On Using Grey Literature and Google Scholar in Systematic Literature Reviews in Software Engineering

AFFAN YASIN [1,2], **RUBIA FATIMA** [1], **LIJIE WEN** [1], **WASIF AFZAL** [3], **MUHAMMAD AZHAR** [4], **AND RICHARD TORKAR** [5]

[1]School of Software, Tsinghua University, Beijing 100084, China
[2]School of Computing, Blekinge Institute of Technology, 371 41 Karlskrona, Sweden
[3]School of Information, Design and Innovation, Mälardalen University, 722 20 Västerås, Sweden
[4]College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China
[5]Software Engineering Division, Gothunberg University, 405 30 Gothenburg, Sweden

Corresponding author: Wasif Afzal (wasif.afzal@mdh.se)

ABSTRACT **Context:** The inclusion of grey literature (GL) is important to remove publication bias while gathering available evidence regarding a certain topic. The number of systematic literature reviews (SLRs) in Software Engineering (SE) is increasing but we do not know about the extent of GL usage in these SLRs. Moreover, Google Scholar is rapidly becoming a search engine of choice for many researchers but the extent to which it can find the primary studies is not known. **Objective:** This tertiary study is an attempt to i) measure the usage of GL in SLRs in SE. Furthermore this study proposes strategies for categorizing GL and a quality checklist to use for GL in future SLRs; ii) explore if it is feasible to use only Google Scholar for finding scholarly articles for academic research. **Method:** We have conducted a systematic mapping study to measure the extent of GL usage in SE SLRs as well as to measure the feasibility of finding primary studies using Google Scholar. **Results and conclusions: a) Grey Literature**: 76.09% SLRs (105 out of 138) in SE have included one or more GL studies as primary studies. Among total primary studies across all SLRs (6307), 582 are classified as GL, making the frequency of GL citing as 9.23%. The intensity of GL use indicate that each SLR contains 5 primary studies on average (total intensity of GL use being 5.54). The ranking of GL tells us that conference papers are the most used form 43.3% followed by technical reports 28.52%. Universities, research institutes, labs and scientific societies together make up 67.7% of GL used, indicating that these are useful sources for searching GL. We additionally propose strategies for categorizing GL and criteria for evaluating GL quality, which can become a basis for more detailed guidelines for including GL in future SLRs. **b) Google Scholar Results**: The results show that Google Scholar was able to retrieve 96% of primary studies of these SLRs. Most of the primary studies that were not found using Google Scholar were from grey sources.

INDEX TERMS Grey literature, Google scholar, software engineering, empirical evaluation, systematic mapping, tertiary study, gray, quality checklist.

I. INTRODUCTION

The Internet has become a vital channel for disseminating and accessing scientific literature for both the academic and industrial research needs. Nowadays, everyone has comprehensive access to scientific literature repositories, which com-

The associate editor coordinating the review of this manuscript and approving it for publication was Resul Das.

prise of both "white" and "grey" literature. The "grey" literature, as opposed to "white" literature, is non-peer reviewed scientific information that is not available using commercial information sources such as IEEE or ACM. A large number of software engineering researchers are undertaking systematic literature reviews (SLRs) to investigate empirical evidence in software engineering. The key reason to include grey literature during information synthesis is to minimize the risk of

any bias in the publication. Using the state of the art non-commercial databases that index information, the researchers can make the rigorous process of searching empirical studies in SLRs easier. This study explains the evidence of grey literature while performing synthesis in Systematic Literature Reviews.

Grey literature (GL) refers to informally published written material, not indexed by major database vendors (such as IEEE Xplore[1] and ACM[2] digital libraries). GL is usually attributed to government, academia, pressure groups, trade unions, industries and is not rigorously peer reviewed [1]. Some examples of GL are reports (progress, market research), theses, conference proceedings, technical specifications and standards, official documents, company white papers, discussion boards and blogs.

Typically at the start of any research endeavor, the first-hand information about a new topic is generally collected through GL. This includes a quick search of the topic on Internet and discussions with peers [2]. GL can offer some advantages, e.g., it can be authored by scholars and scientists and thus is of high quality and detail [1]. It has recent information about a topic of interest and is focused [3]. It is also available earlier than commercially published literature [2].

The growth of Internet has immensely broadened the access to GL [4], [5]. However it has also produced new challenges for researchers: What to include and what not to include in GL? A recent example of such a challenge was faced by a journal article where researchers claimed to identify genes that can predict human longevity with 77% accuracy. This received rapid feedback and enough criticism just after an hour of online publication [6]. The online researchers showed their skepticism about the environment and controls in which the study was conducted.

Over the past decade, the Internet has emerged as an essential source of information for everyone [4]. In scientific community, academic researchers are now equipped with state of the art sources of scientific articles and meta-data research tools for their research. The online presence of scientific communities, discussion boards and blogs owned by notable authors is an important source of up-to-date **scientific information**.[3] However, most of the information published in online communities, blogs and discussion boards is considered as "Grey" by the definition of Grey Literature.

