Systematic Literature Reviews in Software Engineering: Preliminary Results from Interviews with Researchers

Muhammad Ali Babar
Lero, Software Engineering Research Centre
University of Limerick, Ireland
muhammad.alibabar@lero.ie

He Zhang
Lero, Software Engineering Research Centre
University of Limerick, Ireland
he.zhang@lero.ie

Abstract

Systematic Literature Reviews (SLRs) have been gaining significant attention from software engineering researchers since 2004. Several researchers have reported their experiences of and lessons learned from applying systematic reviews to different subject matters in software engineering. However, there has been no attempt at independently exploring experiences and perceptions of the practitioners of systematic reviews in order to gain an in-depth understanding of various aspects of systemic reviews as a new research methodology in software engineering. We assert that there is a need of evidence-based body of knowledge about the application of systematic reviews in software engineering.

To address this need, we have started an empirical research program that aims to contribute to the growing body of knowledge about systematic reviews in software engineering. This paper reports the design, logistics, and results of the first phase empirical study carried out in this program. The results provide interesting insights into different aspects of systematic reviews based on the analysis of the data gathered from 17 interviewees with varying levels of knowledge of and experiences in systematic reviews. The findings from this study are expected to contribute to the existing knowledge about using systematic reviews and help further improve the state-of-the-practice of this research methodology in software engineering.

1 Introduction

A Systematic Literature Review (SLR), also referred as systematic review, is considered one of the key research methodologies of Evidence-Based Software Engineering (EBSE). Systematic reviews have been gaining significant attention from software engineering researchers since Kitchenham, Dyba and Jorgensen's seminal paper on EBSE published in ICSE 2004 [16]. Software Engineering (SE) researchers have been conducting and reporting more and more SLRs on diverse topics such as agile software development [6], regression testing [11], process modeling [26], variability management [5], cost estimation [17], organizational motivators for CMM-based process improvement [25], and statistical power [9]. Researchers have also reported best practise and experiences of conducting and reporting systematic reviews [4, 8, 24, 2]. In addition, the techniques for designing the strategies for assessing the quality of the reported primary studies included in a systematic review have been proposed [7]. Moreover, there has been at least one tertiary study, systematic literature review of systematic literature reviews, reported in [14]. We assert that as the interest of software engineering researcher in systematic reviews is increasing, so should be the need for providing appropriate methodological guidance in designing, conducting, and reporting high quality systematic reviews.

It is a common observation that most of the reported SLRs have been carried out by experienced researchers. Moreover, the lessons learned from and experience reports of performing systematic reviews in software engineering have also been published by either the advocates of EBSE or experienced researchers. A large majority of the reported systematic reviews in software engineering has been carried out by following the guidelines produced by Kitchenham and Charters [13, 15]. Some researchers have also reported their own guidelines for SLRs [3] or have consulted the guidelines from medicine or social sciences [21]. While the available guidelines and lessons learned reports are important and valuable, as they provide the software engineering practitioners and researchers with useful information about different aspects of systematic reviews, there is a vital need for allocating more resources to provide comprehensive guidance and training in performing systematic reviews.

This situation highlights the vital need of allocating more resources for independent studies of the applications of systematic reviews in software engineering in order to build a

978-1-4244-4841-8/09/$25.00 ©2009 IEEE

body of knowledge to improve the current methodological support. Such studies should also aim to synthesize the experiences of and lessons learned by researchers with varying levels of competency of empirical research in general and systematic reviews in particular. To fulfill this need, we are currently carrying out an empirical research program, which aims to gather and disseminate empirical findings about applying systematic reviews. Unlike the previously reported studies of experiences and lessons learned, this research intends to holistically reflect on the real adoption of systematic reviews in software engineering and the challenges that researchers commonly face when applying this research methodology. Specifically, the main goals of this research are to:

1. Collect and reflect on the perceptions and experiences of researchers with varying levels of knowledge and experience of performing SLRs in SE (i.e. advocates, followers, or novices);
2. Report the challenges, strategies, best practices and lessons learned from conducting SLRs in a variety of context irrespective of types and results of SLRs;
3. Empirically evaluate the use and value of the available guidelines for performing SLRs, and provide suggestions;
4. Identify the potential areas for enhancement and improvement of SLR as an important research methodology for EBSE.

