Information and Software Technology 55 (2013) 1341–1354

Contents lists available at SciVerse ScienceDirect

Information and Software Technology

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

Systematic reviews in software engineering: An empirical investigation

He Zhang [a,*], Muhammad Ali Babar [b]

[a] University of New South Wales, Australia
[b] IT University of Copenhagen, Denmark

ARTICLE INFO

Article history:
Available online 4 October 2012

Keywords:
Systematic (literature) reviews
Evidence-based software engineering
Research methodology
Methodology adoption
Mixed-methods research
Tertiary study

Background: Systematic Literature Reviews (SLRs) have gained significant popularity among Software Engineering (SE) researchers since 2004. Several researchers have also been working on improving the scientific and methodological infrastructure to support SLRs in SE. We argue that there is also an apparent and essential need for evidence-based body of knowledge about different aspects of the adoption of SLRs in SE.

Objective: The main objective of this research is to empirically investigate the adoption, value, and use of SLRs in SE research from various perspectives.

Method: We used mixed-methods approach (systematically integrating tertiary literature review, semi-structured interviews and questionnaire-based survey) as it is based on a combination of complementary research methods which are expected to compensate each others' limitations.

Results: A large majority of the participants are convinced of the value of using a rigourous and systematic methodology for literature reviews in SE research. However, there are concerns about the required time and resources for SLRs. One of the most important motivators for performing SLRs is new findings and inception of innovative ideas for further research. The reported SLRs are more influential compared to the traditional literature reviews in terms of number of citations. One of the main challenges of conducting SLRs is drawing a balance between methodological rigour and required effort.

Conclusions: SLR has become a popular research methodology for conducting literature review and evidence aggregation in SE. There is an overall positive perception about this relatively new methodology to SE research. The findings provide interesting insights into different aspects of SLRs. We expect that the findings can provide valuable information to readers about what can be expected from conducting SLRs and the potential impact of such reviews.

1. Introduction

Systematic Literature Review (SLR), more commonly known as systematic review, has emerged as one of the most popular methods of Evidence-Based Software Engineering (EBSE) since Kitchenham, Dybå and Jørgensen reported their seminal piece of work on bringing the evidence-based practice to Software Engineering (SE) in International Conference on Software Engineering (ICSE) [21] in 2004. One of the evidences that the SLR has been gaining significant attention from the SE researchers is that there is a large, and continuously growing, body of published literature reporting SLRs on different topics of SE. The number of reported SLRs has grown so significantly that several researchers have also performed tertiary studies (i.e. systematic review of the SLRs on a particular topic or area). Apart from conducting and reporting over one hundred SLRs, SE researchers have also been focusing on improving the SLR methodology by, for example, providing the techniques for designing the strategies for assessing the quality of the reported primary studies included in a systematic review [9].

We have asserted that as the interest of SE researchers in SLRs is increasing, so should be the need for providing appropriate methodological guidance in designing, conducting, and reporting high quality systematic reviews. Moreover, the lessons learned from and experience reports of performing SLRs in SE have also been published. A large majority of the reported SLRs in SE has been carried out following the guidelines developed by Kitchenham and Charters [18,20]. Some researchers have also reported their own guidelines for SLRs [3] or have consulted the guidelines from medicine (e.g., [17,13]). 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 study this research methodology from different angles.

This paper reports our work aimed at building a body of knowledge about different aspects of SLRs in SE. The work reported in

* Corresponding author. Tel.: +61 2 9376 2227; fax: +61 2 9376 2023.
E-mail addresses: dr.hezhang@gmail.com (H. Zhang), maba@itu.dk (M. Ali Babar).

0950-5849/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.infsof.2012.09.008

this paper takes a multi-perspective view of the SLRs in SE over the past 7 years (2004–2010). The key objective of the work was to provide readers with a broad but real perspective of the adoption of SLRs in SE and its related factors. The intension was to systematically study various aspects of adopting SLR as a research methodology in SE and, more importantly, to gather and understand the perceptions and experiences from the methodology users – *systematic reviewers*, as well as to enhance the communications among SE researchers regarding SLRs and EBSE. To be specific, this study makes the following contributions:

- It provides a holistic status report on the adoption of SLRs in SE research from various perspectives after 7 years' practice.
- It presents a systematic reflection of the methodology adoption based on the methodology outcomes (SLRs) and their producers.
- It also reports an initial comparison of the use and potential value of two literature review methodologies in SE: Systematic Literature Review (SLR) and Traditional Literature Review (TLR).

In order to minimise the potential limitations of applying single research method, this research combined multiple empirical research methods, and integrated them systematically. For the reported research, we employed *semi-structured interviews*, *tertiary study* (using SLR methodology), and *questionnaire based surveys*. The use of mixed-methods approach enabled us to effectively build the evidence upon each other for our goals.

Unlike other 'roadmap' or 'overview' papers, this paper presents and summarises the evidence we collected from the real state of SLR's adoption in SE and the responses from the methodology users without author's subjective inputs. Hence, it is expected to enable readers to make their own conclusion based on these reflections with a minimised bias.

This article is an extended version of the conference paper published in ESEM 2011 [36]. Compared to the original version, the updates and enhancements made in this article are thorough. To be specific, this version (1) includes more description of research method, i.e. design, data collection, and implementation; (2) updates the tertiary study by replicating the search of SLRs by the end of 2010; (3) updates the impact analysis with the latest citations collected for the SLRs by 2010; (4) reports effectiveness of SLRs; (5) reports review types in terms of *systematic review* and *mapping study*; (6) reports quality assessment and roles in SLRs; (7) updates the related work with comparison with the latest tertiary studies in SE; (8) more in-depth discussion about the findings from this study; and (9) improves data presentations.

The rest of this paper is structured as follows. Section 2 sets the research context related to EBSE and SLRs, and defines our research scope and objectives. We enumerate the multi-perspective research questions, and describe our mixed-methods research design and implementation in Section 3. Sections 4–6 address the preliminary answers to the research questions about SLR's adoption in SE respectively. They are followed by the interpretation and discussion in Section 7. In the end, we conclude our study in Section 8.

2. Research context and objectives

This section introduces the context and motivations of the reported research by discussing the role of EBSE and SLRs in SE. We also define the scope of the reported research.

2.1. EBSE and systematic reviews

The main objective of EBSE is reported to be improving decision making for selecting software development technologies (i.e. methods, approaches, and tools) by gathering, evaluating, and synthesising current available evidence from research. EBSE is expected to narrow the gap between research and practice by placing emphasis on scientific rigour as well as practical relevance of research [11].

Systematic reviews are predominantly used for following EBSE. An SLR is "*a means of evaluating and interpreting all available research relevant to a particular research question, topic area or phenomenon of interest*" [11,20]. One of the main goals of a systematic review is to ensure that the review is methodical, repeatable, and thorough. An SLR also attempts to minimise the level of bias that can be prevalent in Traditional (ad hoc) Literature Reviews (TLRs).

There are an increasing number of SLRs being performed in SE since 2004. SE researchers have also provided methodological support by developing guidelines for performing SLRs [20,3] and have reported lessons learned in order to share knowledge and experiences [4,10]. Several researchers have also identified the areas for improving the published guidelines and the needs for supportive techniques (e.g., [38]). The increasing trend to use SLRs highlights the need for providing appropriate knowledge and training in different aspects of SLRs [30]. That means SE researchers need to allocate significantly more resources to develop suitable support system for guiding SE researchers on how to design, conduct, and report high quality systematic reviews in SE and practitioners on how to assess the quality and results of published SLRs on a topic that interests them. At the same time, there is also a need for understanding the use and adoption of systematic reviews in SE and any challenges that researchers are facing while using this methodology.

2.2. Research scope and objectives

Systematic reviews claim rigour and relevance as their strengths and promise to provide the mechanism needed to assist practitioners to adopt appropriate technologies and to avoid inappropriate technologies [21,11]. However, this research is not aimed at exploring the use and adoption of systematic reviews by SE practitioners. Rather the focus of the reported research at this stage is the SE research community.

There are two main reasons that exclude the external validation of SLRs from our research scope at this stage. First, as SLR was introduced to SE as a new research methodology in 2004, many practitioners in software industry have yet to know this research methodology well (this is also true even for many SE researchers). Second, though many SLRs have been reported in SE in the recent years, the number of explored topics and questions are still limited, and the distribution of SLRs over these topics is not even. The technology practitioners may be unable to freely find an SLR, which exactly matches their own interests or questions at present. Given these two constraints, we are not in a position to properly examine the SLRs' external relevance at this stage.

Hence, the objective of the reported research is to carry out an internal validation of SLR methodology within SE research community. Unlike in medical discipline, researchers are expected to be the main users of SLR methodology and evidence-based practice in SE. Therefore, an internal validation of this methodology should necessarily start within SE research community, particularly the literature reviewers irrespective of whether or not they have used SLR.

This research has been motivated by an increasing recognition of the need and importance of providing methodological and technological infrastructures for maximising the exploitation of potential benefits of EBSE in general and SLRs in particular. We assert that any such effort would greatly benefit from a good understanding of the use, value, experiences and challenges involved in

performing SLRs in SE. Thus, we decided to systematically study the adoption and use of SLR methodology in SE using multiple (mixed) research methods such as tertiary study and surveys (interviews and online questionnaires).