The Grey Literature, by Luxembourg definition and **GreyNet community**,[4] is, "Information produced on all levels of government, academics, business and industry in electronic and print formats not controlled by commercial publishing i.e. where publishing is not the primary activity of the producing body". In general, grey literature publications are volatile in nature and lack bibliographic controls such as place and date of publication, details of author and publisher.

These tendencies of grey literature make it difficult to index and categorize it. The grey literature is often referred as "fugitive literature" as it is semi-published and difficult to locate [7], [8]. Grey literature, though not peer-reviewed thoroughly, is still an important source of information [9].

It is worthwhile to note that grey literature, although not peer-reviewed, is often produced by scholars and scientists of their respective fields and is of high quality and detail [9]. According to Soule and Ryan [10], grey literature is becoming a common means for information exchange because it is available on a timely basis than literature published by commercial information sources. For instance, the conference papers are in access to public long before the published articles. Beside these traits, grey literature is focused, has in-depth and up-to-date information about any topic [11]. The growth of Internet has immensely broadened the access to grey literature [4], [12].

Now a days, research on various aspects of grey literature is being undertaken such as one of the recently published studies [13] discusses the argument whether thesis or dissertation are still counted as grey literature (taking in consideration a quality review process for graduation). Furthermore, another group of researchers [14], [15] offer guidelines on how to include online literature/grey literature in research studies, keeping in mind the weaknesses associated with grey literature. Another group of researchers [16], [17] focuses on whether online literature can be used for improving public law or policies. Besides this, research is also being conducted on how online repositories are indexing the grey literature with respect to specific location such as in India [18] and Africa [19]. Our study has multiple objectives and fills the research gap in software engineering by researching (i) the extent of usage of grey literature in systematic literature reviews in software engineering; (ii) categorization strategies and quality assessment criteria for grey literature and (ii) viability of Google Scholar for searching grey literature.

Inclusion of GL is also important to minimize publication bias. Publication bias refers to the problem that the studies with positive results are most likely to be included as primary studies in an SLR than the studies with negative results. Some of the strategies to tackle this issue are to scan for GL, conference proceedings and unpublished results by contacting colleagues and researchers [20]–[22].

With the number of SLRs in SE growing and considering the importance of GL [23], this study investigates the extent of GL use in SE SLRs. As a secondary concern, this study also investigates the extent to which Google Scholar alone is sufficient to find primary studies for an SLR. This tertiary study thus tries to seek answer to the following research questions:

RQ1: What is the extent of usage of GL in SLRs in SE?

RQ1.1: What strategies can be used to categorize GL (non-peer reviewed) and how to assess its quality?

Rationale for RQ1: The Internet is transforming the whole value chain of publishing by offering tools and channels for disseminating and assessing grey literature in the forms