In the initial phase of this program, we have conducted a series of interviews with 17 researchers who have performed systematic reviews on various topics in SE. An initial analysis of the data gathered through these interviews has revealed quite interesting information that is expected to be useful for the researchers and practitioners interested in SLRs. The paper makes following contributions to the EBSE, and broadly to empirical software engineering:

- It describes the design, logistics, and preliminary results of an empirical research program aimed at improving the state of the research and practise of SLRs in SE.
- It identifies the areas that need to be improved in the current guidelines available for designing, conducting, and reporting SLRs in SE.
- It reports the perceptions of researchers (reviewers) about the value of SLRs in SE as compared to other disciplines such as medicine.
- It discusses some of the best practices of performing SLRs, identified by the participants of this study.
- It determines the challenges researchers usually face while performing SLRs and describes the motivations of SLRs for them.

Given the objectives of this paper we structure this paper as follows. Section 2 provides a brief discussion on the context of the research. Section 3 elaborates on the details of the research methodology and logistics. Section 4 presents and discusses the preliminary results from the data analysis performed. Section 5 presents the discussion and limitations of the research. Section 6 closes the papers with outlining the future work.

2 Research Context

This section briefly summarizes the main objectives of performing systematic reviews, and opinions and recommendations to motivate the need for the reported research. A systematic review is "*a means of evaluating and interpreting all available research relevant to a particular research question, topic area or phenomenon of interest*" [10, 13]. Systematic review as a research methodology is widely used in medical research since 1990s, and within that field there are a number of well-documented standards to support its use. The primary reasons for performing systematic reviews are: to summarize the existing evidence concerning a treatment or technology, to identify any gaps in current research in order to suggest areas for further investigation and to provide a framework/background in order to appropriately position new research activities. One of the main goals of systemic reviews is to ensure that the review is methodological, repeatable, and thorough. Systematic reviews also attempt to minimize the level of bias that can be prevalent in ad-hoc literature surveys.

It has been mentioned that the number of software engineering researchers performing systematic reviews has continuously been increasing since 2004. Premier journals and Magazines have special sections for papers based on systematic reviews and a significant number of software engineering conferences seek submissions in this category. To support the adoption and correct application of this research methodology, researchers have also been working for providing methodological support and reporting their experiences and lessons learned as is evident from the guidelines for performing systematic reviews in SE reported in [15, 3] and lessons learned reported in [4, 24, 8].

Nevertheless, several researchers have identified the areas for improving the published guidelines and the needs for supportive techniques. For example, it has been reported that the SLR guidelines provide sufficient support for the reviews that need to synthesize the quantitative data, however, there is not much advice on synthesizing and interpreting qualitative data [6, 8]. A lack of support for dealing with the qualitative data becomes a real problem when

978-1-4244-4841-8/09/$25.00 ©2009 IEEE

dealing with a diverse set of software engineering studies. Researchers have also emphasized the need for providing appropriate advice on how to assess the quality of primary studies as the quality of the results of a systematic review is largely dependent upon the quality of the primary studies reviewed [7, 24]. That appears to be one reason that relatively few researchers are assessing the quality of the studies included in their reviews [6].

It has been widely reported that the current SE specific bibliographic databases are not designed to facilitate systematic reviews [4, 8, 24]. Researchers have been working around these problems by constructing and using different search strings for different digital databases. It requires significant amount of effort to ensure that all the search strings are semantically equivalent if not syntactically. Most of the time researchers find themselves without any precise guidance or advice in the current guidelines or experience reports on how to carry out the search process to deal with the challenges caused by the usages of non-standard terms and diverse nature of the search facilities provided by digital libraries. Deciding about the time frame of a search can also be a challenge for novice researchers.

This discussion on the reported opinions and experiences of various researchers clearly shows that like any other empirical research method, there is an increasing need for providing software engineering researchers and practitioners with appropriate knowledge and training in different aspects of systematic reviews [23]. That means software engineering researchers need to allocate significantly more resources to develop suitable support system for guiding researchers on how to design, conduct, report, and retrieve high quality systematic reviews in SE and practitioners on how to assess the quality and results of systematic reviews being published on a topic that may interest them.

3 Research Methodology

In this section, we describe and discuss the research methodology, data gathering technique and instrument, and data analysis method used for the reported research.

3.1 Interview-Based Survey

We decided to use survey research method to explore the perceptions and opinions of research practitioners about their experiences of conducting SLRs on diverse topics in SE. A survey research method is considered suitable for gathering self-reported quantitative and qualitative data from a large number of respondents [18]. Our survey design was a cross-sectional, case control study. Survey research can use one or a combination of several data gathering techniques such as interviews, self-administered questionnaires and others [19]. We decided to use interviews as the data collection instrument in the initial phase of this research as it appeared to be more appropriate for gathering detail-riched information required to find answers to the questions that motivated our research, as well as to find more interesting threads that may be included in the next phase research.