2.3. Tertiary studies on systematic reviews in SE

As SLR is a relatively new research methodology in SE, merely a very limited number of tertiary studies were reported. In terms of their research questions, we summarise two important series of tertiary studies and their relations to this research.

Kitchenham and her colleagues conducted a continuous tertiary study to provide an overview of the systematic secondary studies (i.e. systematic reviews, mapping studies and meta-analyses) related to EBSE in 2007 [19] and 2009 [23]. Later, by following the same research questions, da Silva et al. updated this thread of tertiary study until the end of 2009 [7]. These three reports identified 120 SLRs in total. Whereas, with the same research questions given, we found there exist different selection criteria applied by the two groups of authors. As a result, some excluded studies in [19,23] (with the reasons reported) were selected as SLRs in [7].

Cruzes and Dybå performed another tertiary study to assess the types and methods of research synthesis in systematic reviews in SE [6]. They restricted their search scope and selection criteria with purpose to investigate the SLRs explicitly influenced by EBSE [21,11] and the SLR guidelines [18,20].

However, these tertiary studies merely focused on the SLRs published in SE. Our research sought both systematic reviews (SLRs) and traditional reviews (TLRs) as a basis for investigating the methodology adoption and impact. We also covered the methodology users' (SE researchers') perceptions and experiences, which are also important for a new methodology adoption and improvement. Hence, our study is able to provide a more holistic view of SLR methodology adoption in SE.

In addition, compared to the SLRs identified in the other tertiary studies [19,23,6], our research systematically produced a more recent and thorough list of the SLRs available to SE researchers and practitioners. Table 1 shows the difference of search scope and results between our tertiary study and the previous ones.

Given the above limitations, a timely research on investigating and reflecting the methodology adoption of SLR in SE to the present is considered to be necessary. Instead of an amended tertiary study of [19,23], this research is a systematically designed empirical study by combining evidence-based practice (tertiary study) and typical empirical methods (surveys).

3. Research methodology

3.1. Research questions

As the intended post-mortem review of the past 7 years' adoption of systematic reviews in software engineering, we investigated this methodology from a multi-perspective. The research questions discussed in this paper are as below.

RQ1. What is the value of SLR for SE? Why did (or did not) SE researchers do SLRs?
RQ2. What SE topics have been addressed by what types of SLRs? What has the influence of SLRs been in SE research?
RQ3. How did SE researchers perform SLRs (in terms of, for example, rigour and effort)?

After the implementation of the research design (described in the rest of this section), the above questions are answered in the following sections.

3.2. Research design

The research design in this study applied a *sequential exploratory* strategy, i.e. a mixed-methods research design that is characterised by the collection and analysis of qualitative data followed by the collection and analysis of quantitative data [5]. Its purpose is to use quantitative data and results to assist in interpreting, confirming and enhancing the qualitative findings when answering the research questions. To be specific to this research, we first generated the preliminary findings based on the qualitative data gathered by interviewing a small but representative sample of the methodology users, then confirmed and enhanced our understanding about the adoption of systematic review by surveying a larger number of SE researchers and reviewing relevant secondary studies.

This research explicitly distinguishes two types of literature reviews: *systematic (literature) reviews* and *traditional (literature) reviews*, which are carried out in an informal or ad hoc style. Accordingly, the literature review performers can be grouped as *systematic reviewers*, who use the SLR methodology for performing literature reviews; and *traditional reviewers*, who undertake the traditional (narrative) style literature reviews.

Our research design was based on one important assumption that each of the (co-)authors of an SLR has participated in at least one phase of the reported SLR (e.g., *reporting review*) as a systematic reviewer. A similar assumption has been made about the traditional reviewers.

Fig. 1 illustrates the used research methods and the connections between them. In order to systematically investigate the use of SLRs in SE, our research design incorporated multiple empirical research methods: *semi-structured interview*, *systematic literature review* (tertiary study), and *questionnaire-based survey*.

3.3. Research methods and data gathering approaches

We have mentioned that we decided to use mixed-methods approach based on the objectives of the reported research. Our research design involved two main research methods, survey and systematic review (tertiary study), and a set of suitable data gathering approaches (i.e. semi-structured interview and questionnaire-based online survey). Survey research can use one or a combination of several data gathering techniques such as interviews, self-administered questionnaires and others [25]. A survey research method is considered appropriate for gathering self-reported quantitative and qualitative data from a suitable number

Table 1
Comparison between our tertiary study with the previous ones.

Tertiary study	Topic/purpose	Search strategy	Time span	Incl. SLRs
Kitchenham et al. [19]	Tertiary study in SE	Manual only	January 2004–June 2007	20
Kitchenham et al. [23]	Updating [19]	Automated only	January 2004–June 2008	33
da Silva et al. [7]	Updating [19,23]	Automated + manual	January 2008–December 2009	67
Cruzes and Dybå [6]	Research synthesis	Automated + [19,23]	January 2005–July 2010	49
This research	Tertiary study in SE and identifying systematic reviewers	QGS-based (manual + automated)	January 2004–December 2010	148

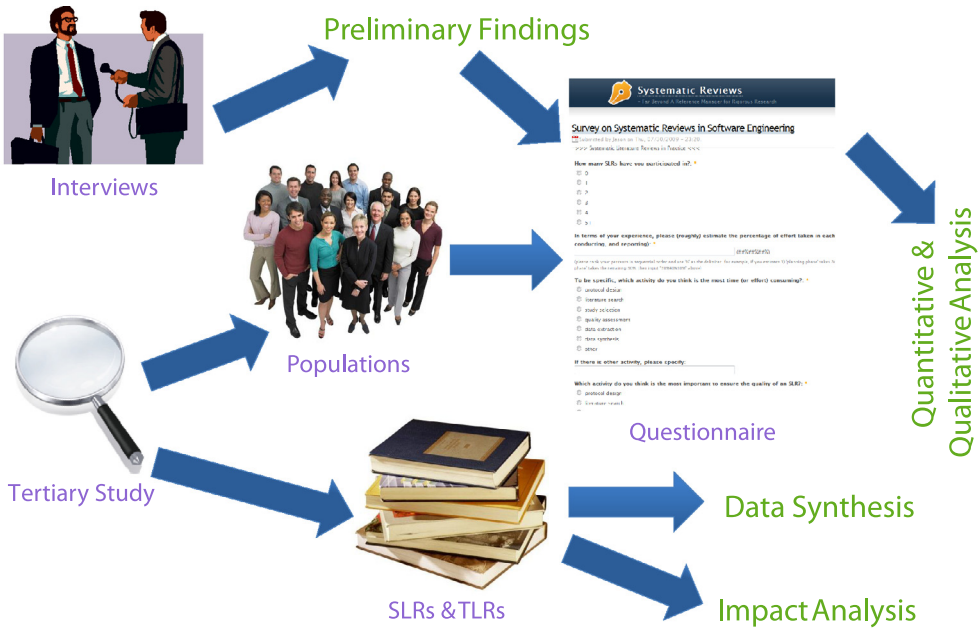

Fig. 1. Research design and methods.

of respondents, i.e. a relatively smaller number in case of interviews and a larger number in case of questionnaire-based online survey [24]. The collected data are more easily amenable for analysis by using simple as well as complex statistical approaches depending upon the research questions and the survey instrument. We decided to use semi-structured interviews and questionnaire-based online survey as the data collection approaches.

We used semi-structured open-ended interviews to collect qualitative data. One reason for choosing the 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. Interviews are considered an effective mechanism of gathering detailed information required to find the answers to the questions that had motivated part of the reported 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 this research. 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 [33]. 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 [33].

We designed our interviewing instrument with the intention of keeping the discussion focused and using the interviewee's time effectively [27]. 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 [27]. However, such interviews may reduce the flexibility and spontaneity because the probing during the interview is kept limited. However, 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 [27]. We also performed a pilot test of our interview instrument and time estimation. Our interviews were designed to take between 70 and 90 min.

At the beginning of this research, we started with exploring the perceptions and experiences of systematic reviewers using semi-structured interviews (the top-left in Fig. 1). The '*preliminary findings*' from the analysis of the data gathered through the first stage of interviews were partly reported in [1].

The interviews were to be followed by a tertiary study for which the research questions were supposed to be based on the '*preliminary findings*'. This phase was a *systematic review* of SLRs reported by the middle of 2009. Different from the tertiary studies reported by Kitchenham et al. [19,23], the search performed in this study had a broader scope (seeking both SLRs and TLRs) and were later updated till the end of 2010. Both types of literature reviews and their corresponding performers, i.e. *systematic reviewers* and *traditional reviewers*, were identified and grouped after the searches (the '*populations*' in the centre of Fig. 1). We extracted and synthesised the data from the identified SLRs only.

Based on the *preliminary findings* from the first stage interview, we designed two questionnaires, one for *systematic reviewers* and the other for *traditional reviewers*. The questionnaires were published as web surveys after a short internal trial. The two groups of reviewers, whose contact details were extracted from the tertiary study, were invited to participate in the surveys. The gathered qualitative and quantitative data were analysed respectively (the rightmost in Fig. 1). In the meantime, the citations of the identified SLRs and TLRs were collected for the impact analysis (the bottom-right in Fig. 1).

Finally, the results gathered from different research methods were put together to generate the findings: value and motivations (by combining the reflection from the interviews and surveys); review types, topics and impact (from the tertiary study and impact analysis), and practice and experiences (from the surveys and tertiary study). These findings provide the evidence-based answers to the research questions of SLR's adoption in SE.