---

[1] http://ieeexplore.ieee.org/Xplore/
[2] http://dl.acm.org/
[3] https://www.nature.com/articles/468867a
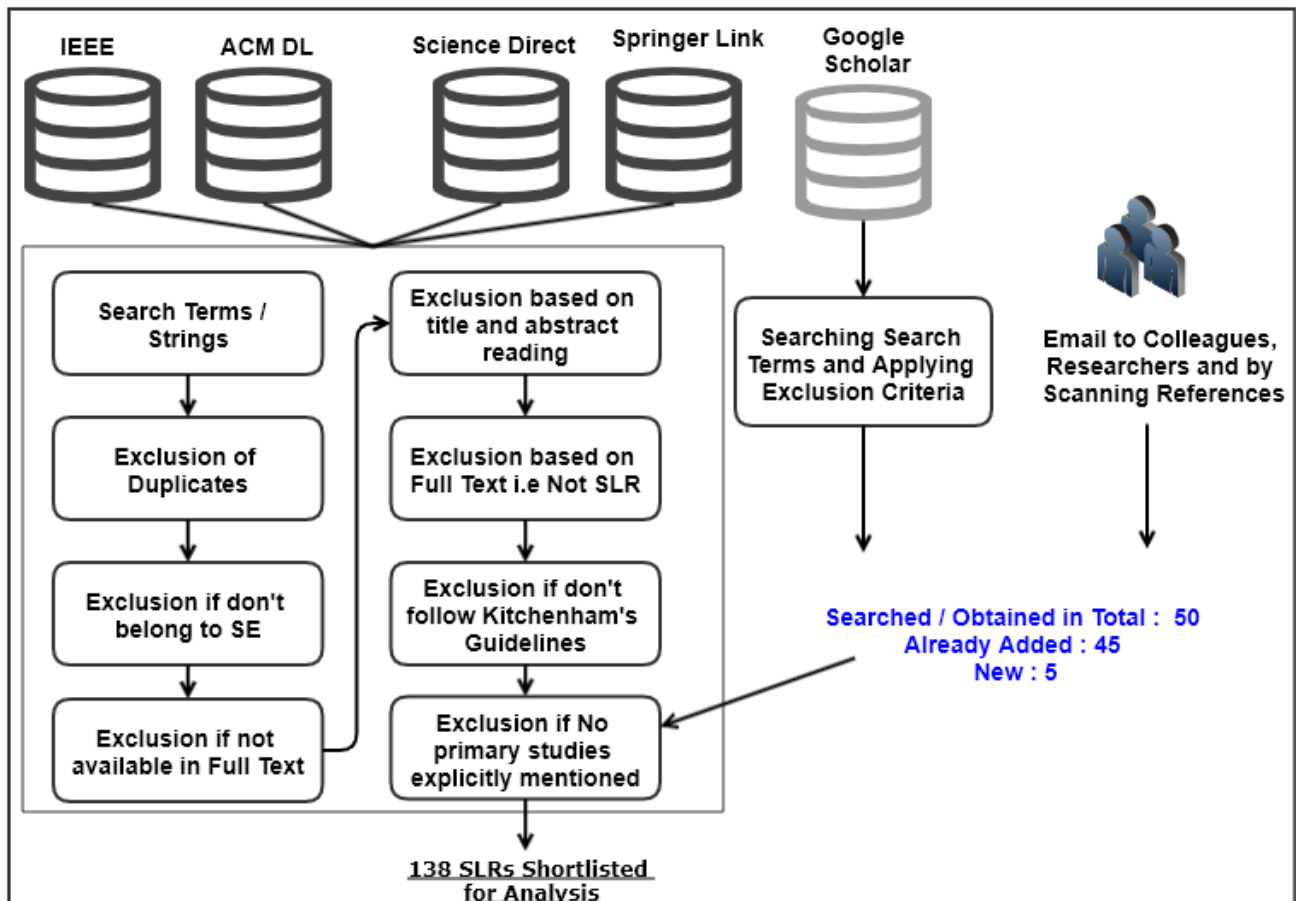[4] http://greynet.org/home.html

FIGURE 1. Research protocol for systematic mapping.

of research blogs, discussion boards and social media. The inclusion of grey literature is inevitable for minimizing publication bias in conducting SLRs in SE. The importance of including grey literature in an SLR demands a study to investigate the evidence of GL being used.

RQ2: Is Google Scholar alone sufficient for searching primary studies in conducting an SLR in SE?

Rationale for RQ2: The process of selecting primary studies for an SLR can be very laborious, time-consuming and rigorous [24]. Manual searches are conducted on different information sources to pile up primary studies. On the other hand, we have Google Scholar [5] [6] that retrieves results from all major databases and orders them on the basis of certain attributes. It is interesting to know if researchers can rely only on Google Scholar for finding primary studies instead of manually searching separately in each of the databases.

The rest of the paper is organized as follows. Section II motivates and explains the research methodologies used, including the important steps in the systematic mapping (sub-section II-A). Section III analyzes and explains the results

acquired from the systematic mapping (sub-section III-A). It thoroughly discusses the characteristics of GL in SLRs including (but not limited to) their forms and origins. Furthermore, sub-section III-B discusses the google scholar indexing results obtained using the systematic mapping. Section IV discusses the proposed categorization strategies and quality evaluation criteria for GL. Threats to the validity of the study are given in sub-section V-A followed by the conclusions in sub-section V-B.

II. RESEARCH METHODOLOGY
A. SYSTEMATIC MAPPING STUDY
This first part of the study is conducted as a systematic mapping study based on the guidelines proposed by Kitchenham [22]. Systematic mapping studies are recommended methods for getting a broad understanding of a research topic and does not involve detailed synthesis as in the case of a systematic literature review (see e.g. [25], [26]). Our methodology is driven by using a predefined protocol that aims to be unbiased by being auditable and repeatable [27]. Our study is also a tertiary study since it collects evidence from secondary studies (i.e. systematic literature reviews in software engineering). Other research methodologies e.g., surveys,

---

[5]https://scholar.google.com/intl/en/scholar/about.html
[6]https://www.nature.com/news/google-scholar-pioneer-on-search-engine-s-future-1.16269

experiments and case studies are not relevant for achieving the goals of this study. Surveys are typically conducted when the use of a technique has already taken place, case studies are mostly suitable for conducting industrial evaluations while experiments are used for quantifying a cause and effect relationship. In our study, we are not assessing any specific technique rather collecting overall evidence of grey literature in SLRs as well as evaluating if Google Scholar alone is able to find primary studies of SLRs.

This systematic mapping study is based on RQ1 and RQ2 given in Section I. The *population* in this study consist of SLRs conducted in SE. *Intervention* includes the use of GL in SLRs within SE. The *comparison* is not applicable in this study as our aim is not to do a comparison. The *outcome* of our interest is the level of usage of GL in SLRs in SE. Our *context* and *types* of primary studies are limited to SLRs.

