Comparing a Rapid Review with a Systematic Review in the Software Engineering Field

Carolline Pena[a,*], Bruno Cartaxo[b], Igor Steinmacher[e], Deepika Badampudi[f], Deyvson da Silva[a], Williby Ferreira[a], Adauto Almeida[c], Fernando Kamei[d] and Sergio Soares[d]

[a] *Universidade Federal de Pernambuco (UFPE), Recife, Pernambuco, Brazil*
[b] *Instituto Federal de Pernambuco (IFPE), Paulista, Pernambuco, Brazil*
[c] *Universidade de Pernambuco (UPE), Garanhuns, Pernambuco, Brazil*
[d] *Instituto Federal de Alagoas (IFAL), Maceio, Alagoas, Brazil*
[e] *Northern Arizona University, Flagstaff, Arizona, Unites States*
[f] *Bleking Institute of Technology, Karlskrona, Sweden*

ARTICLE INFO

Keywords:
Rapid Review
Systematic Review
Software Engineering

ABSTRACT

Context: Rapid Reviews (RRs) are secondary studies aiming to deliver evidence to experts in a more timely manner and with lower costs than traditional literature reviews. Previous studies have shown that experts and researchers are positive toward RRs. However, little is known about how RRs differ from traditional Systematic Reviews (SRs).
Objective: The goal of this paper is to compare a RR with an SR study in terms of their methods (e.g., search strategy, study selection, quality assessment, data extraction, and reporting of findings, in order to understand how optimizing the traditional SR method (to RR) impacts review outcomes.
Method: To achieve this goal, we conducted an SR with the same research questions answered by a pre-existing RR and compared those two studies.
Results: The RR lasted six days, while the SR took one year and two months. The main bottlenecks we identified in the SR are (i) executing the search strategy and (ii) the selection procedure. Together they took ten months. The researchers had to analyze the information from 11,383 papers for the SR compared to 1,973 for the RR. Still, most (~78%) of the papers included in the SR were returned by the RR search, and some papers that could be included were unduly excluded during the RR's selection procedure. Both secondary studies identified the same number of pieces of evidence (30), but the pieces of evidence are not the same.
Conclusion: The SR time frame has shown to be expensive in the experts' context, and the RR was able to identify a considerable number of relevant findings. We do not prescribe RR as an alternative to SR or visa versa. Instead, we believe that RRs and SRs are viable options, and researchers should select the appropriate method based on the research aims and constraints.

1. Introduction

Evidence-Based Software Engineering (EBSE) was introduced in the research community to integrate the best research evidence with practice [1]. Moreover, it is argued that the knowledge aggregated with secondary studies is the most appropriate to be transferred to practice [2]. However, some researchers observed that traditional secondary studies — hereafter Systematic Reviews (SRs) — lack connection with Software Engineering (SE) practice due to many reasons, for instance, the amount of time and effort to conduct them [3, 4], thereby not producing timely results for experts.

Rapid Reviews (RRs) [5] are one of the alternatives to reduce the gap between research and practice. RRs are secondary studies that provide scientific evidence to support experts in making evidence-informed decisions, and usually are conducted in a wider knowledge transfer initiative [6, 7]. RRs adapt SRs' method to convey evidence through more appealing mediums [8, 9], reduced time frames, and lower costs. To make that possible, RRs optimize SRs method by adhering to strategies such as: limiting search sources, having just one person to screen studies, skipping quality appraisal, among others [10].

Although more than 100 RRs are reported in the medical field [11, 12], the SE community has not adopted RRs as extensively. In 2018, Cartaxo et al. [5] reported the first RR in SE. Since then, SE researchers have used RRs on different SE topics, such as Amalfitano et al. [13] explore aspects related to testing of context-aware software systems; De Souza et al.'s [14] investigate how to capture and represent smartness regarding IoT software systems; Ponce et al. [15] overlook strategies to migrate from a monolithic architecture to microservices. Meanwhile, there are also initiatives to evaluate how RRs could be applied in different scenarios. For instance, Pizard et al.'s work [16] about using RRs for EBSE training, as well as another paper from the same group, about the adoption of RRs in governmental agencies [17]. In addition, Rico et al. [18] proposed a guideline for conducting Interactive Rapid Reviews (IRR) in software

*Corresponding author
**Principal corresponding author
cdp@cin.ufpe.br (C. Pena); email@brunocartaxo.com (B. Cartaxo); igor.Steinmacher@nau.edu (I. Steinmacher); deepika.badampudi@bth.se (D. Badampudi); dls4@cin.ufpe.br (D.d. Silva); wsf2@cin.ufpe.br (W. Ferreira); adauto.filho@upe.br (A. Almeida); fernando.kenji@ifal.edu.br (F. Kamei); scbs@cin.ufpe.br (S. Soares)
https://brunocartaxo.com (B. Cartaxo)
ORCID(s):

engineering, shifting from a focus on technology transfer to knowledge exchange.

In this context, we believe it is important not only to have more RRs in SE but also to investigate their intricacies. For instance, a recent study [19] explored the perceptions of the SE research community regarding RRs. One of the findings is that a large group of researchers might adopt RRs as a knowledge transfer method. However, researchers with such a point of view emphasize the importance of studies contrasting RRs with SRs. They argue that it is necessary to understand the impact of optimizing the SRs' method and enable an informed decision on whether or not to conduct a RR and in which context.

Therefore, the **GOAL** of this research is to compare a Rapid Review with a Systematic Review in terms of their methods and findings in order to understand how optimizing the traditional Systematic Review method impacts what we obtain with Rapid Reviews.

To achieve this goal, we conducted an SR with the same research questions of a RR we have previously published [5] and performed the comparison between those two studies. In this paper, we report that comparison, not the SR itself. In addition, we conducted a survey with experts from industry and academia to evaluate the relevance of the findings we obtained with both the secondary studies.

We acknowledge that the purpose of both methods (SR and RR) are different and complementary [5]. Thus, the intention is not to allege that one is better than the other. Also, all claims we make throughout this paper should be taken with due care. As mentioned earlier, there is a need to compare RR and SR methods, and this study is the first step toward understanding the RR or SR methods selection.

Although there are some studies comparing RRs with SRs in medicine [20, 21, 22], this is the first study in the SE field. We also dived more in-depth with our comparison – enabling us to present more detailed insights – when considering the comparisons between RRs and SRs in health-related research fields. Section 2 details how our comparison is more thorough than the ones in medicine.

The main **contributions** of this paper are:

- The first comparison of a RR with an SR in the Software Engineering field. This might be useful for researchers who want to decide which approach to follow, considering their context.
- The results of a survey with experts and researchers about the relevance of the findings from a SR and a RR.
- An approach to compare RRs with SRs. This might be a useful guide to anyone wanting to conduct such kind of comparison in the future.

2. Previous Work Comparing SRs with RRs

To the best of our knowledge, this is the first paper reporting a comparison between an SR and a RR in the SE context. In this section, we present RR and SR comparative studies conducted in health research fields.

Taylor-Phillips et al. [21] conducted a study to compare an SR to two types of RRs, one with all steps optimized and another with fewer steps optimized. The comparison focused on the difference between the studies included in each secondary study and the difference between their findings. The effort to achieve the results also was considered by the number of studies returned by the search. They could not find differences between the SR and the RR with fewer steps optimized. However, the RR with all steps optimized had increased biases and could not identify papers that influenced the findings.

Watt et al. [23] compared the conclusions produced by RRs and SRs. The purpose was to identify how RRs are conducted in medicine compared to SRs on the same topic. The time frame and the conclusions were taken into consideration. The study claims that the results were not essentially different.

On the other side, Van de Velde et al. [22] compared the findings of their RR with an SR conducted by a different research group that published on the same topic, the medicinal use of potato-derived products, and conflicting results were observed between the secondary studies.

The SR identified two trials about potato peels in treating burns but concluded that honey dressings are better than potato peel dressings. The RR found additional evidence indicating that potato peel dressings are better than gauze alone. The comparative study concludes that although potato peel dressings as not applicable to Western countries' settings can be a good alternative for emerging countries.