3.4. Research implementation

According to the research design and methods, this study consists of a number of steps, each of which was supported by one

empirical method. They were sequentially connected to form the research process. One step was based on its former steps, and the outputs from one step might become inputs to the following steps. In the rest of this paper, we use the label numbers in Fig. 2 to indicate individual step of the research process.

This subsection describes the technical process of the implementation of our mixed-methods research project. We provide a brief description of each research method and step carried out in this program.

3.4.1. Semi-structured interviews

Based on the pilot literature search and our knowledge of the researchers active in SLRs, we invited a number of researchers in SE to participate in two stages of interviews. The invitees can be classified into three categories: *advocates* who introduced SLR methodology and evidence-based practice into SE and published many SLRs, *followers* who were experienced SE researchers and had participated in at least one SLR, and *novices* who were research students when performing SLR(s) [1].

In the first stage, we invited 24 researchers, and 17 invitees agreed to be interviewed (cf. [1] for details of the first stage of the interview process). In the second stage, we further interviewed nine more researchers (eight *followers* and one *novice*) identified in the search of SLRs ②. In total, 26 SE researchers were interviewed and all the interviews were transcribed for analysis.

3.4.2. Tertiary study

After the first stage of interviews, we carried out a tertiary study, i.e. an SLR that sought the secondary studies (SLRs and TLRs) between 2004 and the middle of 2009. This tertiary study was later extended ⑧ till the end of 2010 and updated twice in 2011 (January and December) after the data analysis of the interviews and surveys ⑦.

3.4.2.1. Search for SLRs. We employed a Quasi-Gold Standard (QGS) based systematic literature search approach [38], systematically integrating manual and automated search, for identifying the SLRs in SE. For manual search, we screened the publication venues related to empirical software engineering, EBSE, and premier SE venues, such as EMSE, ESEM, EASE, TSE, IST, JSS, ICSE and IEEE Software. It was followed by an automated search through five of the major publishers' digital library portals in SE [37]: IEEE Xplore, ACM Digital Library, ScienceDirect, SpringerLink, and Wiley InterScience. The search string used in automated search is as below, but coded corresponding to the syntax of specific library's search engine.

```
(software OR code OR program) AND (((systematic OR structured OR exhaustive OR controlled) AND (review OR survey)) OR (literature search) OR (meta-analysis) OR (meta analysis) OR (mapping study) OR (scoping study) OR (evidence-based) OR (evidence based))
```

It is the improved search string from the one used in [37]. When applying this string in searching the SLRs in the QGS (from the manual search), its quasi-sensitivity can reach 85%.

The search strategy was developed by the two authors. The first author performed manual and automated search in order, followed by a thorough checking done by the second author. All disagreements on search and selection were solved by the two authors in discussion.

Note that some secondary studies selected (as SLRs) in [19,23,7] were not considered as SLRs in our tertiary study, as we followed a set of stricter criteria (cf. [37]) to appraise if a study is an SLR. For example [35], was excluded from our tertiary study as its random sampling of ICSE papers may be not repeatable by other researchers, which may results in different conclusions.

3.4.2.2. Search for TLRs. Unlike SLRs, the publication of TLRs is much scattered among venues. Hence we applied only automated search to seek TLRs in SE published during the time span through the above-mentioned five digital portals. The following search string was used in searching TLRs.

```
software AND (overview OR state-of-the-art OR road-map OR ((literature OR study) AND (review OR survey)))
```

The purpose of the search of TLRs is to find typical TLRs published in SE, and further to identify representative population of the users of traditional literature review for survey research and impact analysis. Therefore, unlike the search of SLRs, the sensitivity of the above search string was not evaluated because an exhaustive search of TLRs is unnecessary for the research objectives.

We identified a TLR by its usual structure which begins with presenting the earlier work in a particular area and proceeds with the most important past work up to the present [29], and the applied approach that does not use formalised methods of systematic review [28]. In terms of our observation, TLR is often combined with other research methods, such as survey, case study, or framework development. We excluded papers with TLR section if the literature review was not the major contribution. All included TLRs were grouped into two clusters: *full* TLR, which solely reports a literature review; and *partial* TLR, in which literature review is one of the two major contributions reported. Apart from literature review, the other possible contributions of the *partial* TLRs include typically, for example, conceptual model or framework,

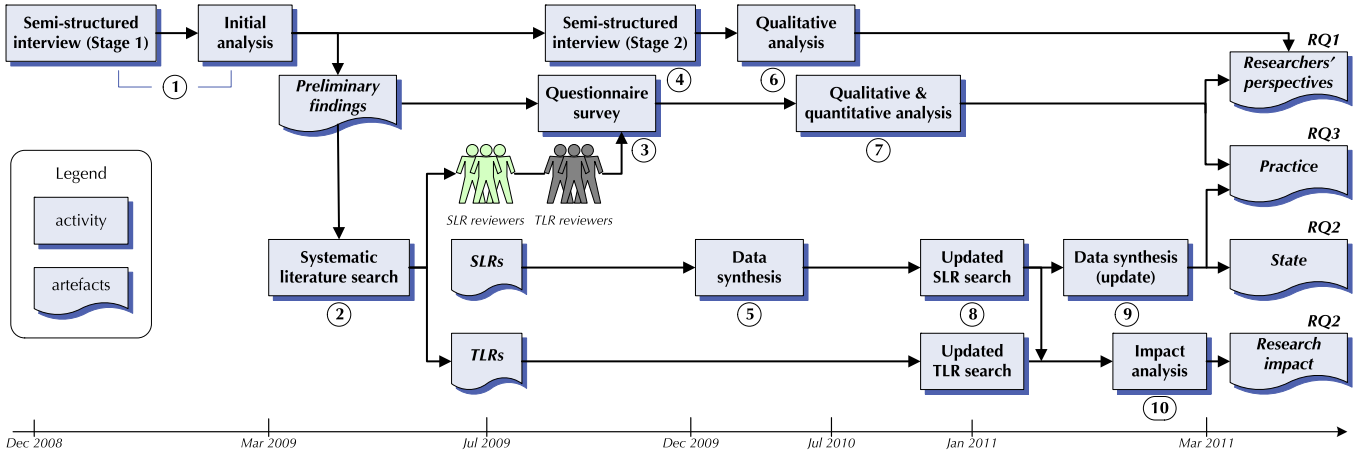

Fig. 2. Research implementation process.

meta-model, case study, survey or interview, experience report, guidelines, and taxonomy or ontology.

3.4.3. Questionnaire based surveys

Based on the preliminary findings from ①, we carried out web-based anonymous surveys using two questionnaires with comparable questions for the *systematic reviewers* and *traditional reviewers* respectively. The questionnaires[1] were developed by following Kitchenham and Pfleeger's guidelines for survey research in SE [24]. We extracted the potential participants of the survey study from the review studies (through the literature search ②) as those reviews were authored by researchers who could make up a sample set very close to our target population. After an internal trial with our colleagues and an external check by SE empiricists, the questionnaires were published online and remained open for gathering responses for a few months.

3.4.3.1. Survey for systematic reviewers. The SLRs identified by our literature search ② were reported by 124 authors. In order to minimise the sampling error and selection bias of the survey study, we maximised the sample size as close as possible to the target population (systematic reviewers in SE) by extracting their contact information from the publications. We found 4 authors without email address in their papers. Though both authors of this paper have also published some SLRs, we excluded ourselves from the survey to prevent potential researcher's bias. As a result, 118 (124-4-2) invitations to participate in the survey were distributed to the identified *systematic reviewers*. We received 12 messages of non-delivery due to unknown email addresses, and 3 auto-replies that mentioned the invitees were unavailable during the survey period.

We received 52 responses to the survey with a response rate of 50% (52/103). Only one author explicitly rejected our survey invitation.

3.4.3.2. Survey for traditional reviewers. During the search of TLRs ②, we even found more authors of (*full* plus *partial*) TLRs than of SLRs. However, after excluding the authors who had also reported SLRs, we were left with 109 *traditional reviewers*. We extracted their contact information from the publications, and sent out 98 (109-11) invitations to this survey to the authors of TLRs (11 authors without email addresses in their papers). We received 14 non-delivery messages and 2 auto-replies due to their unavailability.

We received 27 responses to this survey and the resulting response rate was 33% (27/82). There was again one author who explicitly rejected our invitation.

As mentioned above, the survey questionnaires were designed to collect both quantitative and qualitative data. The gathered quantitative data were entered into MS Excel and MiniTAB for statistical analysis. The qualitative data from the surveys were coded and analysed by the first author using NVivo, and further validated by the second author.

As the surveys were anonymous, no respondents' names and their affiliations were collected in the questionnaires. Fig. 3 shows the geographic distribution of the two groups of invitees to the surveys.

3.4.4. Impact analysis

The SLRs and TLRs in SE were systematically searched ② and updated ⑧ by the end of 2010 twice in 2011. In [36], the SLRs published in 2010 were excluded from impact analysis due to the short observation period for citations, which is no longer an issue in this article. In order to avoid the citation for the contributions other than literature review in the *partial* TLRs, we also excluded them from impact analysis.