1) SEARCH STRATEGY

Our search for primary studies (SLRs in this case) was based on the following steps:

- Identification of alternate words and synonyms for terms used in the research question.
- Use of Boolean `OR` to join alternate words and synonyms.
- Use of Boolean `AND` to join major terms.

We limited our search to papers published between year January 2004 to June 2012. We selected 2004 as the starting year because the guidelines for conducting SLRs in SE were first published in 2004. The search terms used are as following: (i) systematic review (ii) systematic literature review (iii) meta-analysis (iv) empirical evidence (v) empirical studies (vi) empirical study. The use of these search terms led to using the following search queries: empirical studies `OR` empirical study, systematic review `AND` Kitchenham, systematic literature review `AND` Kitchenham, meta-analysis `AND` Kitchenham, (empirical studies `OR` empirical study) `AND` Kitchenham, ''systematic review'' `AND` (software engineering).

The following databases were selected for searching[7]:

- ACM Digital Library
- IEEEXplore
- ScienceDirect
- SpringerLink

We conducted a pilot search before the actual search to verify the strength of search terms. This was an attempt to avoid time being wasted because of inadequately designed search terms [21], [24]. After finalizing the pilot studies, we performed search and if we got more than 90% percent pilot studies using a search term, we retained it. The pilot studies included a total of 37 SLRs representing each year from 2004 to 2012. Out of the 37 pilot studies, 22 were found

from Kitchenham et al.'s paper [30] while 15 more were added by contacting prominent authors.

We used a three-phase strategy for searching, similar to one used in [31]. In the first phase, we searched above mentioned electronic databases. In the second phase of our search strategy, we scanned the reference lists of all the papers found after the search in electronic data bases. We then contacted authors who authored most number of SLRs and also scanned their personal webpages. In the third phase of our search strategy, we used Google Scholar[8] to find any missing SLRs. The detail of the research protocol can be seen in the Figure 1.

2) STUDY SELECTION CRITERIA AND PROCEDURES FOR INCLUDING AND EXCLUDING PRIMARY STUDIES

We included papers that met the following inclusion criteria:

- The paper is an SLR, written following the guidelines given in [22].
- The paper is peer-reviewed.
- The paper language is English.
- The paper is published between year January 2004 and June 2012.

We excluded papers based on the following exclusion criteria:

- Paper is not available in full-text.
- Paper does not belong to SE.
- A shorter version of a similar paper is excluded.
- Editorials, position papers, keynotes, tutorial summaries and panel discussions are excluded.
- Reports of lessons learned, expert judgments, anecdotal reports, and observations are excluded.

3) STUDY QUALITY ASSESSMENT AND DATA EXTRACTION

We did not perform quality assessment as a separate step because one of our inclusion criterion enabled us to only include SLRs that followed guidelines proposed in [22]. This meant that the included studies were of reasonable quality and rigor. We designed a data extraction form to collect information needed to answer our research question. We extracted the full citation details of the SLR, number of primary studies used in the SLR and full citation details of every primary study used in the SLR. Most of the SLRs (primary studies in our case) included a list of primary studies while for others we had to read the full-text to get the list. For each primary study in every SLR, the authors searched for the source of the study (whether GL or indexed elsewhere). The SLRs were divided among the authors for data extraction. The data extraction was cross-checked by an author other than the one extracting.

III. DATA ANALYSIS
A. GREY LITERATURE EVIDENCE: SYSTEMATIC MAPPING
A total of 138 SLRs were selected for data synthesis[9]. These SLRs covered four electronic databases (ScienceDirect, IEEE Xplore, ACM digital library, Springer Link). We present our

---

[7]According to Hasteer *et al.* [28] and Dybå *et al.* [29], these databases cover the most relevant journals, conference and workshop proceedings within SE