The comparisons between RRs and SRs in medicine focus mainly on the differences between the findings they derived and the studies they included. In this research, we take a step further and dive into detail, comparing each aspect of an RR with an SR. More importantly, we mapped each secondary study's decisions and data to explain the reason for each difference between those two studies. For instance, we could track that a specific finding was identified in one secondary study but not in the other due to a particular term in one of the search strings. Similarly, we could determine that a specific finding identified in one secondary study was not found in the other because the paper that provided that finding was excluded on the first step of the selection procedure.

In addition to the fact that our comparison approach is more in-depth than those in medicine, it is also relevant to highlight that RRs and SRs in SE may differ considerably from those in health research fields.

In the SE context, secondary studies tend to be more qualitative-driven. Most SE research needs to consider human aspects inherent to the software development process. Therefore, the synthesis procedure, and consequently the results, tend to suffer from biases related to researchers' subjectivity. This usually does not happen in medicine, where secondary studies tend to synthesize evidence from randomized controlled trials, following meta-analysis principles.

Regarding RRs, one difference between SE and medicine is their time frames and criticality level. RRs in SE are usually applied in an ever-changing context and, in general, not in a critical setting. This is the opposite of what occurs in medicine. Cartaxo et al. [10] mentioned that a RR in SE can be conducted in a few weeks, while in medicine, it usually takes a few months.

Those differences directly influence the way RRs are conducted in SE and medicine. That is why we believe a comparison like the one we present in this paper is important, even considering that there are already other comparisons in health research fields.

3. Method

We conduct this research in three main steps. First, we perform an SR with the same scope as one of our preexisting RRs. Then, in the second step, we compare those two secondary studies regarding the method they followed and the findings they obtained, both from qualitative and quantitative perspectives. And finally, in the third and last step, we conduct a survey with experts from industry and academia, aiming to understand their perceptions concerning the relevance of the findings we obtained with both the RR and the SR.

3.1. The Rapid Review

At the time we conducted this research, there were few RRs published in SE peer-review venues. So, we decided to select an RR conducted by our research group [5]. By following this strategy, we also have access to RR details that enable us to make a richer comparison.

The RR was conducted in the context of a software project in a software development company. The project team adhered to agile practices but was facing problems regarding low customer collaboration. The RR addressed that situation, providing evidence that helped the experts to better understand their problem and, especially, providing strategies to improve customer collaboration.

3.2. The Systematic Review

Since our goal is to compare an SR with a RR, not only regarding their findings but also their methodologies, we started from the same research questions answered by the RR, to make both studies comparable. Nevertheless, from that point on, the SR follows a separated path. To guarantee such independence, the leading researcher of the SR did not read or had any contact with the RR in any way.

We decided that the SR and the RR should follow a different path as early as before creating their protocols. The rationale behind this decision is two-folded. First, although we adopted the same research questions on both the RR and the SR, the SR protocol was created by a researcher who was not part of the RR team to avoid bias due to the learning effect. Second, creating the SR protocol from scratch also reveals what can happen differently when one has less time pressure and no close contact with the experts creating the protocol. The opposite of what happened with the RR's leading researcher when creating the RR's protocol.

Since the goal of this study is not the SR itself but its comparison with the RR, we do not present the entire protocol of the SR here. Although, in Section 4 we discuss the SR steps in detail.

3.3. Comparing the RR with the SR

To conduct the comparison between the two secondary studies, we analyzed each aspect described in the SRs' [24] and RRs' [10] guidelines. We also take into account the comparisons between SRs and RRs in the health field. In summary, we compared the aspects listed in Table 1.

3.4. Survey with Experts

After obtaining the SR results, we administered a survey aiming to investigate how relevant the findings of the SR and RR are from the perspective of specialists from industry and academia. A draft of the survey was created by three researchers and administered as a pilot to the other three researchers, which provided us with comments to improve it.

The recruitment was based on convenience sampling [25], inviting potential participants via email. The participants were informed that the survey was anonymous, for research purposes only, and they could withdraw their participation anytime they wish [26]. Then, we sent the online survey in Google Forms format to the participants.

The participants were also asked to invite any colleague who would fit the required profile. The profile we searched for is either a person working in the industry who has software development experience. Altogether, 39 participants answered our survey.

The survey is composed of two sections. In the first section, there are demographic questions to identify participants' age, level of educational attainment, years of professional experience, current job role, and if they currently work in academia, industry, or both. We decided to keep the demographic section as short as possible, since the second section is somewhat extensive, although very straightforward.

In the second (main) section, we provide a brief context about the problem and the four research questions that drive both our RR and SR, which are related to low customer collaboration in software development teams. Then, we ask the participants to classify, on a five-point Likert scale, how relevant is each piece of evidence we have found in the two secondary studies. For example, one of RR's pieces of evidence that answers RQ1, about the benefits of good customer collaboration, is: *It leads to greater productivity*. Another example is SR's piece of evidence that answers RQ2, concerning problems caused by low customer collaboration: *It causes misunderstandings*.

Table 1
Aspects analyzed to compare the Systematic Review with the Rapid Review.

Aspects	Items
Motivating Problem	The problem that motivated conducting the secondary studies.
Protocol	The procedures documented in the protocol and the level of detail.
Time Frame	The time frame demanded to conduct the secondary studies, and of each of their procedures.
Search Strategy	The search sources; the structure and terms of the search strings; the time span of the searched papers; and the number of papers found with the searches.
Study Selection	The steps composing the selection procedure; how many and who is involved in each step; the number of studies excluded in each step; how many and which studies are included in the end; and, in which step a study included in one secondary study was improperly excluded from the other.
Quality Appraisal	How the quality appraisal was conducted; who was involved in this procedure; and the number of papers excluded due to quality criteria.
Data extraction	How data extraction procedure was conducted and who was involved in it.
Synthesis	How the synthesis procedure was conduced, and who was involved in it.
Findings	The findings and the reasons for some findings to show up in one secondary study but not in the other.
Report	The medium which the findings are presented and the level of detail.

4. Results Comparing the RR with the SR

In this section, we present the convergences and divergences comparing the two secondary studies regarding the method they followed and the findings they obtained, both under qualitative and quantitative perspectives.

4.1. Motivating Problem

The problem that motivated the conduction of the RR emerged from a software development project in a software company. The experts reported having problems related to low customer collaboration. They would like to have access to evidence that could mitigate that problem, which in turn motivated the conduction of the RR. Prior to conducting the RR, the researchers contacted representatives of the development team to discuss problems related to low customer collaboration and to obtain the entire context related to the problem faced by the organization.

We followed the same motivation for conducting the SR. We conducted the SR with aiming to find evidences about issues about customer collaboration in software development. We also wanted to know if the set of findings would be similar or different to the RR findings.

4.2. Research Questions

The RR's research questions were elaborated in close collaboration with experts aiming to provide evidence to help them make informed-decisions in practice. The research questions are:

- **RQ1**: What are the benefits of customer collaboration in software development practice?
- **RQ2**: What are the problems caused by low customer collaboration in the software development practice?
- **RQ3**: What are the challenges in establishing customer collaboration in software development practice?
- **RQ4**: What are the strategies to improve customer collaboration in software development practice?

The SR's research questions are the same as the RR's, aiming to enable a full comparison between them. So, the research questions were our controllable variable.

4.3. Protocol

The RR's protocol[1] has the following main sections: practical problem, research questions, search strategy, selection procedure, quality assessment, extraction procedure, synthesis procedure, and review report. Although there is a specific section for quality appraisal, it only declares that no quality appraisal was conducted. The document has succinct descriptions of each step, and lacks detailed information on how the procedures were performed as well as theoretical justification supporting its methodological decisions. Because of this, it has three pages. A single researcher created the RR protocol and the experts reviewed it to guarantee the RR would be aligned with their needs.

The SR's protocol[2] includes the following main sections: introduction, background, research questions, search strategy, selection process, quality analysis, data extraction, data synthesis, and limitations. The document details each step of the SR method as recommended by Kitchenham et al. [1, 24]. The research decisions are backed by citations to scientific studies.