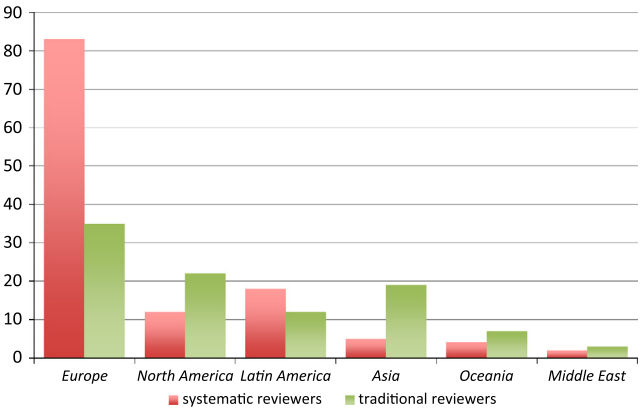

Fig. 3. Geographic distribution of the survey invitees.

The impact analysis reported in this paper completely updates the corresponding part in [36]. The citations received by all SLRs and *full* TLRs published by the end of 2010 were collected through Google Scholar in February 2012, which provided at least 13 months for observing the impact of the both types of literature reviews published by the end of 2010.

4. Value and motivations (RQ1)

This section answers the RQ1 (*What is the value of SLR for SE? Why did or did not SE researchers do SLRs?*) by synthesising the advantages and strengths of SLRs valued by the methodology users; presents the systematic reviewers' motivators for performing SLRs and the de-motivators for "*why did not traditional reviewers do SLRs?*"; as well as reflects the SE researchers' opinion about the effectiveness of SLR in SE. They are from the interviews ① ④ and the online surveys ③ in this research.

4.1. Value of SLRs

The SLR guidelines [20] state several advantages (strengths) of SLRs: (1) well-defined methodology with less bias, (2) effects across a wide range of settings and methods, and (3) possibility of performing meta-analysis. One obvious disadvantage (weakness) of SLR is that it requires considerably more effort than TLRs. Here we compare the reflections from the categories of literature reviewers in SE with respect to the above mentioned claims.

We have analysed the qualitative data gathered from the respondents to open-ended questions, which were used to mainly capture the explanations for responses provided using the Likert scale and different aspects of SLRs in SE. Our analysis of the responses to the question about the value of SLRs in SE reveals that a large majority of the respondents, both *systematic reviewers* (93%) and *traditional reviewers* (89%), identified several values of SLRs; an overwhelming number of respondents were convinced that SLRs provide a systematic way of building a body of knowledge about a particular topic or research question. Other valuable aspects of SLRs reported by the survey respondents were '*more reliable findings based on synthesis of literature*', '*repeatability*', '*identification of problem areas for new research*', and '*a source for supporting practitioners' decisions about technology selection*'. Our analysis also showed that most of the claimed advantages of SLRs [20] were confirmed by the respondents of our survey.

[1] The surveys were published online at http://systematicreviews.org.

Table 2
Motivators for doing SLRs.

Motivator	Adv.	Fol.	Nov.
New research findings and ideas from SLR	0	5	2
Clear statement and structure of state-of-the-art	1	3	1
Learning from studies and getting knowledge	1	2	1
Recognition from the community	0	3	0
Paper publication (e.g., motivated by IST)	0	1	3
Working experience	0	1	0
Learning research skills (SLR methodology)	0	2	0

Table 3
Reasons for not doing SLRs.

De-motivator	Response
Did not know SLRs at the time of past literature review	14(52%)
Systematic reviews are time-consuming	10(37%)
Narrative review is more appropriate to my study	7(26%)
Other	5(19%)

When asked about the potential strengths of SLRs compared to TLRs, most of the respondents reported similar concerns as mentioned in response to the question about the value of SLRs such as a transparent and systematic approach, comprehensive and traceable review of the available literature, identification of more relevant sources of literature, and a basis of drawing reliable and unbiased conclusions. We found that the responses from the traditional literature reviewers also identified similar strengths of SLRs such as '*well-defined process*', '*objective selection of the papers*', and '*traceable and reliable findings*'. These findings identify some of the value and strengths of SLRs perceived by the participants of our surveys.

4.2. Motivators and de-motivators for SLRs

4.2.1. Motivators

Apart from the above-mentioned value and advantages of SLR in SE, the systematic reviewers also shared their more specific motivators. Table 2 enumerates the top motivators and encouragements that SE researchers (*advocates*, *followers* and *novices* [1]) reported for doing SLRs during the interviews ① and ④.

The most important motivators for conducting an SLR are '*getting new research findings and ideas from the results of SLR*' (based upon an overview or synthesis of studies in a particular area/topic), '*clear statement and structure of state-of-the-art*', and '*learning from studies and getting knowledge*', all of which are related to reviewer's research interests. The top motivator is further confirmed by 80% of systematic reviewers in ③ that SLRs sometimes bring them new research innovations *unexpectedly*.

4.2.2. De-motivators

Table 3 shows the responses to the surveys ③ when the traditional reviewers were asked "*why did not you apply SLR methodology to your previous literature review?*". About 50% of the respondents '*did not know SLR methodology when they did the past (ad hoc) literature reviews*'. One quarter of them believed that a narrative review was more suitable for their previous study. Among the 5 respondents chose '*other*', three of them thought '*the study area is relatively small*' and/or '*insufficient data were available*', for which SLR might not be suitable.

Further, when the traditional reviewers were asked "*if you could go back and redo your literature review (survey), would you carry out an SLR instead of traditional (narrative) literature review?*", 74% of the respondents (traditional reviewers) gave positive answer, and showed their intent to carry out an SLR in their future literature reviews.

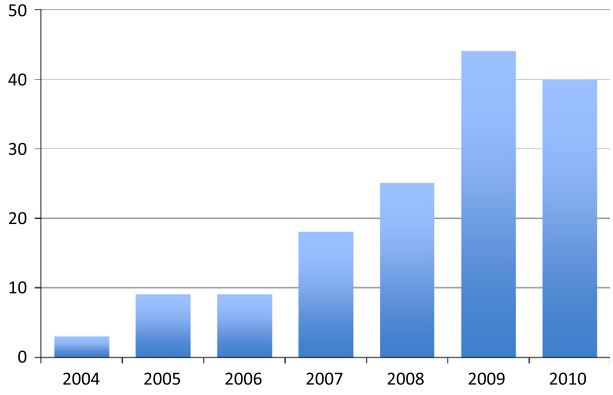

Fig. 4. Numbers of *new* SLRs per year.

4.3. Effectiveness of SLRs

Over half of systematic reviewers (53%) believe that SLRs in SE can be as effective as in other disciplines. About 22% of the respondents were more pessimistic about this questions. The remaining respondents were '*not sure*'.

The traditional reviewers also showed their confidence in SLR's effectiveness in SE. About 50% of the respondents replied '*yes*', even they had not tried SLR methodology themselves. Only 11% of the respondents replied '*no*' to this question.

5. Reviews and impact (RQ2)

This section addresses the RQ2 (*What SE topics have been addressed by what types of SLRs? What has the influence of SLRs been in SE research?*) by summarising the SLRs reported by review type (*systematic review* or *mapping study*), topic (research area) and reporting year. This question also investigates the SLR methodology diffusion and impact in SE. The evidence has been mainly extracted from the tertiary study (②, ⑤, ⑧ and ⑨), and also generated through the impact analysis ⓪, i.e. citation analysis based on the number of time a paper has been cited since its publication.

5.1. SLRs reported in SE

Our systematic search (② and ⑧) found 148 SLRs[2] reported in 160 publications between 2004 and 2010. Note that the search found a number of SLRs also reported as grey publications (e.g., technical reports and theses). However, the number of SLR publication only counts the peer-reviewed articles. Fig. 4 shows the number of *new* SLRs in SE per year has been continuously increasing since its introduction to SE. There were two significant jumps of the number in 2007 and 2009, and a slight drop in 2010.

As mentioned in Section 3, our tertiary study applied a systematic literature search method (QGS-based search [38]) and stricter criteria in carefully identifying SLRs compared to the previous tertiary studies [19,23,7]. As a result, some secondary studies included in those previous reviews were not considered as SLRs in this research. For instance, in [7] some selected reviews conducted systematic search only but missing most of the other necessary elements (normally required in an SLR protocol), e.g., research

[2] The full list of SLRs in SE is available at http://mendeley.com/groups/965201/systematic-reviews-in-software-engineering/papers/.

questions, selection criteria, and/or attributes for data extraction. In addition, some review studies were incorrectly identified as SLRs, but they are in nature other review styles, e.g., [12] reported an integrative review [32] instead of an SLR.

5.2. Review types

The SLR guidelines [20] classify SLRs, to be exact *systematic secondary studies*, into four different types: *(conventional) systematic review* (CSR) that identifies, analyzes and interprets all available evidence related to specific research question(s) in an unbiased and repeatable manner; *(systematic) mapping study* (SMS, or *scoping study*) that is a broad review to identify what evidence is available on the topic (rather than specific research question); *meta-analysis* (MA) that is a form of systematic review where research synthesis is based on quantitative statistical methods; and *tertiary study* (TS) that is a review of secondary studies. In total 148 SLRs, we identified 72 systematic reviews, 72 mapping studies, 3 tertiary studies, and 1 meta-analysis.

Interestingly, our systematic search found exactly equal number of systematic reviews and mapping studies reported in SE. However, when looking into the detailed trend of these two dominant review types over the past years (shown in Fig. 5), it can be observed that the number of mapping studies has grown faster than systematic reviews in the past years.