[8]http://scholar.google.com
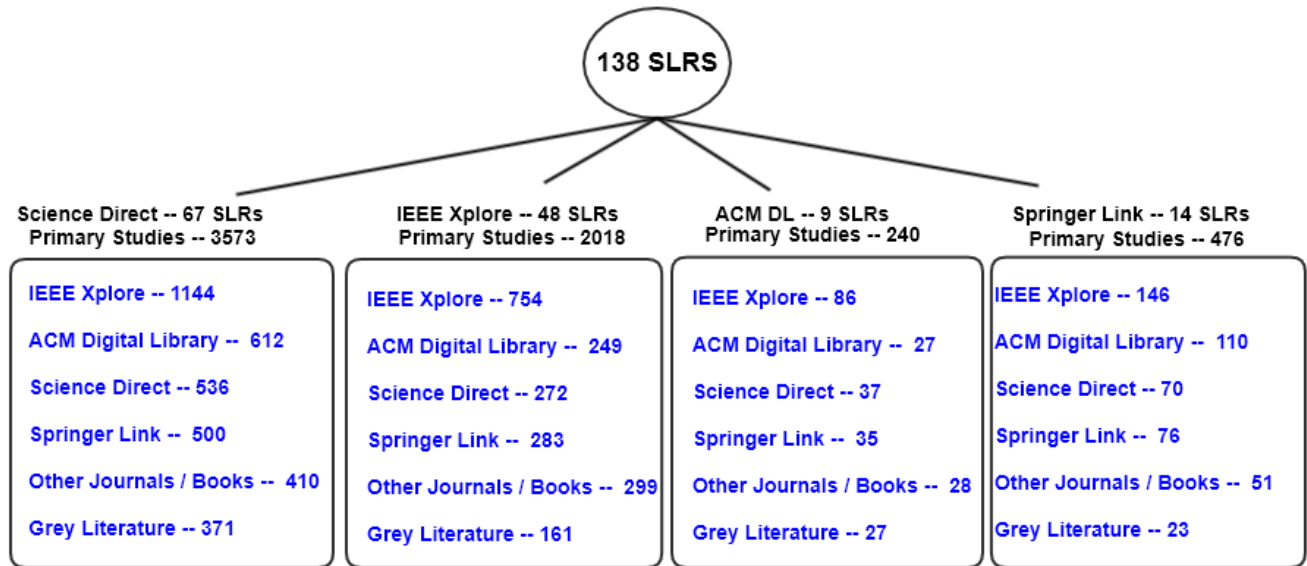[9]The references of primary studies are listed in Appendix.

FIGURE 2. Summary of SLRs and total primary studies with categorization.

TABLE 1. Total SLRs & total primary studies.

Electronic database	Total SLRs	Total primary studies
ScienceDirect	67	3573
IEEE Xplore	48	2018
ACM Dig. Lib.	9	240
Springer Link	14	476
Total	138	6307

TABLE 2. GL evidence in IEEE Xplore.

Primary study source	Total primary studies	Percentage representation %
IEEE Xplore	754	37.36
ACM Dig. Lib.	249	12.34
ScienceDirect	272	13.48
Springer Link	283	14.02
Other journal(s)/books	299	14.82
GL	**161**	**7.98**
Total	2018	100%

TABLE 3. GL evidence in ACM Dig. Lib.

Primary study source	Total primary studies	Percentage representation %
IEEE Xplore	86	35.83
ACM Dig. Lib.	27	11.25
ScienceDirect	37	15.42
Springer Link	35	14.58
Other journal(s)/books	28	11.67
GL	**27**	**11.25**
Total	240	100%

results separately for each database and then, in the end, we will draw the overall picture of grey evidence.

There were a total of 6307 primary studies extracted from 138 SLRs. The total SLRs and the primary studies are given in Table 1 for each database. The detail of the SLRs and primary studies is also shown in Figure 2.

For gathering evidence relating to the use of GL, we classified the total primary studies for every electronic data base according to their source, i.e., whether coming from one of the four electronic data bases (ScienceDirect, IEEE Xplore, ACM digital library, Springer Link), other journals/books or GL.

IEEE SLRs: There were a total of 48 SLRs retrieved from IEEE Xplore, consisting of 2018 primary studies. The classification of these primary studies according to their source is given in Table 2.

ACM SLRs: ACM digital library gave us 9 SLRs consisting of a total of 240 primary studies. Table 3 presents the classification of these 240 primary studies in terms of their source. The number of GL sources stand at 27, making up 11.25% of the total primary studies for SLRs found in ACM digital library.

Science Direct SLRs: For ScienceDirect, the 67 SLRs gathered a total of 3573 primary studies. The classification

of these primary studies according to their source is given in Table 4. The percentage of GL is lowest as compared to other sources of primary studies.

Springer SLRs: There were a total 476 primary studies extracted from 14 SLRs of Springer Link database. 23 primary studies were classified as GL, making up 4.83% of the total number of primary studies (Table 5).

In summary, out of 6307 primary studies in 138 SLRs, 582 (9.23%) were classified as GL. 4920 primary studies (78%) were from the four major databases (ScienceDirect, IEEE Xplore, ACM digital library, Springer Link).

TABLE 4. GL evidence in ScienceDirect.

Primary study source	Total primary studies	Percentage representation (%)
IEEE Xplore	1144	32.02
ACM Dig. Lib.	612	17.13
ScienceDirect	536	15.00
Springer Link	500	13.99
Other journal(s)/books	410	11.47
GL	**371**	**10.38**
Total	3573	100%

TABLE 5. GL evidence in springer link.

Primary study source	Total primary studies	Percentage representation %
IEEE Xplore	146	30.67
ACM Dig. Lib.	110	23.11
ScienceDirect	70	14.71
Springer Link	76	15.97
Other journal(s)/books	51	10.71
GL	**23**	**4.83**
Total	476	100%

TABLE 6. Frequency of GL use.