3.1.1 Data Collection Technique and Instrument. We used semi-structured open-ended interviews to collect qualitative data. One reason for choosing interview as the data gathering technique was to gather as much information as possible from the interviewees as our target population, especially senior researchers, were expected to be interviewed only once for this research. The open-ended nature of the questions within the structured nature of the interviews was expected to help us to systematically collect useful data for the study. It is considered that open ended questions allow for a variety of responses and fit better with the aim of getting an '*insider view*' of a situation [24]. Open ended questions are also expected to help researchers to avoid introducing any of his or her own preconceptions and protect the validity of the data.

We designed our interviewing instrument with the intention of keeping the discussion focused and using the interviewee's time effectively [20]. Our interviewing instrument consisted of a set of open-ended questions carefully worded and arranged into six different sections. The structure of the interviewing instrument was designed with the intention of taking each respondent through the same sequence and asking each respondent the same questions with essentially the same words. The benefit of using the open-ended questions in an interview is that a researcher can obtain data that are systematic and thorough [20]. Whereas, such interviews may reduce the flexibility and spontaneity because the probing during the interview is kept limited. Nevertheless, we planned to address this limitation by using frequent probes during the interviews. Hence, we planned to focus not only on the '*What*' questions but also '*How*' and '*Why*' probes in response to the answers to the designed questions. Elaboration probes are used to keep an interviewee talking more about a subject [20].

We ran a pilot of our interviewing instrument and it was estimated that an interview would take between 70 and 90 minutes.

3.2 Data Source

Since the main goal of our research program is to gain an in-depth understanding of different aspects of applying SLR in SE by exploring the experiences, opinions, and perceptions of the users of SLR in SE, our target population for data gathering consisted of practitioners (i.e. researchers) of SLRs in SE. Based on a literature search and our own awareness, we identified 24 researchers, who appeared to

978-1-4244-4841-8/09/$25.00 ©2009 IEEE

be active practitioners of SLRs in SE based on sampling their publications on SLRs in SE. We classified the identified researchers into three categories:

Advocates Researchers who introduced SLR methodology and EBSE in SE, and have published many SLRs they conducted in the past years;

Followers Researchers who have participated in planning, conducting and reporting one or more SLRs;

Novices PhD research students who have experiences in performing SLRs.

An email invitation to participate in our research was sent to the identified researchers from both researchers. Our email briefly described the research project, its goals, and the nature of the commitment required by the potential participants.

We received replies from 21 invitees. Seventeen of them agreed to be interviewed during the time period that we had allocated for the data collection stage of this study. Two responded that they were interested but could not be available in the initial phase of this research. Other two invitees showed interests in the research, however, declined for some personal reasons. Three of the invitees never responded despite our two reminders. Hence, the findings reported in this paper are based on the analysis of the data gathered from seventeen interviewees, who represented seven countries (i.e., Australia, Brazil, China, United Kingdoms, Ireland, New Zealand, and Norway). Table 1 shows the number of interviewees for each of the categories and the forms of interviews.

Table 1. Types of interview by interviewees

	Advocates	Followers	Novices
Face to face	0	1	6
Telephone	3	6	0
Skype	0	1	0

3.3 Study Procedure

The interviews were conducted between December 2008 and February 2009. The interviewees are from ten research organizations (research institutes or universities). Thirteen interviews were conducted in English, and four were conducted in Mandarin.

A few days (2 or 3 days) before the scheduled interview, each participant was provided, via email, with a document outlining the main themes to be covered during the interview. We encouraged each of the participants to use that document to do some preparation and reflection before the interview. We also asked them to gather some facts and figures about their respective SLRs in order to facilitate the discussion during the interview session. In order to assure the participants of confidentiality and privacy of the data gathered during the interviews, we also sent a statement stating how the data will be protected. We repeated that statement to each interviewee before starting the interview. Participants had been informed that we intended to record the interviews with their permission. However, they could have chosen not to be recorded. In the latter case, they had been forewarned that the interview duration might be longer as the researchers would have been expected to take more detailed notes during the interviews. The participants were also assured that their data would not be accessible to anyone except the researchers. Moreover, we explicitly made it clear to the participants that the research team would not share the data with anyone in a way that could reveal the opinions and views of individuals. We sought permission from each participant for recording the interview at the beginning of each interview. There were no objections by any of the interviewees to recording the interviews.

We used mainly two forms of interviews: face-to-face and teleconference. For the teleconference, all but one interview were conducted using telephones. One interview was conducted via Skype on the desire of the interviewee. We did not observe any difference between using telephone and Skype for the interviews. Most interviews were recorded with two digital recorders, one for each researcher. The researchers also took extensive notes of the discussions. Each interview lasted approximately 90 minutes on average. The same questions and format were used for all the interviews. However, some of the interviews were conducted by one researcher alone. All of these interviews were conducted with the interviewees classified in the category of novices in SLRs.