Two formats are suggested for reporting SLRs in the guidelines [20]: technical report or a section of thesis, and journal or conference paper. In this research, we merely selected the peer-reviewed SLR publications. For article types, we found the SLRs were published in common peer-reviewed formats, i.e. journal (or transaction) article, conference (full or short) paper, workshop paper, poster, and book chapter. The most published article types of SLRs are journal articles (66) and conference full papers (62). In terms of the number of SLRs published, *Information and Software Technology* (IST) is the top venue and published 30 SLRs in the search time span (2004–2010). Table 4 shows the article types of SLRs per year.

5.3. Research topics

Our systematic search of SLRs updates the research topics identified in the previous tertiary studies [19,23,7]. Table 4 presents a concise landscape of the reported SLRs on a variety of topics in SE per year. We identified over 30 SE research topic areas addressed by SLRs. Among them, *global development*, *cost estimation*, *requirements engineering*, *empirical methods*, and *agile development* are the most investigated topics attracting systematic reviewers' research interests.

For future SLR topics, in response to online survey, systematic reviewers and traditional reviewers indicated 38 topics in SE on which they expect to see further SLRs conducted and reported. Table 5 lists the top suggested topics (with 4 or more votes) for future SLRs. The most expected topics include *requirements engineering*, *software process improvement*, and *agile development*. Note that 12 respondents showed their interests in SLRs on any topics in SE as, for example, "*we need good overviews of our knowledge*".

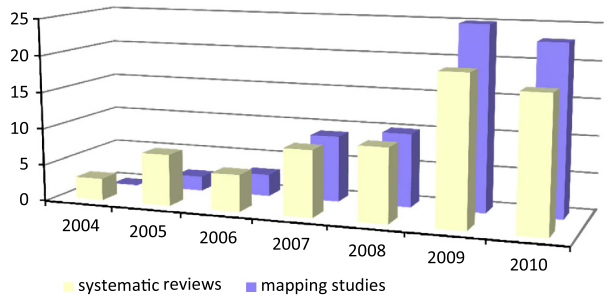

Fig. 5. Systematic reviews vs. mapping studies per year.

5.4. Impact of SLRs

To investigate SLR's diffusion and impact in SE, we collected the distribution of the *systematic reviewers* and citations of the SLRs *vs.* TLRs in the impact analysis ⓪.

5.4.1. Dissemination among researchers

Based on the systematic reviewers extracted from the literature search ② and ⑧, we examined the diffusion of SLR methodology among the populations.

Fig. 6 shows the geographic distribution of systematic reviewers from 2004 to 2010. As a newly introduced research methodology, it has convinced a large number of SE researchers in Europe. Nevertheless, the numbers of SLR users in the other regions were still quite low.

5.4.2. Impact on researches

The citation information about both SLRs and TLRs was collected through Google Scholar in February 2011, and updated in February 2012. Fig. 7 explicitly compares the average citations received by the SLRs and the *full* TLRs per publishing year. It is obvious that the average citations of SLRs are higher than *full* TLRs in all years except 2009. This phenomenon to certain extent reveals the higher impact of SLRs on SE research compared to TLRs. By looking into the 2009 data, we found that several TLRs (literature surveys) published in the premier venues (e.g., ACM Computing Surveys) received very high citations that lift the average citation of TLRs in 2009.

Fig. 8 presents the statistical distribution of the citations received by the SLRs from 2004 to 2010 (collected from ⑦). It is noted that the standard deviation of the citation for SLRs varies significantly, which may imply the quality of SLRs was not so stable or some topics (or research questions) were not so attractive to other researchers.

In order to investigate the above deviation among the reported SLRs, we further compared the citations received by SLRs but published in two major article types, i.e. journal and conference. Though our search found the close numbers of SLRs published in journals (66) and conferences (62) during 2004–2010, Fig. 9 shows the big difference of the impact between the two main publication types. The average citations of the SLRs published in journals are much higher than in conferences in every single year, which probably gives one reason resulting the significant overall deviation.

From our analysis ⓪, the most 'influential' (cited) SLRs, that received over 100 citations, are shown in Table 6. It is noted that all of them are journal publications.

6. Practice and experiences (RQ3)

Methodological rigour is one major claimed strength of EBSE [21]. It is also known that SLR requires considerable amount of effort and expertise compared to TLR [26]. Accordingly, rigour and effort are two distinct characteristics of SLR in practice. These characteristics always need to be balanced in the course of performing an SLR. In addition, unlike TLR, teamwork is necessary for controlling and minimising the potential bias during review. However, there is no guidance on forming a review team that includes people with the required diversity [26] in the SLR guidelines for SE [20].

This section concentrates on the above three distinct characteristics of SLR to investigate the systematic reviewers' experience in performing SLRs (i.e. RQ3, *How did SE researchers perform SLRs?*).

Table 4
Overview of SLR studies and publications in software engineering research by topic and year.

Rank	Review topics	2004	2005	2006	2007	2008	2009	2010	Sum
1	Global development					◊◊◊	◆[a]◊◊▽	◆[a] ◆[a]◊◊◊◊◊◊▽	13/16
1	Cost estimation	◆	◆◊◊▽	◆◊	◆◆◆[a]◊◊		◊▽		13/14
3	Requirement engineering			◊	◊	◊	◆◆◆◊◊[a]▽▽	◆◆◆	12/13
4	Empirical methods		◆◊▽	◆	◆◆▽	◆▽	◆▽		11/11
5	Agile development					◆	◆◊◊	◆◊◊◊▽▽	10/10
6	Inspection and testing	◆		◆	▽	◆[a]	◆◆[a]◊◊	◆◆◊	9/11
6	Software product lines					◊▽	◆◊◊◊◊◊	◆	9/9
8	Software process improvement				◆◊	◆◆◊	◊	◆◊	8/8
9	Software architecture			▽		▽	◆[a]▽	◆◊	5/6
10	Software process modelling			▽		◊	◊	◊◊[a]	4/5
10	Open source development						▽◊[a]	◆◊◊	4/5
10	Software measurement			◊		◆◊		◆	4/4
10	Program analysis						◆◆◊	◊	4/4
14	Model-based development					◊	◆[a]	◊▽	3/4
14	Tertiary study						◆	◆[a]◊▽	3/4
14	Software maintenance				◆	▽	◊		3/3
14	Software tools				◆▽		◆		3/3
14	Software security					◆◊		◆	3/3
14	Web engineering		◊		◊	▽			3/3
14	Software outsourcing						◆◊◊		3/3
14	Human-aspects (e.g., motivations)					◆	◆◆		3/3
22	Software design				◊		◊		2/2
22	Unified modelling language					▽	◆		2/2
22	Software evolution				◆			◊	2/2
22	Aspect-oriented programming						◊	◆	2/2
22	Business process			◊				◆	2/2
22	SE research in general	◆				◆			2/2
	Other topics		◊		◆	◆	◆	◆◆	6/6
Total	(*new/all SLR reports*)	3/3	9/9	9/9	18/19	25/26	44/50	40/44	**148/160**

◆ journal paper or book chapter, ◊ full conference paper, ▽ workshop or short paper.
[a] Update/extension of previous SLR report.

Table 5
Expected research topics for future SLRs.

Rank	Topic to be reviewed	Votes
1	Requirements engineering	15
2	Software process improvements	10
3	Agile software development	9
4	Software testing	8
4	Software cost estimation	8
6	Software process (modelling)	7
7	Software metrics/measurement	4
7	Software quality	4
7	Software product lines	4

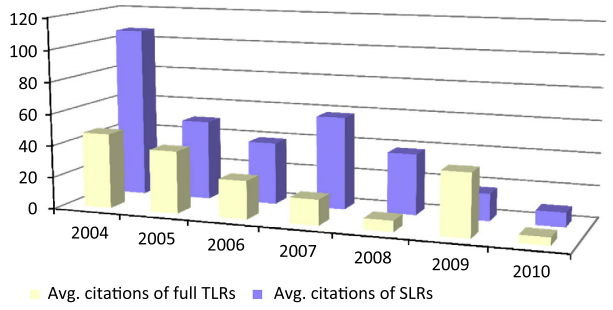

Fig. 7. Average citations of SLRs vs. *full* TLRs per year.

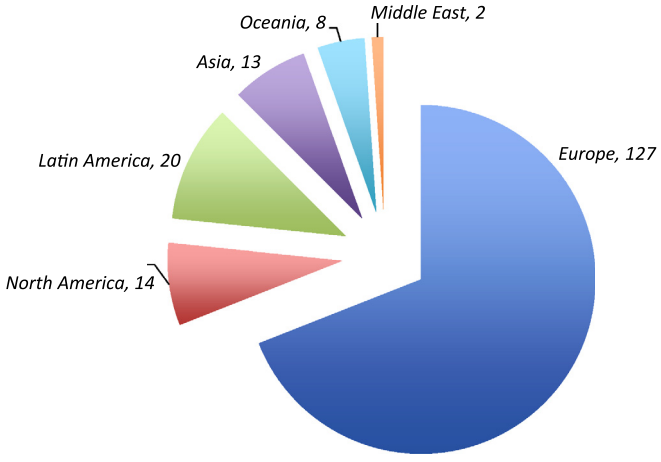

Fig. 6. Geographic distribution of *systematic reviewers*.