[1] RR's protocol: https://bit.ly/3FzwiwZ
[2] SR's protocol: https://bityli.com/xCaJdqFm

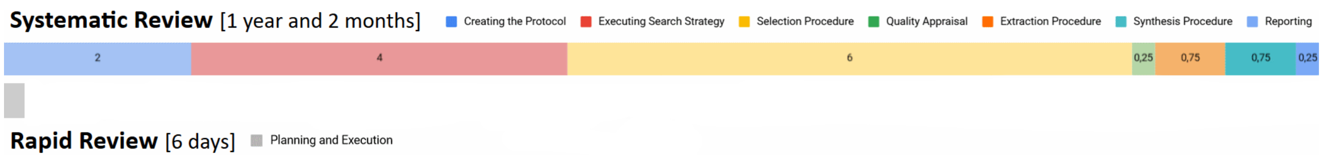

Figure 1: Time frames of both the secondary studies. The numbers in the systematic review time frame are the duration of each activity in months. The rapid review has no detailed time frame for each activity due to its low duration and tangled nature.

4.4. Stakeholders Role

Although only one researcher led the entire RR process, it was conducted in close collaboration with experts who were immersed in a practical problem. This is not usually observed in SRs [4]. In addition, the protocol of the RR does not describe how many experts collaborated with the RR. Thus, we need take a look at the paper that reports the RR [5], which says that five experts were involved.

In the SR protocol, on the other side, we could find only bibliographic discussions aiming to justify the SR execution. In addition, to avoid bias, nine researchers participated in the paper selection procedure, and some other steps (e.g., quality analysis, extraction procedure, and synthesis procedure) were revised by at least one researcher.

Both researchers who led each of the secondary studies are from the same software engineering research group.

4.5. Time Frame

Regarding the duration of the two secondary studies, the results are quite contrasting. Figure 1 summarizes the time frame of both the SR and the RR.

The RR's planning, execution, and reporting took approximately six days of dedication of a single full-time researcher. The process was so short, and the activities so tangled in one another that it was not possible to measure the duration of each RR activity independently.

The SR's planning, execution, and reporting spanned a year and two months. The SR's leading researcher has worked continuously on it, but the dedication time was not tracked fine-grained enough to measure the effort in hours or days, as with the RR.

Although we could not investigate the time frames of each RR's steps in detail, we could do that for the SR. This is important because it shows the main bottlenecks of the SR method, revealing why such a study takes so much time and effort when compared to a RR. Following we detail the duration of each of the SR's steps:

- **Creating the Protocol [2 months]**: We created the SR protocol from scratch, not based on the RR's one. The first author defined and edited the protocol, which has been revised several times by three researchers, two senior researchers and a PhD student. The various back and forth inherent to the revision process and the researchers busy agendas are the main reasons for the time we spent creating the protocol.
- **Executing the Search Strategy [4 months]**: Three months and three weeks for manual search, and one week for automatic search. This period includes extracting the metadata of all found papers and organizing it on the selection spreadsheets. Such activity is very tedious and time-consuming. The manual search demands more effort because one has to extract the metadata of each paper, one by one. The automatic search, on the other hand, could be partially supported by some search engines that enabled us to extract search results metadata in a very straightforward way. The first author executed this steps under the supervision of a senior researcher.
- **Study Selection [6 months]**: The SR's first selection step (i.e., title and abstract) lasted three months and a half. Although analyzing just papers' titles and abstracts is straightforward,the high number of papers (11,973) returned by the search sources prolonged this step. The second selection step (i.e., complete paper content) lasted two and a half months. One of the major causes of delay occurred due to the dependency on researchers outside the research group to conduct the selection procedure in pairs. It involved seven researchers with limited availability. Finding a common schedule for the conflict meetings with the pairs was difficult.
- **Quality Assessment [0.25 month]**: This step was quite straightforward, taking just one week to analyze 63 papers. This step was reviewed by a senior researcher.
- **Data Extraction [0.75 month]**: It took three weeks, and it includes creating and filling a spreadsheet with the data extracted to answer each of the SR's research questions from the 63 included papers.This step was reviewed by a senior researcher.
- **Synthesis [0.75 month]**: It took three weeks to apply the thematic analysis and synthesize the extracted data. This step was reviewed by a senior researcher.
- **Reporting [0.25 month]**: It took one week to write down the document reporting the results of the SR. This step was reviewed by a senior researcher.

4.6. Search Strategy

The RR adopts a single strategy (automated search only) and a single search source (one search engine only) approach. The goal is to make the search process more straightforward. The single search source is the Scopus search engine.

The SR adopts a multi-strategy and multi-source search approach taking into account Kitchenham's guidelines [24]. It is based on both automated and manual searches. We used various search engines (Scopus), digital libraries (IEEExplorer Digital Library, ACM Digital Library, Springer Link),

Table 2
Secondary Studies Search Strings.

RR Search String	SR Search String
(customer OR "product owner" OR stakeholder) **AND** (collaborat* OR participat* OR cooperat* OR relation* OR involvement OR engagement) **AND** ("software development" OR "software engineering" OR "software project")	(customer OR client OR "product owner") **AND** (collaboration OR participation OR contribution OR cooperation) **AND** ("software engineering" OR "software development")

Table 3
Secondary Studies Search Sources

RR Search Sources	SR Search Sources
Automatic Search Scopus	**Automatic Search** ACM DigitalLibrary, IEEExplorer Digital Library, Scopus, Springer Link **Manual Search** EASE, EMSE,ESEM,ICSE,IST

journal issues (EMSE - Empirical Software Engineering Journal and IST - Information and Software Technology), and conference proceedings (ESEM - Empirical Software Engineering and Measurement Conference, EASE - Evaluation and Assessment in Software Engineering Conference, ICSE - International Conference on Software Engineering). The intention is to have exhaustive coverage.

Table 3 presents the sources used in the search strategy for each one of the secondary studies. Although the SR took place a year and a half after the RR, they both covered studies published on the same time span. This tries to guarantees a fair comparison. We are aware that more search sources may generate more evidence. However, an RR proposes the omission/abbreviation of traditional SR steps. And, our primary goal is identifying the difference between both and what we are losing or gaining from this omission/abbreviation.

The RR's and SR's search strings are also different. However, none of the convergences and divergences between them are intentional. The protocols were defined independently, as we already mentioned. When two different people conduct two secondary studies about the same topic, separately, it is expected that parts of the protocol are somewhat different, including the search strings. That kind of scenario is exactly what we want to reproduce.

If we used the same RR search string on the SR, it would neither be a fair nor a real scenario comparison. RR's search string is supposed to be defined in close collaboration with experts on a tight schedule. SR's search string, on the other side, is usually defined based on the author's expertise and with plenty of time for calibration and pondering. Those are two very different settings.

Even though, the RR and SR occurred to have the same main terms (<customer> AND <collaboration> AND <software engineering>), the synonyms for each term are slightly different. Table 2 presents the strings of the secondary studies used to search in the automated search sources.

The RR's search process returned 1,973 papers, while the SR returned 12,297 papers (11,383 non-duplicated), of which 3,601 came from the automatic search and 8,696 from the manual search, without excluding duplicates. The difference between the RR and SR is very significant, but most of the studies found in the SR came from the manual search. This happens because, to have high coverage, we considered all studies published in all journal issues and conference proceedings of the five SR's manual search sources.

4.7. Selection Procedure

Figure 2 depicts the selection procedure of the RR and the SR side-by-side.

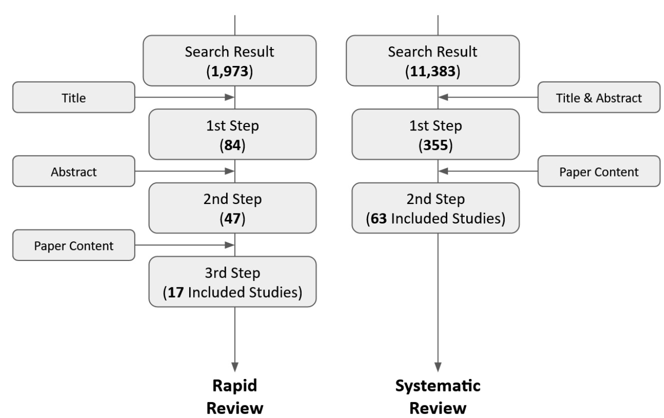

Figure 2: Selection procedure of both the secondary studies.