3.4 Data Analysis Process and Methods

The data analysis step involves transcribing the recorded discussion. The transcribed data can be analyzed using one or more of the qualitative data analysis techniques reported in [22]. All four interviews by Mandarin were translated into English for analysis. The interviews with 17 participants resulted in approximately 28 hours of audio recording and extensive notes. This is a large amount of qualitative data and each recorded hour usually takes 6 to 8 hours of transcription work. We have decided to follow a staged process of transcribing the recorded data. For the first stage, we have identified a set of questions, mentioned earlier, whose answers have been transcribed for analysis. Rest of the recording will be transcribed at a later stage.

After transcribing the responses to the selected set of

978-1-4244-4841-8/09/$25.00 ©2009 IEEE

questions, content analysis was used to analyze the data following the procedures described in [19]. All transcriptions were entered into NVivo [1] for qualitative analysis.

4 Preliminary Results

Each interviewee in this study has participated at least one SLR in software engineering. Some example questions for the results reported in this paper are listed in Appendix.

4.1 Working with Guidelines

4.1.1 Methodology acquisition. The practitioners of SLR in SE follow some reference as 'guidelines' when performing their SLRs. The most influential SLR guidelines in SE have been developed by Kitchenham and Charters [13, 15], which are recognized by almost all interviewees. In addition, some research organizations have adapted their own guidelines for SLR, such as the technical report by Biolchini et al.'s [3].

In addition to Kitchenham's guidelines, all *advocates* also learned and followed the guidelines and materials from other disciplines, especially medicine and sociology. Another useful reference from social science (Petticrew and Roberts' [21]) was recommended to SE community by three of our interviewees independently. *Advocates* suggested reading any general guidelines as a starting point, and then using "*more detailed guidelines for specific areas (where) your interest is and relevant*", however, "*the problems you have may be really specific or difficult, you couldn't get support in guidelines, because they are not covered*".

Apart from the above guidelines, many *followers* and all *novices* reported that they also learn SLR through reading the papers or technical reports reporting SLRs and relevant experience reports in SE. All *followers* except one have direct connections to at least one advocate. In contrast, all *novices* were not able to access the guidance from any *advocate* directly.

4.1.2 Improvement suggestions. Table 2 summaries the improvement for the SLR guidelines suggested by the practitioners. The detailed instructions about '*how to assess study quality in different types of SLRs*' is the most needed information in the currently available guidelines. More examples of good SLRs and useful experiences are also expected by many interviewees.

4.2 Value of Systematic Reviews

4.2.1 Value to software engineering. All interviewees provided positive feedback on the value of SLR in SE.

Advocates argue that software engineering practitioners need "*not only a body of knowledge, but a body of evidence about how about the methods ... (and) SLR should be based on a body of evidence*"; it can help judge what we actually know, based on empirical studies; it is also necessary to make some summaries, which "*should be a good way to identify where more research is needed, where less is needed*". In EBSE, mapping study should (such as [12]) be "*a starting point to identify sources of material in particular topic area*".

Compared to the value to industry, most interviewees recognize SLR is more valuable to academia. *Followers* reflected SLR "*does provide quite sound basis for arguing and presenting cases and date in SE*", it is "*a rigorous way to help us identify whether the gap in the literature*". It also "*provides a new way to get data ... if we can't get access the (real) data itself*".

Compared to traditional literature review, "*it's wonderful to make literature reviews properly done, to make it really scientific, precise, and other people can repeat*"; "*it's a very objective method, and defines a formal process for literature reviews, then results can be comparable and reliable*". However, "*it's only valuable if (the authors) list papers (included), ... even it has retrieved only a small number of papers*".

Some *novices* consider SLR could bring similar value to industry, particularly for technology adoption, but the value also depends on the application domain and research questions.

4.2.2 Value to novices. Almost all interviewees believe SLR can benefit entry level researchers. Regarding the appropriateness for novices, particularly first-year PhD students, however, they gave different answers. Table 3 shows the answers given by the participants from different categories.

Table 3. Is SLR appropriate for novice?

Appropriate?	Advocates	Followers	Novices
Yes	1	3	4
No	0	1	0
It depends	2	4	2

Only one *follower* did not think it's appropriate for novices, because "*PhD students should do a lot of literature reviews, but doing SLR is quite different thing, it's very focus and (needs) very well-defined research questions*".