Primary study source	No of SLRs	SLRs with grey primary studies	Freq. of GL use (%)
IEEE Xplore	48	36	75
ACM Dig. Lib.	9	6	66.67
ScienceDirect	67	55	82.09
Springer Link	14	8	57.14
Total	138	105	76.09

We have noticed that most of the grey literature that has been included as primary studies in SLRs are conference proceedings and technical reports. In order to further analyze the extent of GL use in SLRs, we define certain indicators:

- **Frequency of GL use**: The proportion of SLRs with GL, out of all the SLRs examined.
- **Frequency of GL citing**: The proportion of primary studies as GL, out of all the primary studies examined.
- **Intensity of GL use**: The intensity of GL use is the average number of grey primary studies in SLRs with GL. It is calculated by dividing total grey primary studies by total SLRs with grey primary studies.

1) FREQUENCY OF GL USE
Table 6 shows that 76.09% (105 SLRs) of the total SLRs have used GL for their primary studies. The Table 6 also presents the frequency of GL use in primary studies per database.

2) FREQUENCY OF GL CITING
We see from Table 7 that 582 primary studies were identified as GL out of 6307 primary studies. The Table 7 also presents the frequency of GL citing in primary studies per database.

TABLE 7. Frequency of GL citing.

Primary study source	Total primary studies	Grey primary studies	Freq. of GL citing (%)
IEEE Xplore	2018	161	7.98
ACM Dig. Lib.	240	27	11.25
ScienceDirect	3573	371	10.38
Springer Link	476	23	4.83
Total	6307	582	9.23

TABLE 8. Intensity of GL use.

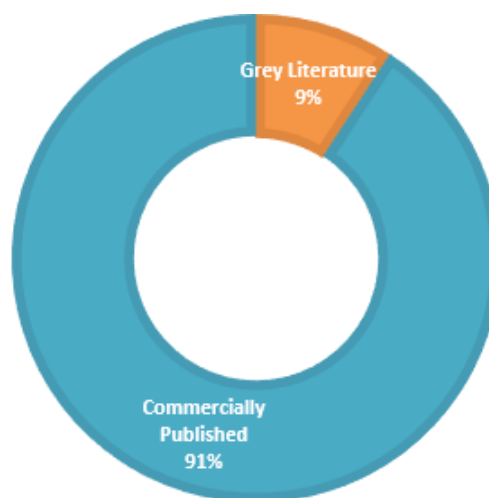| Primary study source	Grey primary studies	Total SLRs with GL	Intensity of GL use
IEEE Xplore	161	36	4.47
ACM Dig. Lib.	27	6	4.5
ScienceDirect	371	55	6.74
Springer Link	23	8	2.88
Total	582	105	5.54

FIGURE 3. Total grey evidence found in software engineering SLRs.

3) INTENSITY OF GL USE
Table 8 shows the intensity of GL use indicator for each database. We see that the intensity of GL use in 105 SLRs is 5.54.

4) TOTAL GREY EVIDENCE FOUND USING SYSTEMATIC MAPPING
A total of 6307 primary studies included in 138 SLRs are investigated. We have found out that 582 primary studies are from grey sources. The percentage of grey evidence is around 9.22% in the selected 138 SLRs of Software Engineering. Figure 3 shows the extent to which grey literature has been used in SLRs in Software Engineering (SE).

5) CHARACTERISTICS OF GL IN SLRs
While the inclusion of GL in synthesizing evidence is important, the GL source should be traceable. During this study, we noticed a small percentage of GL without proper bibliographical control (such as missing date of write-up and

TABLE 9. Ranking of GL documents.

Type of GL	No. of primary studies	Percentage %
Conference papers	252	43.30
Technical Reports	166	28.52
Thesis Dissertation	71	12.20
Workshops/Seminars	44	7.56
Guidelines/Lecture Notes	36	6.18
Preprints	13	2.23
Total	582	100

TABLE 10. Origin of GL documents.

Origin of GL	No. of primary studies	Percentage %
Universities	223	38.32
International organizations	120	20.62
Research institutes/ labs/ scientific societies	171	29.38
Government organizations	21	3.61
Others	47	8.07
Total	582	100

TABLE 11. Publication year of GL documents.

Year	No. of grey primary studies	Percentage %
1970-1990	15	2.58
1991	5	0.86
1992	10	1.72
1993	5	0.86
1994	10	1.72
1995	11	1.89
1996	5	0.86
1997	6	1.03
1998	34	5.84
1999	20	3.44
2000	15	2.58
2001	24	4.12
2002	34	5.84
2003	53	9.11
2004	43	7.39
2005	86	14.78
2006	29	4.98
2007	58	9.96
2008	48	8.25
2009	29	4.98
2010	20	3.44
2011	10	1.72
Missing	12	2.06
Total	582	100

missing company name). We recommend that the GL should have at least the following information: name(s) of authors, date of write-up and name of sponsoring company.