The RR selection procedure was conducted by a single researcher in three steps. In the first step, the researcher analyzes only the title of the 1,973 papers returned by the search, leaving 84 papers after excluding those clearly out of the RR's scope. In the second step, the researcher excludes the papers that remained from the previous step based on their abstracts, leaving 47 papers. From those papers, five could not be found. The RR's leading researcher did not try to contact the papers' authors due to RR's time constraints. So, those five papers were excluded before the next step. Finally, in the third and last step, the researcher excludes the remaining papers by reading their entire content, leaving only papers that were included in the RR, which are 17.

In addition, no quality appraisal is conducted in the RR; consequently, no paper is excluded due to it.

On the other hand, the SR selection procedure comprised two steps plus the quality appraisal. In the first step, we analyzed the title and abstract of the 11,383 non-duplicated papers, resulting in the selection of 355 papers.

In the first selection step of the SR both the papers' titles and abstracts were analyzed. In theory, that would reduce the probability of unduly excluding papers, in contrast to what may occur in the first step of RR's selection procedure, which selects based on the paper's title only. That assumption has proved to be true, at least in our case; we discuss the results in Section 6. This first selection step also passed through a validation involving four researchers grouped into three pairs. The first author paired with the others. Due to the high number of studies, we conducted the validation in pairs for 10% of the papers (1,139), which were randomly chosen. We verified the agreement of pairs with Kappa statistics [27], obtaining two pairs with a moderate agreement and one with a substantial agreement. The remaining papers were conducted by the first author only.

Similar to the RR study, we also could not find the documents with the full content of some of the SR's papers, 29 to be more precise. However, unlike with the RR, in the SR study, we tried contacting the papers' authors. We have succeeded in obtaining 12 out of the 29 papers.

In the second selection step of the SR, we conducted the paper analysis fully in pairs to reduce selection bias. The papers were excluded analyzing their entire content, resulting in the selection of 63 papers.

The next step (quality appraisal) does not exactly fit in the selection procedure. No paper was excluded, which makes the SR end up with 63 included studies.

Now we explore how the RR and the SR converge and diverge regarding their included papers. Figure 3 depicts a Venn Diagram to represent the quantitative of both the secondary studies' included papers.

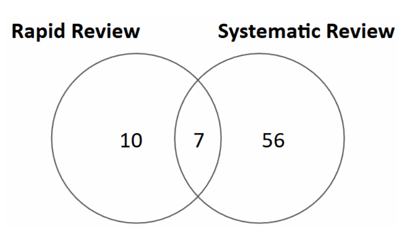

Figure 3: The number of papers included in the rapid and the systematic review.

One would probably expect that the RR's included papers would be a subset of the SR's. However, it has been shown to be false in our case. Following, we explore the reasons why some papers (10) were included in the RR but not in the SR:

- **Absence of terms on the SR's search string [8 papers]**: Most of the papers that were included in the RR but not in the SR occurred due to the absence of the term "stakeholder" on the SR's search string. That term is present on the title and/or abstract of those papers, that were not even returned by the SR search procedure.
- **Selection bias during the 1st step of the SR's selection procedure [1 paper]**: Only one paper was excluded from the SR during the first selection procedure step (title and abstract) but included in the RR. The only reason we see to explain that is selection bias.
- **Selection bias during the 2nd step of the SR's selection procedure [1 paper]**: Just one paper excluded in the SR during the second selection procedure step (paper content) but included in the RR. The reason to explain that is selection bias, as in the item before.

Now we explore the other way around. The reasons why some papers (56) were included in the SR but not in the RR:

- **Absence of some search sources on the RR [14 papers]**: Those papers were not even returned by the RR search procedure based on Scopus only. In the SR they emerged from other search engines, are ACM Digital Library, IEEExplorer, and SpringerLink.
- **Selection bias during the 1st step of the RR's selection procedure [38 papers]**: Most of the papers included in the SR but not in the RR were returned by the RR's search source. However, they were excluded during the first phase of the RR's selection procedure (title only) due to selection bias. Excluding papers based on title only may have influenced this number.
- **Selection bias during the 3rd step of the RR's selection procedure [4 papers]**: The reason to explain these unduly exclusion in this step (paper content) is selection bias.

4.8. Quality Appraisal

Although there is a section about quality appraisal in the RR protocol, it only states that it does not conduct such kind of procedure.

In contrast, we conduct a quality appraisal of the primary studies in the SR. We followed the 11 criteria method by Dyba and Dingsoyr [28]. The studies were given scores for each criterion. The quality appraisal was performed by a single researcher and reviewed by another one. In the end, no paper was excluded based on the quality assessment.

4.9. Extraction Procedure

The RR extraction procedure was performed by a single researcher. The extraction spreadsheet has the following information about the included papers: metadata about the papers; and evidence from the papers that answers the research questions.

The SR extraction procedure was performed by a single researcher and reviewed by another one. The SR's extraction spreadsheet has not only the same data of the RR's — metadata about the papers and evidence from the papers that answer the research questions — but also methodological and contextual data of the included papers.

4.10. Synthesis Procedure

Both the RR and the SR adopted thematic analysis as a synthesis method. This is just a coincidence since the methodological decisions of the secondary studies are independent. The nature of the collected data (i.e., qualitative) also reduces the space of possibilities. The RR synthesis was conducted by a single researcher only. In contrast, the SR synthesis was conducted by a single researcher and reviewed by another one.

Despite both secondary studies adopting the same synthesis method, the guidelines they followed are different. The RR adopted Fereday and Cochrane's [29] guidelines, which define the following steps: familiarization with the data, creation of initial codes, search for themes, review of themes, and construction of the report. While the SR adopted the guidelines of Cruzes et al. [30] and Braun and Clarke [31], which comprises the following steps: reading data, identifying codes, forming themes, reviewing themes, and subsequently, building theme descriptions.

4.11. Report

The RR's findings are reported through an Evidence Briefing [32], which is a one-page document that presents the main findings of an empirical research. The goal of such kind of medium is to present the results to experts more appealingly and straightforwardly compared to the traditional research paper format. In addition, there was a workshop together with the experts involved in the RR to present the Evidence Briefing and discuss the findings. According to them, the experts mentioned that, although the Evidence Briefing was easy to read, they could not understand some points just reading it. However, those points became clear during the workshop discussion. Still, regarding the RR's Evidence Briefing, it has a URL to access the list of primary studies that composed the RR. Nonetheless, there is no way in the Evidence Briefing to know from which primary study a specific finding comes from, which is a downside.

4.12. Findings

In this section, we present how the findings of the RR and the SR converge and/or diverge, and why. To explore the differences, we have to mention some of the concepts related to the topic of the secondary studies, which is "customer collaboration" in software development projects. However, we do not dive into details — as the RR and SR do — since this is not the goal of this paper.

The RR identified 30 findings, summing up all the findings that answer each of the four research questions. The SR, coincidentally, has identified the same number of findings. However, the findings are not all the same. 12 out of the 30 RR's findings are not identified in the SR, while 16 out the 30 SR's findings are not identified in the RR.

The relation between the secondary studies' findings is N-N. So, a finding of one of the secondary studies may correspond to more than one finding of the other. To illustrate, take a look at Figure 7.

Another peculiarity is that, in some cases, a finding does not immediately seems to be covered by another. For example, take a look on Figure 7. The finding "Risk assessment", which answers RR's RQ4 about strategies to improve customer collaboration, matches the SR's finding "Use of agile methodology". They do not look the same at first sight. But the latter comprises the former when reading the SR's findings in detail, which is out of the scope of this paper. This occurs because, in general, the findings of the SR are more high level compared to the RR's ones, sometimes covering more than one RR finding. In Section 6 we discuss this occurred.