The supporters *believe first-year PhD students have to do SLRs*. The reasons include "*[students] can get hold of what have been done in the area they pursue PhD and find really the gaps*"; "*[SLR] provides an essential research skill to novices*", "*and helps to organize the work, helps to make things clear, [supervisor] cannot control students*

978-1-4244-4841-8/09/$25.00 ©2009 IEEE

Table 2. Improvement suggestions for SLR guidelines

Topic	Description	Adv.	Fol.	Nov.
Quality assessment	*Needs to know how to assess the quality of evidence; the current quality criteria could not be applied to all SLRs.*	0	2	3
Experience/examples	*Some experiences should be grouped for different topics; real examples of good protocols could be helpful.*	1	2	1
Simplified version	*Some kind of pocket version guide for people who are reviewing SLR papers; a simplified version is needed for novices.*	0	2	2
Quantitative analysis	*More references of statistic methods should be included in guidelines; more details about how to do meta-analysis are expected.*	0	1	1
Qualitative analysis	*Most of guidelines are relevant to quantitative studies and analysis, however in SE, we also have to deal with qualitative studies, like case studies.*	1	0	0
Protocol template	*Need to improve review protocol templates, to describe how to fill the protocol, which depends on the type of SLR.*	0	1	0

quite wide and quite well"; it also helps avoid missing important papers and ignore low-quality ones. In practise, they suggested it can be started with mapping study.

Most *Novices* welcome this new methodology. Compared to an ad-hoc literature review, "*its process has been clearly defined in the guidelines and specified in protocol*", which makes it easy to be followed by students. They also said "*it's a really strong tool, to help them find out the research opportunities*". Nevertheless, they perceive lower productivity and more difficulties experienced in their SLRs, compared to skilled researchers.

Table 4 summaries the frequency of factors considered by the skeptics that may influence the appropriateness.

Table 4. Factors influencing the appropriateness to novices

Factors	Frequency	%
Experience needed	3	43%
Too much time & effort	3	43%
Work with experts	2	29%
Get focused	2	29%
Domain knowledge needed	1	14%

4.2.3 Compared with other disciplines. SLRs have been widely adopted in medicine discipline, where EBSE borrowed the idea from. Nonetheless, the practitioners of SLR in SE have quite different perspective of the effectiveness compared to other disciplines, especially medicine. As shown in Table 5, almost half of interviewees think '*it could be*', but the other nearly half '*do not think so*'.

Table 5. SLR's effectiveness in SE vs. medicine

As effective as in medicine?	Adv.	Fol.	Nov.
It could be	1	5	2
I don't think so	2	1	4
Hard to say/I don't know	0	1	0

Advocates stated "*we should recognize that SLRs in other disciplines should be and have to be different from medicine*". In SE, "*there are lots of side-effects should be reported, procedure, and so on*", so "*it will be as useful as they are in sociology, politics and economics*", except medicine.

Though most *followers* are optimists, they also recognized the difference of SLRs between SE and other disciplines. "*In medicine domain, they're dealing with experiments, but we don't run the experiments the same way in SE necessarily ... so questions need to be chosen very carefully in SE to make sure SLR is actually suitable tool to use.*" Apart from difference in experiments, some issues in current SE research have emerged as well, such as the literature databases (libraries) and abstract quality in papers. But one *follower* encouraged "*we started in 2004, now we have a lot of SLRs published ... some days before we have almost none evidence, but now we have some, I think it's much better than zero*".

The pessimists directly questioned the quality and quantity of primary empirical studies (evidence) in SE, and accordingly argued the reliability of SLR results is lower than in medicine. Even many PhD students thought SLR is a good research methodology, they did not think it would be

978-1-4244-4841-8/09/$25.00 ©2009 IEEE

as effective as in medicine, because "*the context or environment in SE is not mature and controllable enough (like in medicine)*".

4.3 Experience and Best Practice

4.3.1 Problems related to SLRs. Roughly speaking, there are three major problems reflected in the interviews: database, publication, and education.

Databases The current literature databases or digital libraries provide web search interfaces, which several interviewees complained about. "*It is the most frustrating thing with us, because it needs a lot of work, a lot of noise risk you.*" "*When I search ACM, we got one result, when I search two days after, I got different results. Sometime the paper includes, sometime the (same) paper not included. So what to do?*"

Publication Some people who are reviewing papers that claim as an SLR do not know well what is SLR, which results in some low quality SLRs being published in conferences and journals. Some of them do not have explicit research questions, well-defined search strategy and selection process, and so on.