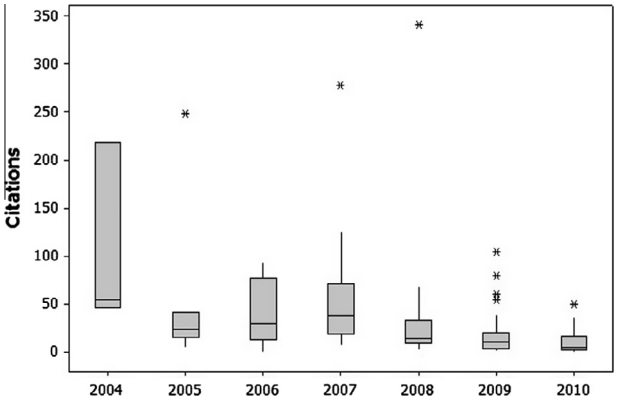

Fig. 8. Citations of systematic reviews per year.

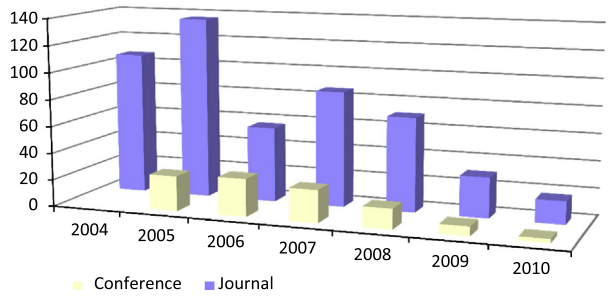

Fig. 9. Average citations of SLRs between journal and conference.

We distilled the results and findings focusing on these three aspects, which are mainly collected from activity ⑤, ⑦ and ⑨, in order to benefit the current and future systematic reviewers in SE.

6.1. Rigour

Rigour has been claimed to be one of the top strengths of SLRs as it is expected to ensure quality. We were interested in revealing how the methodological rigour is implemented in SLRs.

6.1.1. Critical activity

Many systematic reviewers (35%) believe that *protocol design* is the most important activity to ensure the rigour of an SLR. It is followed by *study selection*, *quality assessment*, and *data extraction*, which received 18%, 14% and 14% support respectively. Only one respondent thought *data synthesis* is the most critical for SLR quality.

6.1.2. Bias control

In order to minimise the potential bias during the review process, *peer-review* is the most common method used by 80% systematic reviewers. Many reviewers also sought help from *external checkers* (33% responses), conducted *self-review* (29%), or validated the agreements by *statistical techniques* (24%), e.g., Kappa.

6.1.3. Quality assessment

Compared to TLRs, quality assessment is a distinct and critical step in SLRs to investigate the strength of individual piece of evidence and weight their importance in data synthesis [20]. The tertiary study ⑤ found about 44% SLRs in SE explicitly reported their quality assessment (QA) results or included the checklists for evaluating the quality of primary studies. Among the respondents to the survey ③ who ever performed quality assessment in their SLRs, 50% defined the QA checklists by themselves. Some systematic reviewers tried to adopt the QA systems from other disciplines, medicine in particular, to SE, such as GRADE (10%), DARE (2%), and MOOSE (2%). Nevertheless, there was neither observation from the tertiary study nor reflection from the survey that a commonly accepted quality assessment system has been formulated for SLRs in SE.

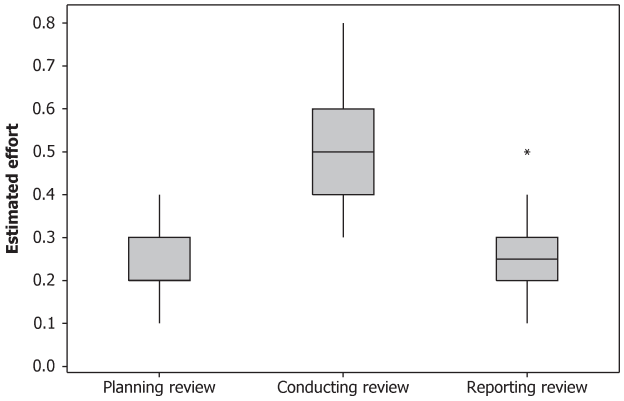

Fig. 10. Effort consumed over SLR phases.

6.2. Effort

In estimating the effort (time) for performing an SLR, Allen and Olkin [2] present a regression formula for determining the number of hours as a function of the number of references returned (x),

$$\text{hours} = 721 + 0.243x - 0.0000123x^2 \quad (1)$$

As this formula was based on empirical observations on SLRs in medical discipline, its accuracy in SE needs to be further assessed. However, there was no reported guidance for estimating the required effort distribution across the three phases of an SLR in SE.

6.2.1. Estimation by phase

The box-plot in Fig. 10 shows the distribution of the respondents' (*systematic reviewers*) estimation of effort consumed over the three phases of SLR according to their experiences. It indicates that the *conducting* review phase takes around half of the effort of undertaking an SLR. In the other half, the *planning* and *reporting* review phases roughly share an equal quarter of the overall effort.

6.2.2. Time-consuming activity

Among all activities of SLR process, from *protocol development* through *data synthesis*, the most time-consuming activity (through ③) voted by the systematic reviewers was *data extraction* (43% responses) followed by *study selection* (27%).

6.2.3. Rework

Through ①, many interviewees mentioned the iterative rework between study selection and data extraction, especially for novices. It can be interpreted that for some primary studies the final selection decision has to be made when reading the full text in data extraction, which may trigger the refinement of the study selection criteria and further the data extraction form. As a result, some refined or new attributes have to be recollected from the reviewed studies. This issue was further investigated with the survey ③.

Table 6
The most cited systematic reviews in software engineering.

Rank	Ref.	Author(s)	Topic	Venue	Year	Citation
1	[8]	T. Dybå and T. Dingsøyr	Agile software development	IST	2008	339
2	[15]	M. Jørgensen and M. Shepperd	Software cost estimation	TSE	2007	277
3	[31]	D. Sjøberg et al.	Controlled experiments in SE	TSE	2005	248
4	[14]	M. Jørgensen	Expert estimation of development effort	TSE	2004	218
5	[16]	H. Kagdi et al.	Mining software evolution repository	JSME	2007	125
6	[22]	B. Kitchenham et al.	Cross- vs. within-company cost estimation	TSE	2007	115
7	[19]	B. Kitchenham et al.	Tertiary study in software engineering	IST	2009	101

Only 18% systematic reviewers in SE never experienced rework between these two activities.

6.3. Teamwork

To enhance its rigour as well as minimize the potential bias, an SLR is normally undertaken by more than one reviewers. Hence, setting up a review team is always a necessary step before starting an SLR. We explored this aspect in this study.

6.3.1. Team size

From the surveys ③, more than half of our respondents (51%) believe that an ideal SLR team should consist of three members. Equally 18% of the respondents regarded either two or four members can form a desirable team. The reason can be explained by the findings from the interviews ① and ④ that a '*too small*' team size (e.g., single reviewer) is difficult to control the potential bias; a '*too large*' size, on the other hand, may lead to a much higher communication and coordination overhead.

6.3.2. Team distribution

In the interviews ① and ④, our interviewees addressed their procedure of performing SLR in either local (within organisation) or distributed (cross organisations) settings. A local team may make communications '*simple*' and '*convenient*', while a distributed team could ensure the '*expertise*' and reinforce '*independence*' of the individual's reviews.

We further examined the systematic reviewers' preference in a larger pool (the surveys ③). The results ⑦ show that many people (36%) are inclined to a local team setting. However, a large number of our respondents (47%) would like to make decision between the two settings based on the tradeoff (of the above mentioned benefits) in the real context.

6.3.3. Roles in SLR

In response to our survey ③, most of the respondents (86%) ever acted '*principle reviewer*' in their previous SLRs, who took the majority of activities from planning to reporting. Other common roles in SLRs include '*secondary reviewer*' (54%), '*internal checker*' (38%), and '*external checker or panel*' (15%). The *checkers or panel* who possess domain expertise specific to the review topic are responsible to review the artifacts and outcomes during review process, e.g., review protocol, selected studies, and extracted data. Our respondents also nominated a few other roles, such as *project leader or administrator* who makes final decision on conflicts and options to '*avoid endless discussions*'; *statistical analyst* who provides help with data analysis and synthesis; and *research methodologist* who '*consults on SLR topics with reviewers*' (e.g., PhD supervisor).

7. Interpretation and discussion

We have adopted a mixed-methods approach to gain an in-depth understanding of different aspects of the status of adoption, perceived and real advantages, and practice of SLRs in SE since its introduction in 2004. We expected the use of mixed-methods approach to provide us with complementary research strategies to produce robust findings by systematically gathering and analysing the required data [34]. We assert that a mixed-perspective overview of adoption and practice of SLRs in SE over the past 7 years can help SE researchers and practitioners to understand the perceived value, current or potential impact, and practices of a research methodology that has been reported to have significant influence on practitioners' and policy makers' decisions in several other disciplines.

7.1. Value and motivations

The proponents of SLRs in SE claim that SLR is a systematic approach to building a body of knowledge about a particular topic or research question(s) by its practitioners, and to identifying problems for future research and support decision making and technology selection. Our findings reveal that many *traditional reviewers* also agree with various value of SLRs claimed by the *advocates* and confirmed by *followers* of SLRs in SE, and are also willing to try SLR in their future literature reviews.