Following there the reasons why some findings were identified on one secondary study but not on the other, and vice-versa:

- **Divergence due to the search strategy [RR 1 finding and SR 5 findings]**: In this case, the papers that originated the findings on one secondary study were not even returned by the other's search. This is a coverage problem on the search strategy. Either due to differences between search sources or search strings. This situation occurred with five findings that were identified in the RR, but not on the SR. While just one finding was identified in the SR, but not on the RR for that same reason.
- **Divergence due to the selection procedure [RR 5 findings and SR 1 finding]**: In this case, the papers that originated the findings on one secondary study were returned by the other's search. However, those papers were unduly excluded during the selection procedure. This situation occurred with just one finding that was identified in the RR, but not on the SR. On the other side, 11 findings were identified in the SR, but not on the RR for that same reason.
- **Divergence due to the synthesis procedure [RR 4 findings and SR 6 findings]** : In this case, the papers that originated the findings on one secondary study were returned by the other's search, and even included after the selection procedure. However, the findings were not identified in the synthesis procedure of the latter. This situation occurred with five findings that were identified in the RR, but not on the SR. While four findings were identified in the SR, but not on the RR for that same reason.

Following we present, in more detail, the differences between the findings of the RR and the SR per each research question.

Differences between the findings of the secondary studies regarding RQ1 - Benefits of customer collaboration

The RR identified nine benefits of customer collaboration, while the SR found six. Three benefits were identified in both the secondary studies. The others are disjoint. Figure 4 shows the benefits identified on both the RR and the RR, side-by-side.

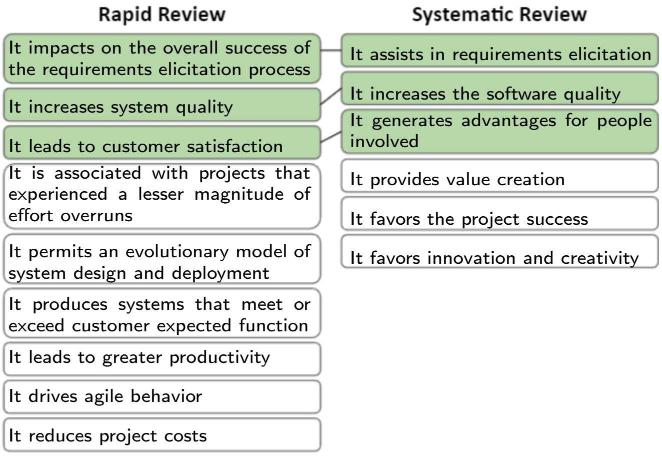

Figure 4: Comparison of customer collaboration benefits (RQ1) identified on the secondary studies.

There are two reasons why some findings that answer RQ1 were identified in the RR but not in the SR: divergence due to the search strategy (4 findings), i.e. the search string could not find one or more papers that originated the findings; and divergence due to the synthesis procedure (2 findings), i.e. the papers that originated the findings were found and included, but the findings could not be identified during the synthesis procedure. While, on the opposite side, the reasons why some findings that answer RQ1 were identified in the SR but not in the RR are: divergence due to the selection procedure (3 findings), i.e. papers were unduly excluded in the first selection step.

Differences between the findings of the secondary studies regarding RQ2 (Problems caused by low customer collaboration)

The RR identified five problems caused by low customer collaboration, while the SR found six. Four problems identified on the SR are not covered by the RR, while there are only two the other way around. Figure 5 shows the problems identified on both the RR and the RR, side-by-side.

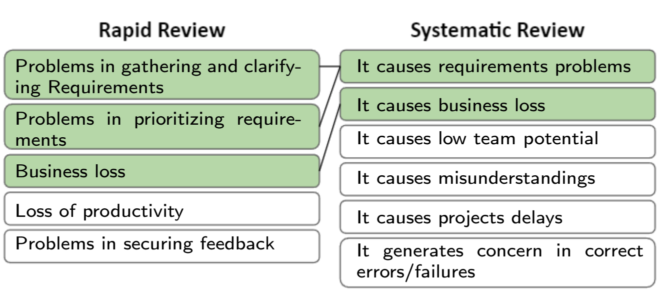

Figure 5: Comparison of problems caused by low customer collaboration (RQ2) identified on the secondary studies.

The reason some findings that answer RQ2 were identified in the RR but not in the SR is the divergence due to the synthesis procedure (2 findings). While the reason some findings that answer RQ2 were identified in the SR but not in the RR is the divergence due to the selection procedure (4 findings).

Differences between the findings of the secondary studies regarding RQ3 (Challenges to establish customer collaboration)

The RR identified five challenges to establish customer collaboration, while the SR found eight. Three challenges were identified in both the secondary studies. Figure 6 shows the challenges identified on both the RR and the RR, side-by-side.

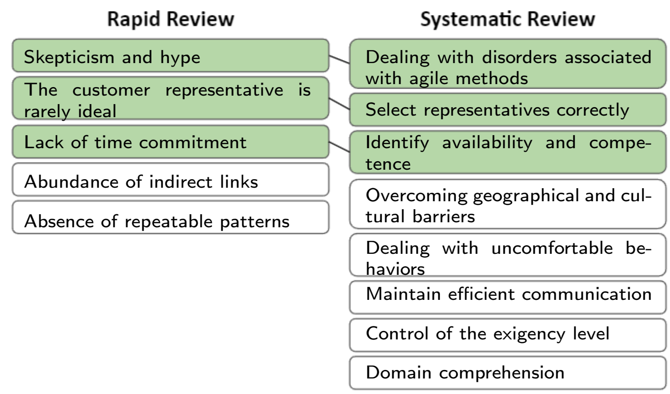

Figure 6: Comparison of challenges to establish customer collaboration (RQ3) identified on the secondary studies.

The reasons why some findings that answer RQ3 were identified in the RR but not in the SR are the divergence due to the selection procedure (1 finding) and the divergence due to the synthesis procedure (1 finding). While the reasons why some findings that answer RQ3 were identified in the SR but not in the RR are the divergence due to the selection procedure (2 findings) and the divergence due to the synthesis procedure (3 findings).

Differences between the findings of the secondary studies regarding RQ4 (Strategies to improve customer collaboration)

The RR identified 11 strategies to improve customer collaboration, while the SR found 10. Despite that, five strategies identified on the SR are not covered by the RR, while three are not covered the other way around. Figure 7 shows the strategies identified on both the RR and the RR, side-by-side.

The reasons some findings that answer RQ4 were identified in the RR but not in the SR are: divergence due to

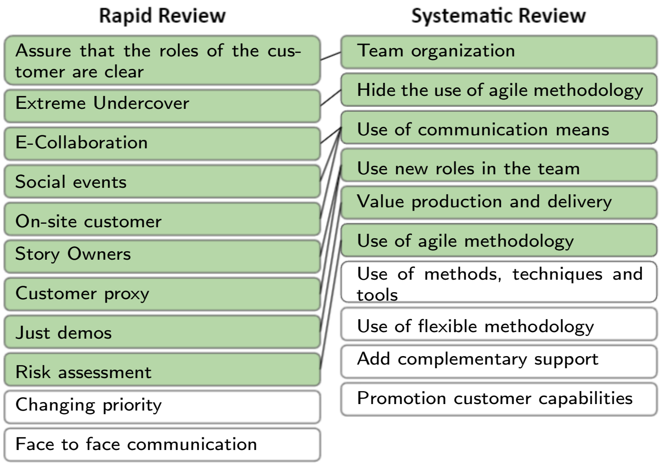

Figure 7: Comparison of strategies to improve customer collaboration (RQ4) identified on the secondary studies.

the search strategy (1 finding); and divergence due to the synthesis procedure (1 finding). While the reasons why some findings that answer RQ4 were identified in the SR but not in the RR are the divergence due to the search strategy (1 finding), the divergence due to the selection procedure (2 findings), and the divergence due to the synthesis procedure (1 finding).

Finally, Figure 12 shows a bigger picture of the main quantitative data comparing the SR with the RR.