Education For example, "*some PhD student worked in his SLR, but his supervisor who have never done one before, you couldn't believe the kinds of questions come up. It's like try to help someone write Java application, where you haven't done really coding with Java?*"

4.3.2 Best practices. Table 6 summaries the best practices for doing SLRs that are supported by more than one interviewee. They are from SLR practitioners and for current and future practitioners of SLRs.

In addition, two *followers* recommended the ideal team size for SLR is to get three people involved. If less, it might be difficult to avoid subjective bias; if more, it may take much time and effort in communication, coordination and getting agreement, particularly in a distributed working environment.

4.3.3 Productivity improvements. Most interviewees gave positive response to the question associating productivity improvements compared to their previous SLRs. The improvements mainly attribute to learning from experience and getting more confidence. "*The effort at beginning is on understanding how to write protocol, what SLR really needs ... after that, you concentrate on the analysis of your data.*" Compared to the first time SLR, "*more time and effort (are required) for learning and exercising*", "*it could be half time now, because we had already done that, we know how to do things.*"

Though some interviewees said "*I don't know*" or "*possible*" to productivity improvement, they affirm the rigor of their research has improved. But they remind "*if the (subject) domain changes, then perhaps no significant improvement*".

4.4 Challenges and Fulfillment

4.4.1 Challenges. Table 7 summaries the issues challenging or frustrating SLR practitioners. The most challenging thing (to the interviewees) is time and effort taken in SLRs. The capability of search engines (of digital libraries) and design of search strings are another major challenge.

Table 7. Most challenging things in SLRs

Challenge	Adv.	Fol.	Nov.
Time/effort consuming	2	3	1
Searching literature	1	3	1
Guiding students	1	2	0
Defining research questions	0	2	1
Too much rework	0	1	0
Study selection	0	1	0
Getting agreement	0	1	0
Lack of guidance	0	0	1
Lack of domain knowledge	0	0	1
Writing protocol	0	0	1
Rejection of paper	0	0	1

4.4.2 Fulfillment. Table 8 enumerates the encouragements and fulfillment to the researchers when doing SLRs. The most important motivators for conducting an SLR are '*getting new findings from the results of SLR*', and '*learning from studies and getting knowledge*', which both are related to reviewer's research interests.

5 Limitations

As the first reflection of user's experiences and perspectives of systematic review, our study also has some limitations. Our study has explored the perceptions and views of software engineering researches about their experiences of applying SLRs in SE through semi-structured interviews. The interviewees might have reported whatever they could remember. That means our results are based on the recollection of the interviewees. This is a well known weakness of retrospective interviews. However, we have full confidence in our findings because we collected data from re-

978-1-4244-4841-8/09/$25.00 ©2009 IEEE

Table 6. Best practices from SLR practitioners

	Best practise	Adv.	Fol.	Nov.
1.	*Make your research questions as concrete and explicit as possible, keep focus on them and a narrow world. Don't waste lots of time on irrelevant literature*	1	3	2
2.	*Read the guidelines, e.g. Kitchenham's guidelines, make sure your understand, and then follow the guidelines; but never expect the guidelines give all answers to the problems*	1	2	1
3.	*Find and read good SLR examples, experiences, and protocols from others as many as you can.*	0	3	1
4.	*Expect protocol take a long time, allocate appropriate time for it, and expect changes. Get your protocol validated externally, as a low-quality protocol may lead to a lot of rework. Share the protocols within community.*	0	2	2
5.	*Do pilot review, it is necessary especially when you're not familiar with the domain for SLR*	0	1	3
6.	*Have somebody with experience in conducting SLRs involved or being in touch, make them available to consult to, and ask them check your questions and results.*	0	3	0
7.	*Do bookkeeping, record as much as you can during the review.*	0	2	1
8.	*You should have good reasons for everything you do; you should be willing to do it. Don't stop thinking, and be very careful about what you're doing*	1	0	2
9.	*Clarify criteria for search, selection and quality as much as you can, and as good as you can.*	0	2	1

Table 8. Encouragements and fulfillment in SLRs

Challenge	Adv.	Fol.	Nov.
New findings from SLR	0	5	0
Learning from studies/getting knowledge	1	2	1
Recognition from community	0	3	0
Paper publication	0	1	2
Working experience	0	1	0
Learning research skills	0	1	0

searchers who have been trained and involved in applying SLR methodology to a diverse set of topics in SE.

This validity of the findings of a qualitative study based on interviews can also be threatened by the inaccuracies or incompleteness introduced into what was heard. We tried to minimize this risk by audio-taping all the interviews using two separate recorders. The transcriptions of the interviews were verified with the notes taken. Moreover, we tried to have both researchers present in most of the interviews.