Most of the *systematic reviewers* believe that SLRs provide the mechanism of achieving reliable and traceable findings that can help practitioners to make more informed decisions. Our research has also revealed that SLR users expect to identify new issues in the studied area and innovative ideas. This can be considered the most important motivator for performing SLRs. The participants of our research realise that there are significant differences between medical and software engineering disciplines that is why it is difficult to expect of EBSE to have similar results as evidence-based medicine has produced in the near future. There are some very obvious reasons for this sort of perception, for example the quality of the outcomes of systematic reviews depends upon the quality of the primary studies. It is being widely recognised that a large number of papers on any topic in software engineering is of relatively low quality; there are hardly any standards and/or protocols that are widely acceptable and/or used by a particular community. That means the design, analysis, and reporting of studies on same and/or similar topics may be carried out in different fashions. Moreover, the research questions in medicine are usually quite narrowed and the studied variables can concretely identified and studied. Furthermore, software engineering researchers interested in carrying out systematic literature reviews do not have any concrete discipline specific guidelines for assessing the quality of the primary studies included in a review. Rather this activity is completely ignored or carried out on an ad hoc basis. These kinds of factors make the identification and synthesis of the evidence quite difficult, and in some circumstances impossible. Moreover, software engineering community does not have commonly used terminology for keywords related to a particular topic and the abstracts are usually poorly written. These kinds of factors impact the search, evaluation, and selection process of the relevant papers in the first instance. While there are several efforts geared towards addressing these kinds of factors (e.g., [38]) and SLRs have played a significant role in highlighting the above-mentioned problems, it should be completely understandable that SLRs in SE will not achieve the results that can be considered similar to the results achieved in medicine discipline in the near future.

It is a positive surprise that a majority of the participants were quite optimistic about the effectiveness of SLRs in SE as in other disciplines, such as sociology and education. However, we assert that even the achievement of effectiveness of SLRs in SE similar to these disciplines may not be possible if some of the above-mentioned factors are not satisfactorily addressed; for example, having SE authors to use relevant keywords which are widely used and providing guidelines for defining, assessing, and using quality criteria for assessing the quality of the primary studies. Another important factor is making search engines supportive of SLRs.

We also found the reasons for researchers choosing TLRs over SLRs. It is an interesting finding that half of the respondents were not aware of SLRs when they carried out their literature review. We were not surprised on this finding as systematic review is a relevantly new research methodology in SE which is still going through its dissemination and adoption phases. It is difficult to expect a new research methodology in a discipline to gain widespread awareness in such a short time. However, we have been observing that the general awareness about SLRs in SE is increasing

with several conferences and journals having allocated special tracks/sections, and themes on this topic. Compared to *systematic reviewers*, we received a relatively small number of responses to our survey invitation from the *traditional reviewers*. Apart from the reason that lack of awareness about SLRs among traditional reviewers may not have motivated a large number of them to particulate in our study, this can also be considered an indication that *systematic reviewers* are very enthusiastic about this particular research methodology and are willing to share their perspectives and experiences in order to help improve the methodology.

7.2. Reviews and impact

While it is becoming increasingly accepted that more and more researchers are conducting and reporting SLRs since their introduction in SE, our research has found some interesting numbers, for example, there have been 160 peer-reviewed publications based on almost 148 SLRs. We have not included the grey literatures published based on SLRs. These many publications based on a particular research methodology in SE in a short period is quite impressive and if this trend continues, then SLRs (*conventional systematic reviews* and *systematic mapping studies*) may become one of the most used research methodologies in SE compared with other commonly used research methods such as controlled experiments, surveys, and action research. That means there will be not only an acute need but also huge justification for allocating significant amount of resources for developing the supportive infrastructure for appropriate use and reporting of this research methodology as other researchers have indicated [30].

We have also observed an interesting trend that is an increased shift in favour of *systematic mapping studies* (SMSs) compared with *conventional systematic reviews* (CSRs). SMSs are aimed at providing a broad map of the area of research being investigated. Hence, this kind of studies are not meant to rigourously assess the quality of the primary studies; nor are they supposed to emphasise evidence synthesis. SMSs can also help researchers to decide whether or not the resources required for a systematic review will be justified as if there are low number of primary studies available on a particular topic and a cursory inspection reveals some apparent problems in the quality of even that small number of studies, it may not be justified to go ahead and kick off an SLR which may end up with only a dozen of studies with questionable quality.

The reports based on SLRs tend to be quite long as there are quite a lot of details about different aspects of the methodology used and the results found. Hence, it is quite natural that such studies are meant to be published in archival nature of venues which can allocate the required space such as journals and technical reports. Our findings have revealed that there are almost equal number of journal and conference papers published on SLRs. However, the SLRs published in conferences may not be considered of high quality as a reader is unlikely to find the information that is required to assess the quality of the methodological use and the reliability of the results reported in the conference based SLRs. Moreover, the space in a conference paper may not allow the authors to report the full list of the papers included in the reported study. One suggestion around this shortcoming is to provide this kinds of information online. However, we strongly suggest that a comprehensive report on an SLR should be published in peer-reviewed journals and authors should avoid having incremental publications (which is commonly acceptable practice) in case of SLRs unless there are some solid reasons for publishing an SLR in a conference. And there is a strong incentive for researchers to target high quality journals for reporting SLRs as it is common observation that the papers published in high quality journals are likely to have much more citations than papers published in other venues; and this study has also provided an evidence that the SLRs published in journals gets more citations than the ones which are published in conferences.

This research has also identified more than 30 topics on which SLRs have been conducted and reported. This indicates that the methodology is being widely used as far as the research topics are concerned. However, the distribution of SLRs over these topics was not yet even. A large majority of the reported SLRs have reported on a small number of topics such as *cost estimation*, *requirements engineering* and *empirical methods*, particularly *global development* in recent years. One apparent reason for this concentration can be the nature of research methods used in the reported studies on these topics. Many of the studies reported on these topics tend to use one or more empirical research methods. That means the researchers working on these topics are usually familiar with different kinds of empirical methods and appreciate their value. Moreover, SLRs can be an important and readily available source of secondary data when the sources for primary data are not available in specific context. This finding also highlight the importance of widening the range of topics on which SLRs are being conducted as there are huge amount of peer-reviewed literature published on other topics in software engineering, for example, software testing, program analysis, formal methods, and architectural description languages. However, SLRs on these kinds of topics may require close collaboration between subject matter specialists and methodological experts as neither of them may be fully knowledgeable and skilled in designing and conducting high quality reviews. Moreover, the participants of our study have also identified a few topics on which they would like to see more SLRs conducted and reported. These topics include requirements engineering, software process improvement, and agile development. These topics can provide motivation to researchers working in these areas for conducting more and high quality reviews in future.

7.3. Practice and experiences

Our research shows SLRs, compared to the TLRs reported in the same period, have made a wider research influence in terms of citations. Whereas, systematic review, as a new research methodology, has not been widely adopted by the researchers outside Europe.

An SLR requires more rigour and effort. The rigour is based on its teamwork, bias control and systematically defined research process. With respect to their experiences, the users of SLR suggested an ideal team size of three, and *designing protocol* as the most important activity to ensure rigour and repeatability of SLR. However, no commonly accepted quality assessment system for SLRs in SE was observed or reflected through our research. Based on the statistics, *conducting review* is the most effort-consuming phase in SLR, especially the *data extraction* activity. The effort distribution reflected by the experienced *systematic reviewers* may help the future reviewers estimate and plan their own SLRs.

7.4. Limitations

This research has some limitations that we consider worth mentioning. We have used a mixed-methods approach which is expected to help produce more robust results which are based on the combination of complementary empirical research methods for data triangulation. However, we need to be aware of the limitations of the individual research methods, e.g., interview and questionnaire-based survey. Our study has explored the perceptions and views of SE researchers about their experiences of applying SLRs in SE research through semi-structured interviews. That means our results are based on the recollection of the interviewees. We tried to minimise this risk by audio-taping all the interviews with the interviewees' permission. The transcriptions of the

interviews were verified with the notes taken. Moreover, we tried to have both authors present in most of the interviews.

A shortcoming of the questionnaire-based survey sessions is that respondents are provided with a list of potential reasons and expected benefits of a particular technology (i.e. SLR methodology in our case) and are asked to select from that list. This approach may limit the respondents to consider only those options and questions provided to them. However, we tried to address this issue and provided the respondents an opportunity to share their perceptions and experiences by seeking explanation for most of the responses with open-ended space. We gathered significant amount of qualitative data. A relatively small number of responses to our survey invitation from traditional reviewers can be another limitation of this study.

Considering the wide distribution of traditional literature reviews in SE publication venues, we only performed an automated search to capturing TLRs. This might result in an incomplete set of the identified TLRs compared to the outputs of our '*exhaustive*' search of SLRs based on the QGS approach [38]. However, the purpose of searching TLRs in this research was for identification of representative traditional reviewers and for impact analysis in terms of *average* citations, rather than to produce a complete list of TLRs. Hence, we have confidence in the validity of our findings about the impact that was based on a large sample set of TLRs retrieved from automated search.

Generalizability can be another risk. However, we tried to manage this by involving the participants from different organisations and located in different parts of the world. It should also be noted that a large majority of the participants (i.e. interviewees and survey respondents) reported similar experiences and lessons. It increases our confidence in the findings of this study. One of the main limitations of the mixed-methods approach is the huge amount of time and effort required for planning, executing, and analysing each phase of our research. Nevertheless, we believe that such investment is necessary in order to achieve reliable and robust results.

