and applicability of the findings.

The analysis effort was guided more by practitioners’ needs rather than focusing on the scope and relevance of the output. During the reporting phase, the results proved valuable within the immediate team, fostering spontaneous knowledge sharing and future study plans. Eventually, both reviews were documented and published as scientific contributions.
Our study identifies the benefits and challenges of conducting RRs with practitioners. Engaging practitioners in RRs yields several benefits. Firstly, it facilitates terminology alignment between researchers and practitioners, ensuring more precise communication and understanding. Secondly, these collaborations act as pilot initiatives, setting the stage for future joint efforts and partnerships. Additionally, practitioners gain an overview of the field through the RR process, enhancing their awareness and understanding of key issues and developments. Lastly, the outcomes of RRs are not just theoretical; they are actively used for various purposes, such as raising awareness among stakeholders and formulating new projects.

However, this collaborative approach also presents challenges. The level of involvement and expectations between practitioners and researchers need careful management and alignment to avoid misunderstandings and dissatisfaction. There might also be instances where the RR outcomes do not fully match the needs and expectations of the practitioners, necessitating adjustments and refinements to the review process.

Furthermore, we provided recommendations for researchers and practitioners when conducting RRs. The recommendations derived from the experiences and insights from the two case studies are intended to support and enhance the conducting of RRs by researchers and practitioners in the future, serving as a supplement to existing guidelines.

In future research, we plan to collaborate with researchers and practitioners to conduct more RRs and study the impact of RRs on the research and practitioners’ context. Overall, as seen in the cases in this study, fostering an exchange between researchers and practitioners is a promising way to increase the relevance and applicability of software engineering research.

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

Data will be made available on request.

Acknowledgments

This work has been supported by ELLIIT, a Strategic Area within IT and Mobile Communications, funded by the Swedish Government. The work has also been supported by a research grant for the GIST project (reference number 20220235) by the Knowledge Foundation in Sweden. We would like to thank the researchers and practitioners who participated in the interviews for their time and valuable insights. Additionally, we extend our gratitude to the anonymous reviewers who participated in the interviews for their time and valuable insights.

The work has also been supported by a research grant for the GIST project (reference number 20220235) by the Knowledge Foundation in Sweden. We would like to thank the researchers and practitioners who participated in the interviews for their time and valuable insights.

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

Data will be made available on request.

Acknowledgments

This work has been supported by ELLIIT, a Strategic Area within IT and Mobile Communications, funded by the Swedish Government. The work has also been supported by a research grant for the GIST project (reference number 20220235) by the Knowledge Foundation in Sweden. We would like to thank the researchers and practitioners who participated in the interviews for their time and valuable insights. Additionally, we extend our gratitude to the anonymous reviewers who participated in the interviews for their time and valuable insights.

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.