We also paid careful attention to the validity of the interpretation performed based on the transcribed data and notes. One way was to minimize the amount of interpretations and speculations at this stage. We intend to interpret the findings in light of the data that we plan to acquire, e.g. papers of the interviewees, in order to apply triangulation for gaining rich understandings of the findings. Generalizability can be another risk. However, we tried to manage this by selecting the interviewees from different organizations and located in different parts of the World. So far the external generalizability is concerned, qualitative studies are usually considered weak in this respect and we do not claim the generalizability of the findings from this research. Nevertheless, it should be noted that a large majority of the interviewees reported similar experiences and lessons, which increases our confidence in the findings of this study.

6 Summary and Future Work

The objective of our research on EBSE is to contribute to the growing body of evidence on how to perform systematic reviews in software engineering. We plan to follow two-pronged strategy to achieve this objective: developing techniques and tools to complement the existing SLR guidelines, as well as gathering and reporting the opinions and experiences of the practitioners of systematic reviews in software engineering. Especially we want to increase the understanding and knowledge about performing high quality systematic reviews. We are also interested in finding appropriate means to reduce the time and resources required for effectively and efficiently carrying out systematic reviews without compromising the quality. This paper reports the preliminary findings from the initial phase of this research program that is one of the most important steps towards these goals.

The results presented in this paper are useful for researchers interested in gaining knowledge about different aspects of performing SLRs in SE. For example, the results describe different sources that the interviewees used to learn about the process of performing systematic reviews. The re-

978-1-4244-4841-8/09/$25.00 ©2009 IEEE

searchers classified in different categories for this research described different sources of knowledge about SLRs. It has also identified the areas of improvements to the current guidelines. Researchers interested in improving the methodological aspect of systematic reviews can direct their research effort to fill the identified gaps. The findings also highlight the value of performing secondary studies in a systematic and rigorous manner as even novice researchers and PhD students are of the view that systematic reviews are valuable to software engineering research and practise. These findings should provide researchers further evidence to make a business case for seeking resources and funds for performing systematic reviews and improving methodological support. Moreover, the findings about the challenges caused by low quality primary studies should provide yet another motivation to researchers to continuously work on improving the methodological and reporting rigor of reported studies. This study has also identified a few best practices that are expected to be useful for researchers intending to undertake systematic reviews. These practices are based on the accumulated experiences of the interviewees of this research. Additionally, researchers can also benefit from the knowledge of the factors that encourage the interviewees to undertake the arduous task of conducting systematic reviews. It is also hoped that the results of this study will stimulate researchers to carry out the kind of studies reported in this paper as well as contribute to the methodological aspects of systematic reviews.

The future work in this line of research includes the analysis of the data on the other questions that we asked from the interviewees. We also intend to perform triangulation by extracting the data from the papers published by the interviewees on systematic reviews. We also intend to identify the similarities and differences among the factors reported by researchers with varying levels of knowledge and experiences of systematic reviews. Moreover, we have also identified another set of respondents that would be invited to participate in the second phase of data collection.

7 Acknowledgments

The authors give our great thanks to all interviewees who accepted our invitations for their time and inputs. Without your dedication this study would never have been completed. We are in the process of findings out whether or not the participants would like to be acknowledged by names.

This work was supported, in part, by Science Foundation Ireland grant 03/CE2/I303 1 to Lero - the Irish Software Engineering Research Centre (www.lero.ie).

References

[1] Nvivo version 7. http://www.qsrinternational.com/, January 2009. Provalia Research.

[2] M. T. Baldassarre, D. Caivano, B. Kitchenham, and G. Visaggio. Systematic review of statistical process control: An experience report. In *11th International Conference on Evaluation and Assessment in Software Engineering (EASE'07)*, UK, April 2007. British Computer Society.

[3] J. Biolchini, P. G. Mian, A. C. C. Natali, and G. H. Travassos. Systematic review in software engineering. Technical report, Universidade Federal do Rio de Janeiro, 2005.

[4] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil. Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software*, 80(1):571–583, 2007.

[5] L. Chen, M. Ali Babar, and C. Cawley. Evaluation of variability management approaches: A systematic review. In *Accepted by 13th International Conference on Evaluation and Assessment in Software Engineering)*, Durham, UK, April 2009. British Computer Society.

[6] T. Dyba and T. Dingsoyr. Empirical studies of agile software development: A systematic review. *Information and Software Technology*, 50(9-10):833–859, 2008.

[7] T. Dyba and T. Dingsoyr. Strength of evidence in systematic reviews in software engineering. In *Proceedings of 2nd International Symposium on Empirical Software Engineering and Measurement (ESEM'08)*, Germany, October 2008. ACM.