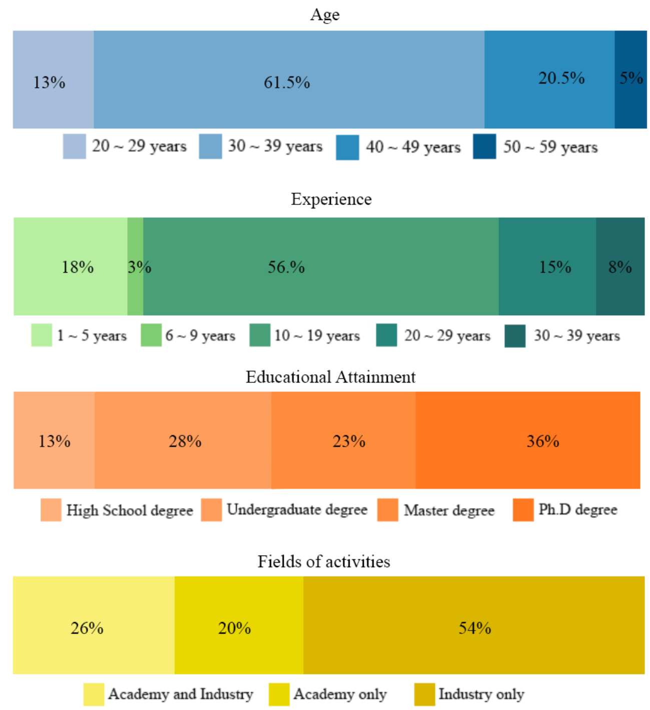

Figure 8: Participants survey demographic data.

5. Results of the Experts Survey

Figure 8 summarizes the demographics data regarding the participants of the survey.

We had 39 participants in our survey. The ages were diverse, so we transformed it into age groups. Most of the participants belong to the group 30 - 39 years (~61,5%). Followed by the group of 40 - 49 years (~20,5%), and the group of 20 - 29 years (~13%). The smallest percentage belong to the group of the 50 - 59 years (~5%).

Concerning the participants' educational attainment, ~36% of them have a Ph.D. degree, followed by an undergraduate degree (~28%), a master degree (~23%), and a high school degree (~13%). Also, most of the participants are from Industry (~54%), followed by those who work in both Academy and Industry (~26%) and Academy Only (~20%).

The participants' current role are very diverse, so we grouped them. Leadership/Management are ~44% of our sample, followed by Professors/Researchers with ~28%, and Engineers/Developers, with ~26%. About 2% preferred not to share.

Regarding the participants' professional experience in years, we also grouped them since the values are very diverse. Most of the participants (~56%) have 10 - 19 years of experience; the others are participants with 1 - 5 years of experience (~18%), 20 - 29 years of experience (~15%), 30 - 39 years of experience have (~8%) and 6 9 years (~3%) .

One may believe that older and/or more experienced specialists would be more critical and, in consequence, would assign lower relevance to the secondary studies findings. Thus, we ran Spearman's rank correlation tests. One, between the participants' age and the level of relevance they assign to the findings of both the secondary studies (rho = 0.0804); and another considering the participants' professional experience and also the level of relevance they assign to the findings of both the secondary studies (rho = 0.0684). We could not find significant correlation in any case.

That being said, our main hypothesis is that the participants would attribute a lower level of relevance to the RR's pieces of evidence when compared to the SR's findings, even though they do not know from where (SR or RR) the pieces of evidence come from.

Therefore, we performed a two-tailed Wilcoxon signed-rank hypothesis test. Again, we could not find significant difference (W = 567795, p-value = 0.07129,) between the relevance attributed to RR's and SR's findings. Despite the fact that the p-value does not enable us to reject the null hypothesis, its value is quite close to the threshold (0.05). So, we generated a violin plot with the data, in order to analyze not only with inferential but also with descriptive statistics. Figure 9 presents the violin plot that reveals data shape and dispersion.

As one can see, the relevance of RR's and SR's findings looks quite similar, both more concentrated on the positive side. However, a slight difference can be observed when plotting the data as a stacked bar chart, as shown in Figure 10.

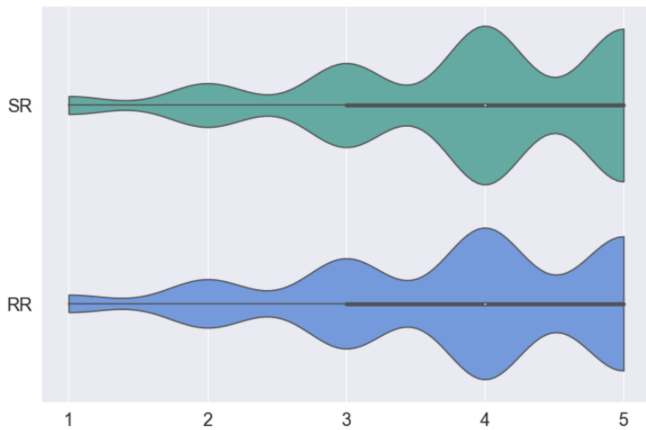

Figure 9: Shape and dispersion of the level of relevance the specialists assigned to the SR's findings compared to the RR's findings.

The SR has a few more positive values, and less neutral and negative values, although the difference is clearly negligible.

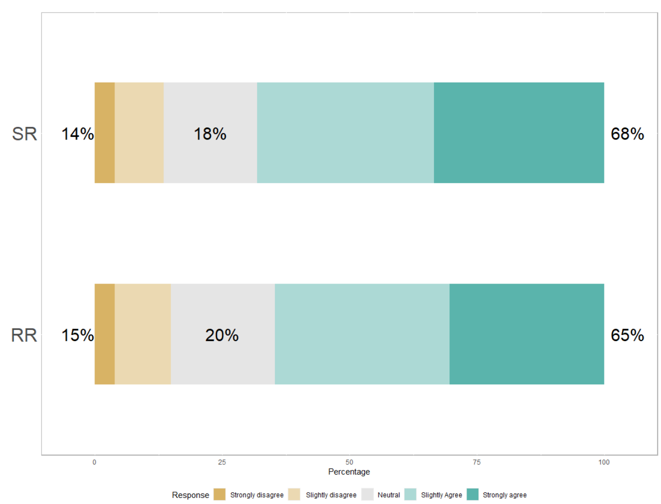

Figure 10: Level of relevance the specialists assigned to the SR's fndings compared to the RR's findings.

When we look at each piece of evidence separately, as we can see in Figure 11, one can note that the top 10 most relevant findings (slightly agree + strongly agree), according to the respondents, are a mix of findings shared by both studies (RR and SR) and the evidence found by SR only. Five pieces of evidence come from both studies, and five from the SR only.

Additionally, the top ten least relevant findings, according to the respondents, are composed of two pieces of evidence from both studies, four findings from the RR, and four from the SR. The evidence with the worst evaluation among all is shared by both the RR and the SR.

Figure 11: Level of relevance the specialists assigned for each of the secondary studies' findings.

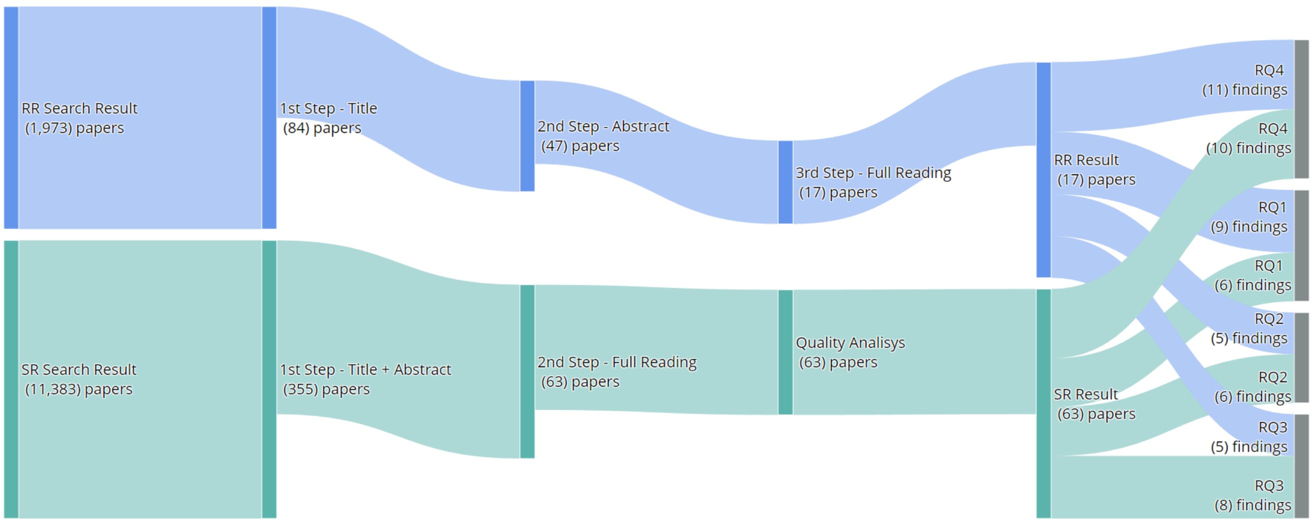

Figure 12: Procedures of both the secondary studies.