6) FORMS OF GL CITED

The distribution analysis of GL with respect to forms of document is shown in Table 9. The GL is classified into 7 categories: conference papers, technical reports, theses/dissertations, workshop/seminar papers, guidelines/lecture notes and preprints. These categories are described briefly below:

- Conference papers: The conference papers not indexed in the four major databases (ScienceDirect, IEEE Xplore, ACM digital library, Springer Link) are taken as GL.
- Technical reports: Includes reports such as research reports, internal progress and review reports and scientific reports.
- Theses/dissertations: Includes academic theses done at undergraduate and postgraduate levels.
- Workshop/seminar papers: Includes working papers from research groups and committees, typically presented in workshops and seminars.
- Guidelines/lecture notes: Includes company white papers and guides to help readers understand and solve a problem.
- Preprints: Includes draft of a scientific paper that has not yet been published in a peer-reviewed scientific journal.

We see that conference papers are the most cited (43%) GL document type in SLRs followed by technical reports (25.2%) and theses/dissertations (12.4%).

7) ORIGIN OF DOCUMENTS

Table 10 shows the number of grey primary studies by origin type. We classify the origin of grey primary studies as being produced by universities, international organizations, research institutes/labs/scientific societies, government organizations and others. We see that the universities and research institutes/labs/scientific societies are the biggest producers of GL documents covering ∼68% of the total grey primary studies. We also noticed that the grey studies produced by universities, international organizations and research institutes/labs/scientific societies contain well-formed bibliographical details and are highly accessible.

8) DATE OF PUBLICATION

We found 12 (∼2%) grey primary studies that did not provide date of publication. The breakdown of grey primary studies with year of publication is given is Table 11. Majority of grey primary studies included in SLRs can be found in recent past. Almost 48% (280) of included grey primary studies were published in the last 5 years.

B. GOOGLE SCHOLAR INDEXING: SYSTEMATIC MAPPING

ScienceDirect SLRs: A total of 67 SLRs were selected from ScienceDirect database. There were total of 3573 primary studies in all the SLRs. We searched the 3573 primary studies in Google Scholar indexing database. We came up with 3383 studies as hit and 190 studies as miss. Total 94.6% percent of primary studies were found in the Google Scholar. The more granular breakdown of each information source primary studies is tabulated in Table 12.

IEEE SLRs: There were a total of 48 SLRs retrieved from IEEE that consisted of 2018 primary studies. We searched
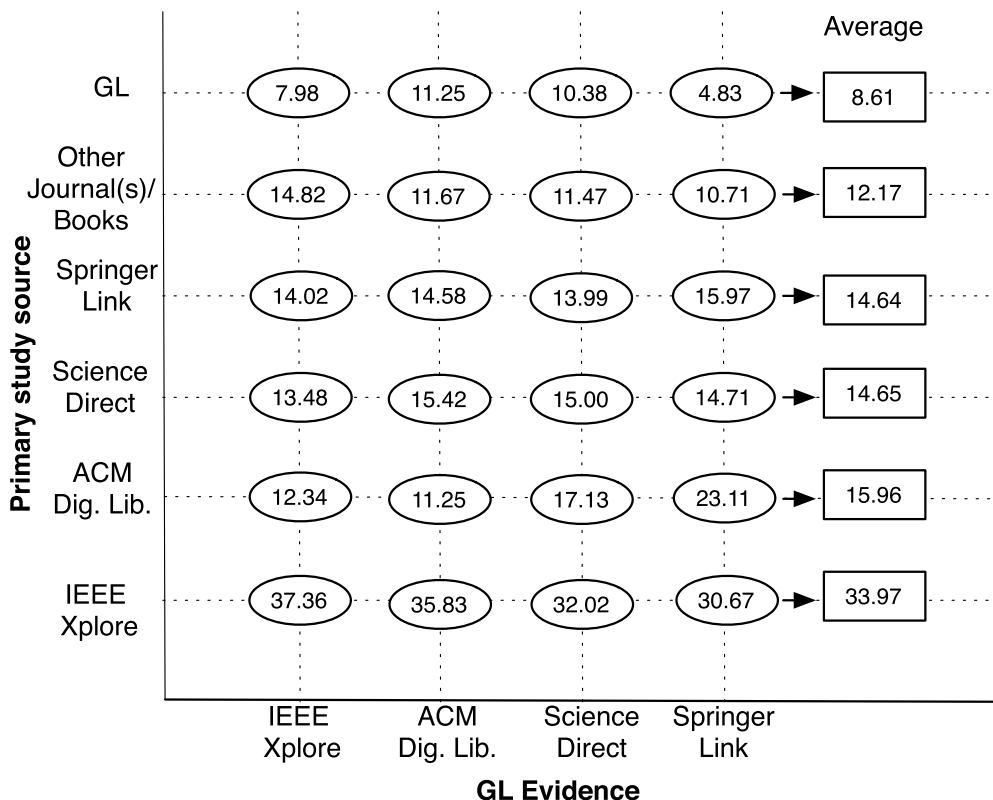
FIGURE 4. GL evidence in the four electronic databases (the numbers in the bubbles are percentages).
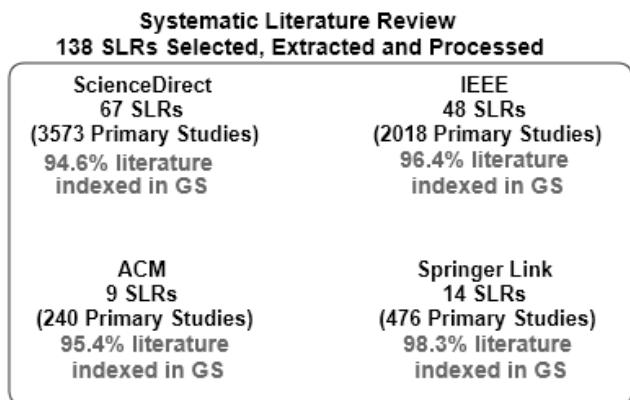
Systematic Literature Review
138 SLRs Selected, Extracted and Processed