[8] T. Dyba, T. Dingsoyr, and G. K. Hanssen. Applying systematic reviews to diverse study types: An experience report. In *Proceedings of 1st International Symposium on Empirical Software Engineering and Measurement (ESEM'07)*, pages 225–234, Madrid, Spain, September 2007. IEEE Computer Society.

[9] T. Dyba, V. B. Kampenes, and D. I. Sjoberg. A systematic review of statistical power in software engineering experiments. *Information and Software Technology*, 48(8):745–755, 2006.

[10] T. Dyba, B. Kitchenham, and M. Jorgensen. Evidence-based software engineering for practitioners. *IEEE Software*, 22(1):158–165, 2005.

[11] E. Engstrom, M. Skoglund, and P. Runeson. Empirical evaluation of regression test selection techniques: A systematic review. In *Proceedings of 2nd International Symposium on Empirical Software Engineering and Measurement (ESEM'08)*, Germany, October 2008. ACM.

[12] M. Jorgensen and M. Shepperd. A systematic review of software development cost estimation studies. *IEEE Transactions on Software Engineering*, 33(1):33–53, 2007.

[13] B. Kitchenham. Procedures for undertaking systematic reviews. Technical report, Keele University and National ICT Australia, 2004.

[14] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman. Systematic literature reviews in software engineering: A systematic literature review. *Information and Software Technology*, 51(1):7–15, 2009.

978-1-4244-4841-8/09/$25.00 ©2009 IEEE

Table 9. Example interview questions

Session	Example question
Working with Guidelines	
	What are the source(s) that you learned SLR from?
	What were the key documents to guide the execution of your SLR?
	Are you able to access any other kinds of help or advice (such as your colleagues) except the guidelines for your SLR?
Values of SLRs	
	Do you think SLR in SE will be as effective as in other disciplines? Why?
	What is the value to people conducting SLRs?
	Do you think SLR is appropriate for novices, esp. first-year research student? Please explain.
Experiences & Best Practices	
	If your colleagues are planning their own SLR, what are the best practices you want to share with them?
	From your point of view, what are the major reasons caused the problems in your systematic review*
	Will you perform another SLR in the near future? If so, how much do you expect to improve your productivity?
Challenges and Fulfillment	
	What encouraged and motivated you most in your SLRs?
	What frustrated you most in your SLRs?

[15] B. Kitchenham and S. Charters. Guidelines for performing systematic literature reviews in software engineering (version 2.3). Technical report, Keele University and University of Durham, 2007.

[16] B. Kitchenham, T. Dyba, and M. Jorgensen. Evidence-based software engineering. In *Proceedings of 26th International Conference on Software Engineering (ICSE'04)*, pages 273–284, Edinburgh, Scotland, UK, May 2004. IEEE Computer Society.

[17] B. Kitchenham, E. Mendes, and G. H. Travassos. Cross- vs. within-company cost estimation studies: A systematic review. *IEEE Transactions on Software Engineering*, 33(5):316–329, 2007.

[18] B. Kitchenham and S. Pfleeger. principles of survey research, part 1 to 6. *Software Engineering Notes*, 2001-2002.

[19] T. lethbridge. Studying software engineers: data collection techniques for software field studies. *Empirical Software Engineering*, 10:311–341, 2005.

[20] M. Patton. *Qualitative Evaluation and Research Methods*. Sage Publication, Inc., 1990.

[21] M. Petticrew and H. Roberts. *Systematic Reviews in the Social Sciences: A Practical Guide*. Wiley Blackwell, 2005.

[22] C. Seaman. Qualitative methods in empirical studies of software engineering. *IEEE Transactions on Software Engineering*, 25(4):557–572, 1999.

[23] D. Sjoberg, T. Dyba, and M. Jorgensen. The future of empirical methods in software engineering research. In *Proceedings of International Conference Software Engineering, Future of Software Engineering Track*, 2007.

[24] M. Staples and M. Niazi. Experiences using systematic review guidelines. *Journal of System and Software*, 80(9):1425–1437, 2007.

[25] M. Staples and M. Niazi. Systematic review of organizational motivation for adopting cmm-based spi. *Information and Software Technology*, 50(7-8):605–620, 2008.

[26] H. Zhang, B. Kitchenham, and D. Pfahl. Reflections on 10 years of software process simulation modelling: A systematic review. In *Proceedings of International Conference on Software Process (ICSP'08)*, pages 345–365, Leipzig, Germany, May 2008. Springer.

Appendix

Example Interview Questions (see above Table 9)

978-1-4244-4841-8/09/$25.00 ©2009 IEEE