6. Discussion

Regarding the **protocol**, the SR's one is more detailed, which fosters replication and scrutiny. On the other hand, the RR protocol avoids theoretical justifications that might not be relevant to experts, saving their scarce time.

Regarding the **search strategy**, the RR seems to have a better cost-benefit when compared to the SR, at least under a quantitative perspective. Most (~78%) of the papers included in the SR (63) were returned by the RR search (49). Even though the RR search returned one order of magnitude fewer papers (1,973 vs. 11,383) than the SR. This, in turn, cuts off much of the effort spent on RR's selection procedure, while still guaranteeing a considerable similarity of relevant papers with the SR. The manual search was a great villain in this context. Just two out of 8,696 papers from the manual search ended up being included in the SR. The manual search is also responsible for more than 90% of the four months spent executing the entire SR search strategy. But a search strategy cannot be reduced to the search sources only. The search strings also can play an important role. Due to the presence of one particular term (i.e., stakeholder), the RR could include ten papers that were not even found by the SR search string. In the end, a manual search strategy seems to be not promising, even for an SR, unless the topic is critical and exhaustive coverage is mandatory.

Regarding the **selection procedure**, it has shown to be the Achilles' heel of the RR. Although the RR search could find 49 of the 63 papers included in the SR, 42 of them were unduly excluded during the RR selection procedure. Most of those unduly exclusions (38 out of 42) occurred during the first step of the RR selection procedure, where the researcher excluded papers based on papers' titles only. Nevertheless, one important aspect to take into consideration is that the SR selection procedure took six months, while the entire RR was conducted in six days. Here we are in a trade-off situation that choosing an option is a matter of pondering the research goals and the target audience of the secondary study. The number of papers excluded during the RR selection procedure may be high for a rigorous academic-focused SR. However, on the other side, six months and nine researchers to conduct just the SR selection procedure may also be unacceptable from expert' perspective. Still, regarding the selection procedure, we took a step further and investigated how the unduly paper exclusions impacted the findings of both secondary studies. And again, it shows us that the RR lost a fair number of findings compared to the SR, due to divergences in the selection procedure. In summary, the RR could not identify 11 findings that the SR could, while the SR could not identify only one finding that the RR could.

Regarding the **synthesis procedure**, this is probably the most subjective activity of a secondary study, even when one follows a systematic method like thematic analysis. Despite such subjectivity, the number of findings identified by one secondary study but not by the other, due to the synthesis procedure, is quite similar. The RR could not identify 4 out of the 30 findings of the SR due to divergences in the synthesis procedure. While the SR could not identify 5 out of the 30 findings of the RR, for the same reason. In addition, with the results from the survey we conducted with 39 specialists from academia and industry, we could see that there is no significant difference, regarding relevance of the findings, from the SR when compared to the ones from the RR. This may contradict the initial belief that evidence from SR are necessarily more relevant than from RRs. Still regarding the synthesis procedure, while the SR identified more coarse-grained findings – sometimes even comprehending more than one RR finding – the RR identified more fine-grained ones, which makes the RR's findings more specific and straightforward. This may have happen because the expert' influence guided the RR towards more specific findings, but it is just a conjecture for future work. Furthermore, it is also necessary take into consideration that the SR' findings are presented more in-depth than the RR's pieces of evidence, with descriptions and contexts, and this can also be beneficial to experts.

Regarding the **time frame** to conduct the secondary studies, the difference is severe. The SR took one year and two months, while the RR took only six days. The main bottleneck on the SR is concentrated in two activities: executing the search strategy, and the selection procedure. Together they took ten months. Those activities are also less intellectual when compared to others. So, researchers spend a lot of time and effort in an SR with activities that do not explore their full potential. Experts also cannot wait for months and sometimes years to make their decisions in real-world scenarios. Thus, considering such an expressive difference between RR's and SR's time frames, and the fact that the differences between the findings are quite balanced, we believe the software engineering research community should accept RRs as a valid secondary study method, among the others we already have. At least when the goal is delivering evidence promptly to experts.

Regarding the obvious question that comes to anyone's mind — should we conduct RRs or not? — we believe it is also obvious that the answer is not definitive. It may vary according to different contexts. In particular, it lays on another question — is the difference between the RR and the SR acceptable? — which is also hard to answer and very context-dependent. All that being said, we believe a difference in time frame from six days (RR) to one year and two months (SR) is so high, and perhaps impractical in many industry scenarios, that RRs should at least be considered in such contexts. Additionally, the positive perceptions of experts towards RRs reported by Cartaxo et al. [5] also reveal that experts are willing to cope with any threat to validity inherent to RRs, as long as they are aware of them. The findings of both secondary studies demonstrate that SRs and RRs can bring relevant contributions to professional practice. Also, they can be complementary since we could find unique pieces of evidence with both of them. The context will define which method can be applied.

Finally, one should consider all the results of this research with due care. First, we compare just one RR with one SR. Many other comparisons need to arise until we are confident about the real impact of flexibilizing SR's method. Second, both RRs and SRs have high methodological variability. So, a RR or an SR that adopts different strategies regarding any of their procedures might reveal a different impact on their results. For instance, a RR that does not apply any systematic synthesis method would probably exhibit considerably different findings compared to the RR we analyzed in this research, which adopted thematic analysis. And third, as much systematic a secondary study may be, it still inherently exhibits some level of subjectivity in most of its activities. Many aspects may interfere with the results.

7. Conclusions

This paper shows that the time frame to plan, perform, and report a Rapid Review (RR) compared to a traditional Systematic Review (SR) is severe. Six days for the former and one year and two months for the latter. We also could observe the main bottlenecks on SRs are executing the search strategy and the selection procedure. RR's search strategy also has been shown to be more cost-effective than the SRs. It returned most of the papers included in the SR but taking considerably less time and effort. On the other side, the selection procedure has shown to be the RR main drawback. Many papers that could be included were unduly excluded during the RR's selection procedure. In the end, both the RR and the SR could identify the same number of findings to answer their shared research questions, but not without some differences.

We reinforce with these results that RRs and SRs are inherently different and complementary. The time and cost to conduct an SR can be prohibitive in experts' contexts. Thus, at least in such situations, a RR may be an adequate choice. Moreover, a RR may be executed in the experts' context as a previous low-cost step before deciding to invest in a high-cost SR.

However, there is a need for more comparisons like this to draw stronger conclusions. This is the first step towards other initiatives alike. As soon as there are more RRs out there, we plan to compare them with pre-existent SRs. We also believe it is viable to have a comparison between two pre-existent RR and SR, even when they start from different research questions. As long as they share some characteristics, it is possible to have partial comparisons that together may accumulate a relevant body of knowledge around the actual differences between RRs and SRs, as there are in medicine.

7.1. Limitations and Threats to Validity

The RR was conducted by the same research group that conducted the SR. This inherently exhibits desired and not desired issues. The desired one is that researchers in our research group deeply participated in the RR, so they could provide all the details that otherwise would be hard to discover just by reading the original publication. On the other side, the fact that our research group conducted the RR may introduce bias when conducting the SR. To mitigate this bias, as mentioned in Section 3.2, the leading researcher of the SR did not read or had any contact with the original RR in any way, at least not until the end of the entire SR.