8. Conclusions and future work

Our long-term research goal is to build an empirically supported body of knowledge to improve the adoption and use of SLRs in SE with the objective of supporting practitioners' decision making for selecting software technologies. Furthermore, we also intend to contribute to the development of scientific and technological support to exploit the full potential of SLRs. We are approaching these goals by firstly concentrating on gaining an in-depth understanding of researchers' perspectives about and motivations for conducting (or not conducting) SLRs, capturing the current status of the adoption and impact of SLRs in SE research, and also studying the practices being followed for conducting SLRs.

This research has gathered empirical evidence to advance the knowledge about different methodological aspects and logistics involved in *planning*, *conducting*, and *reporting* SLRs in SE. The findings provide support to several advantages and strengths of SLRs claimed by the advocates of this methodology based on the perceptions of the users of systematic review as well as of the users of traditional (ad hoc) review. The results also provide SE researchers and practitioners an evidence-based understanding of different aspects and potential value of SLRs for decision making.

Systematic review is a relatively new research methodology for SE researchers. The appropriateness and application of SLR in SE have yet to be fully explored and assessed. However, there is clear evidence that more and more researchers support the move to perform secondary studies in a systematic and rigourous manner as even *traditional reviewers* are of the view that systematic reviews are valuable to SE research and practice. Researchers can also gain motivation for continuously improving the methodology and reporting rigour of their studies in order to deliver high quality SLRs. This study has also identified a few best practices that are expected to be useful for researchers intending to undertake systematic reviews. Additionally, the findings that published SLRs appear to be more influential than TLRs should provide satisfaction to those who have invested significant amount of time and effort in conducting SLRs and motivate those who have been contemplating to apply this methodology. We also expect that our initiative to study different aspects of research methodology will stimulate researchers to carry out similar studies for increasing the understanding of the value and the adoption of systematic review in SE.

Acknowledgments

We would like to record our thanks to all those who have participated in this research and shared their perspectives and experiences. We also sincerely thank the editors and reviewers of this article for their pertinent comments and suggestions that help us improve the quality of this report.

References

[1] M. Ali Babar, H. Zhang, Systematic literature reviews in software engineering: preliminary results from interviews with researchers, in: 3rd International Symposium on Empirical Software Engineering and Measurement (ESEM'09), IEEE, Lake Buena Vista, FL, October 2009, pp. 346–355.
[2] I.E. Allen, I. Olkin, Estimating time to conduct a meta-analysis from number of citations received, Journal of the American Medical Association 282 (7) (1999) 634–635.
[3] J. Biolchini, P.G. Mian, A.C.C. Natali, 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, M. Khalil, Lessons from applying the systematic literature review process within the software engineering domain, Journal of Systems and Software 80 (1) (2007) 571–583.
[5] J. Creswell, Research Design: Qualitative, Quantitative and Mixed Methods Approaches, third ed., Sage Publications, 2009.
[6] D.S. Cruzes, T. Dybå, Research synthesis in software engineering: a tertiary study, Information and Software Technology 53 (5) (2011) 440–455.
[7] F.Q. da Silva, A.L. Santos, S. Soares, A.C.C. França, C.V. Monteiro, F.F. Maciel, Six years of systematic literature reviews in software engineering: an updated tertiary study, Information and Software Technology 53 (9) (2011) 899–913.
[8] T. Dybå, T. Dingsyr, Empirical studies of agile software development: a systematic review, Information and Software Technology 50 (9–10) (2008) 833–859.
[9] T. Dybå, T. Dingsyr, Strength of evidence in systematic reviews in software engineering, in: 2nd ACM-IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM'08), Kaiserslautern, ACM Press, Germany, October 2008, pp. 178–187.
[10] T. Dybå, T. Dingsyr, G.K. Hanssen, Applying systematic reviews to diverse study types: an experience report, in: 1st International Symposium on Empirical Software Engineering and Measurement (ESEM'07), IEEE, Madrid, Spain, September 2007, pp. 225–234.
[11] T. Dybå, B. Kitchenham, M. Jrgensen, Evidence-based software engineering for practitioners, IEEE Software 22 (1) (2005) 158–165.
[12] B. Haugset, G. Hanssen, Automated acceptance testing: a literature review and an industrial case study, in: Agile 2008 Conference (AGILE'08), IEEE, Toronto, Canada, August 2008, pp. 27–38.
[13] J.P. Higgins, S. Green, The Cochrane Collaboration, Cochrane Handbook for Systematic Reviews of Interventions (version 5.1), Technical Report, 2011.
[14] M. Jrgensen, A review of studies on expert estimation of software development effort, Journal of Systems and Software 70 (1–2) (2004) 37–60.
[15] M. Jrgensen, M. Shepperd, A systematic review of software development cost estimation studies, IEEE Transactions on Software Engineering 33 (1) (2007) 33–53.
[16] H. Kagdi, M. Collard, J. Maletic, A survey and taxonomy of approaches for mining software repositories in the context of software evolution, Journal of Software Maintenance and Evolution: Research and Practice 19 (2) (2007) 77–131.
[17] K.S. Khan, G. ter Riet, J. Glanville, A.J. Sowden, J. Kleijnen, Undertaking Systematic Review of Research on Effectiveness, CRD's Guidance for those Carrying out or Commissioning Reviews, Technical Report, NHS Centre for Reviews and Dissemination, University of York, 2001.
[18] B. Kitchenham, Procedures for Undertaking Systematic Reviews, Technical Report, Keele University and National ICT Australia, 2004.

[19] B. Kitchenham, O.P. Brereton, D. Budgen, M. Turner, J. Bailey, S. Linkman, Systematic literature reviews in software engineering: a systematic literature review, Information and Software Technology 51 (1) (2009) 7–15.
[20] B. Kitchenham, S. Charters, Guidelines for Performing Systematic Literature Reviews in Software Engineering (version 2.3), Technical Report, Keele University and University of Durham, 2007.
[21] B. Kitchenham, T. Dybå, M. Jrgensen, Evidence-based software engineering, in: 26th International Conference on Software Engineering (ICSE'04), IEEE, Edinburgh, Scotland, UK, May 2004, pp. 273–284.
[22] B. Kitchenham, E. Mendes, G. Travassos, Cross versus within-company cost estimation studies: a systematic review, IEEE Transactions on Software Engineering 33 (5) (2007) 316–329.
[23] B. Kitchenham, R. Pretorius, D. Budgen, O.P. Brereton, M. Turner, M. Niazi, S. Linkman, Systematic literature reviews in software engineering – a tertiary study, Information and Software Technology 52 (8) (2010) 792–805.
[24] B.A. Kitchenham, S.L. Pfleeger, Principles of Survey Research: Part 2 6. ACM SIGSOFT Software Engineering, Notes, 2002–2003, pp. 27–28.
[25] T. Lethbridge, S. Sim, J. Singer, Studying software engineers: data collection techniques for software field studies, Empirical Software Engineering 10 (3) (2005) 311–341.
[26] J.H. Littell, J. Corcoran, V. Pillai, Systematic Reviews and Meta-Analysis, Pocket Guides to Social Work Research Methods, Oxford University Press, 2008.
[27] M.Q. Patton, Qualitative Research & Evaluation Methods, third ed., Sage Publications, 2002.
[28] M. Petticrew, H. Roberts, Systematic Reviews in the Social Sciences: A Practical Guide, Wiley Blackwell, 2006.
[29] G. Rugg, M. Petre, The Unwritten Rules of PhD Research, Open University Press, 2004.
[30] D.I. Sjøberg, T. Dyba, M. Jorgensen, The future of empirical methods in software engineering research, in: International Conference on Software Engineering, Future of Software Engineering Track, IEEE, 2007.
[31] D.I. Sjøberg, J.E. Hannay, O. Hansen, V.B. Kampenes, A. Karahasanovic, N.-K. Liborg, A.C. Rekdal, A survey of controlled experiments in software engineering, IEEE Transactions on Software Engineering 31 (9) (2005) 733–753.
[32] R.J. Torraco, Writing integrative literature reviews: guidelines and examples, Human Resources Development Review 4 (3) (1995) 356–367.
[33] M. Walsh, Research Made Real – A Guide for Students, Nelson Thornes, 2001.
[34] M. Wood, J. Daly, J. Miller, M. Roper, Multi-method research: an empirical investigation of object-oriented technology, Journal of Systems and Software 48 (1) (1999) 13–26.
[35] C. Zannier, G. Melnik, F. Maurer, On the success of empirical studies in the international conference on software engineering, in: 28th International Conference on Software Engineering (ICSE'06), ACM, Shanghai, China, May 2006, pp. 341–350.
[36] H. Zhang, M. Ali Babar, An empirical investigation of systematic reviews in software engineering, in: 5th International Symposium on Empirical Software Engineering and Measurement (ESEM'11), IEEE, Banff, Canada, September 2009, pp. 87–96.
[37] H. Zhang, M. Ali Babar, On searching relevant studies in software engineering, in: 14th International Conference on Evaluation and Assessment in Software Engineering (EASE'10), BCS, Keele, England, April 2010.
[38] H. Zhang, M. Ali Babar, P. Tell, Identifying relevant studies in software engineering, Information and Software Technology 53 (6) (2011) 625–637.