Regarding the comparison between the RR and the SR, there is least one situation that may bias the results we obtained. The RR leader had slightly more experience conducting secondary studies when compared to the leader of the SR. So, we cannot affirm if or how this fact impacts the time frame and other steps of the secondary studies.

In addition, a limitation that needs to be considered is that we only compare one RR with one SR. Thus, there is a need for more comparisons to draw stronger conclusions about the differences between RRs and SRs, but this is the first step towards other initiatives alike.

Regarding the survey, we followed the guidelines proposed to increase the validity of surveys in SE [33]. First, we defined what we wanted to evaluate – the pieces of evidence provided by the secondary studies – and then we created a draft survey. Next, we conducted a pilot with three researchers and based on their feedback, we fixed and improved the survey.

Acknowledgments

We would to thank Professor Breno Bernard Nicolau de França and Professor André Luís de Medeiros Santos for their valuable suggestions for this research. Carolline Pena is supported by FACEPE grants IBPG-0052-1.03/20. This work is partially supported by INES (www.ines.org.br), CNPq grant 465614/2014-0, FACEPE grants APQ-0399-1.03/17, and APQ/0388-1.03/14, CAPES grant 88887.13641 0/2017-00. Sergio Soares is partially supported by CNPq, grant 306000/2022-9.

CRediT authorship contribution statement

Carolline Pena: Leading, conceptualization, execution, writing, and revising. **Bruno Cartaxo:** Supervising, conceptualization (comparison and survey), execution (comparison and survey), writing, and revising. **Igor Steinmacher:** Conceptualization (survey), and revising. **Deepika Badampudi:** Execution (SR), and revising. **Deyvson da Silva:** Execution (SR). **Williby Ferreira:** Execution (SR). **Adauto Almeida:** Execution (SR). **Fernando Kamei:** Execution (SR). **Sergio Soares:** Supervising, conceptualization (comparison and survey), and revising.

References

[1] B. Kitchenham, T. Dyba, and M. Jorgensen. Evidence-based software engineering. In *Proceedings of the 26th International Conference on Software Engineering*, ICSE '04, pages 273–281, 2004.

[2] J. N. Lavis, D. Robertson, J. M. Woodside, C. B. McLeod, and J. Abelson. How can research organizations more effectively transfer research knowledge to decision makers? *Milbank quarterly*, 81(2):221–248, 2003.

[3] Edgar Hassler, Jeffrey C. Carver, Nicholas A. Kraft, and David Hale. Outcomes of a community workshop to identify and rank barriers to the systematic literature review process. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, EASE '14, pages 31:1–31:10, New York, NY, USA, 2014. ACM.

[4] R. E. S. Santos and F. Q. B. d. Silva. Motivation to perform systematic reviews and their impact on software engineering practice. In *2013 ACM / IEEE International Symposium on Empirical Software Engineering and Measurement*, pages 292–295, Oct 2013.

[5] B. Cartaxo, G. Pinto, and S. Soares. The role of rapid reviews supporting decision-making in software engineering practice. In *22nd International Conference on Evaluation and Assessment in Software Engineering (EASE)*, 2018.

[6] B. Cartaxo, P. Pinto, and S. Soares. Towards a model to transfer knowledge from software engineering research to practice. *Information and Software Technology*, 2018.

[7] Deepika Badampudi, Claes Wohlin, and Tony Gorschek. An evaluation of knowledge translation in software engineering. In *2019 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, pages 1–6. IEEE, 2019.

[8] D. Chambers and P. Wilson. A framework for production of systematic review based briefings to support evidence-informed decision-making. *Systematic reviews*, 2012.

[9] S. Khangura, K. Konnyu, R. Cushman, Jeremy Grimshaw, and David Moher. Evidence summaries: the evolution of a rapid review approach. *Systematic Reviews*, 2012.

[10] B. Cartaxo, G. Pinto, and S. Soares. *Contemporary Empirical Methods in Software Engineering*, chapter Rapid Reviews in Software Engineering. Springer, first edition, 2020.

[11] Andrea C. Tricco et al. A scoping review of rapid review methods. *BMC Medicine*, 2015.

[12] A. C. Tricco et al. Rapid reviews to strengthen health policy and systems: a practical guide. *World Health Organization. http://apps. who. int/iris/bitstream/10665/258698/1/9789241512763-eng. pdf*, 2017.

[13] D. Amalfitano et al. A rapid review on testing of context-aware contemporary software systems. 2019.

[14] B.P.ça de Souza, R. C. Motta, and G. H. Travassos. Towards the description and representation of smartness in iot scenarios specification. In *Proceedings of the XXXIII Brazilian Symposium on Software Engineering*, pages 511–516, 2019.

[15] Francisco Ponce, Gastón Márquez, and Hernán Astudillo. Migrating from monolithic architecture to microservices: A rapid review. In *2019 38th International Conference of the Chilean Computer Science Society (SCCC)*, pages 1–7. IEEE, 2019.

[16] Sebastián Pizard, Diego Vallespir, and Barbara 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)*, pages 1–13, 2022.

[17] Sebastián Pizard, Fernando Acerenza, Diego Vallespir, and Barbara Kitchenham. Assessing attitudes towards evidence-based software engineering in a government agency. *Information and Software Technology*, page 107101, 2022.

[18] Sergio Rico, Nauman Bin Ali, Emelie Engström, and Martin Höst. Guidelines for conducting interactive rapid reviews in software engineering–from a focus on technology transfer to knowledge exchange. 2020.

[19] B. Cartaxo, G. Pinto, B. Fonseca, M. Ribeiro, P. Pinheiro, and S. Baldassarre, M. T.and Soares. Software engineering research community viewpoints on rapid reviews. In *2019 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, pages 1–12. IEEE, 2019.

[20] Ahmed M Abou-Setta et al. Methods for developing evidence reviews in short periods of time: A scoping review. *PloS one*, 2016.

[21] Sian Taylor-Phillips et al. Comparison of a full systematic review versus rapid review approaches to assess a newborn screening test for tyrosinemia type 1. *Research Synthesis Methods*, 2017.

[22] Stijn Van de Velde et al. Medicinal use of potato-derived products: conclusions of a rapid versus full systematic review. *Phytotherapy Research*, 2011.

[23] Maddern G., Norderhaug I., Hailey D., et al. Rapid reviews versus full systematic reviews: an inventory of current methods and practice in health technology assessment. *Int J Technol Assess Health Care.*, 2008.

[24] B. Kitchenham and S Charters. Guidelines for performing systematic literature reviews in software engineering, 2007.

[25] S. Merrian and E. Tisdell. *Qualitative Research: A guide to Design and Implementation*. The Jossey-Bass higher and adult education series, fourth edition, 2016.

[26] Deepika Badampudi, Farnaz Fotrousi, Bruno Cartaxo, and Muhammad Usman. Reporting consent, anonymity and confidentiality procedures adopted in empirical studies using human participants. *e-Informatica Software Engineering Journal*, 16(1):220109, July 2022. Available online: 22 Jul. 2022.

[27] Anthony Viera Joanne Garrett. Understanding interobserver agreement:the kappa statistic. *Journal of the Society of Teachers of Family Medicine*, 2005.

[28] Tore Dybå and Torgeir Dingsøyr. Empirical studies of agile software development: A systematic review. *Information and Software Technology*, 50(9):833 – 859, 2008.

[29] Jennifer Fereday and Eimear Muir-Cochrane. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods*, 2006.

[30] D. S. Cruzes and T. Dyba. Recommended steps for thematic synthesis in software engineering. In *2011 International Symposium on Empirical Software Engineering and Measurement*, pages 275–284, 2011.

[31] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[32] Bruno Cartaxo, Gustavo Pinto, Elton Vieira, and Sérgio Soares. Evidence briefings: Towards a medium to transfer knowledge from systematic reviews to practitioners. In *Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, ESEM '16, pages 57:1–57:10, New York, NY, USA, 2016. ACM.

[33] J. Linåker, S. M. Sulaman, R. Maiani de Mello, and M. Höst. 2015.

Author biography with author photo. Author biography.

Author biography with author photo. Author biography. Author biography. Author biography. Author biography. Author biography. Author biography. Author biography. Author biography. Author biography. Author biography. Author biography. Author biography.