ScienceDirect 67 SLRs (3573 Primary Studies) 94.6% literature indexed in GS	IEEE 48 SLRs (2018 Primary Studies) 96.4% literature indexed in GS
ACM 9 SLRs (240 Primary Studies) 95.4% literature indexed in GS	Springer Link 14 SLRs (476 Primary Studies) 98.3% literature indexed in GS

FIGURE 5. Data synthesis google scholar: results.

TABLE 12. Science direct SLRs & google scholar.

Electronic database	Found in Google Scholar	Not found in Google Scholar
IEEE	1129	15
ACM	602	10
ScienceDirect	530	6
Springer Link	491	9
Journals / Books	329	81
Grey Literature	302	69
Total	3383	190

the 2018 primary studies in Google Scholar (GS). A total of 1946 primary studies were found using GS and 72 primary studies were not found. Overall 96% of primary studies were

TABLE 13. IEEE SLRs & google scholar.

Electronic database	Found in Google Scholar	Not found in Google Scholar
IEEE	751	3
ACM	249	0
ScienceDirect	272	0
Springer Link	277	6
Journals / Books	264	35
Grey Literature	133	28
Total	1946	72

found using Google Scholar. The results of Google Scholar findings are tabulated below in Table 13.

ACM SLRs: We retrieved 9 SLRs consisting of total 240 primary studies. There were total 27 grey sources used as primary studies in SLRs selected from ACM database. We searched 240 primary studies on Google Scholar. Out of these 240 primary studies, we were able to found 229 primary studies using Google Scholar. So, overall we were able to find about 95% of total primary studies of ACM SLRs using Google Scholar. The results of Google Scholar finding are shown in Table 14.

Springer Link SLRs: There were a total of 476 primary studies extracted from 14 SLRs of Springer Link database. 23 primary studies were found to be from grey sources. We searched 476 primary studies on Google Scholar. Out of these 476 primary studies, we were able to find 468 primary

TABLE 14. ACM SLRs & google scholar.

Electronic database	Found in Google Scholar	Not found in Google Scholar
IEEE	85	1
ACM	27	0
ScienceDirect	37	0
Springer Link	35	0
Journals / Books	25	3
Grey Literature	20	7
Total	229	11

TABLE 15. Springer link SLRs & google scholar.

Electronic database	Found in Google Scholar	Not found in Google Scholar
IEEE	146	0
ACM	110	0
ScienceDirect	70	0
Springer Link	75	1
Journals / Books	48	3
Grey Literature	19	4
Total	468	8

studies using Google Scholar. So, overall we were able to find about 98% of total primary studies of Springer Link SLRs using Google Scholar. The results of Google Scholar finding are shown in Table 15.

Summary of Google Scholar Results: We searched for the 6307 primary studies in Google Scholar and we came up with 6026 primary studies as hit. Only 281 primary studies were not found using Google scholar. The GS hit percentage is 95.5, which if we round, becomes 96 percent. Going into more detail, we noticed that 281 primary studies that were not found by GS, most of the primary studies were grey sources. Around 38.4% of the primary studies that were not found in Google Scholar were grey literature. We believe that this is because of that fact that grey literature is volatile in nature. Also, this can be because of the fact that sometimes the grey literature is not published in electronic formats or is not published over the web at all.

IV. DISCUSSION

Internet is an obvious choice for searching GL as it attracts a much broader audience [32]. Open access journals are increasing in numbers and are another source for GL. There is an increasing number of data which is generated at informal platforms, such as researchers producing personal opinions, reports and articles over social media, personal websites and blogs. Therefore to utilize this information in a proper manner, we suggest simple strategies to categorize GL based on various attributes. These strategies are a result of our experience and knowledge gained while investigating grey evidence in SLRs in SE.

The strategies presented in this Section have their pros and cons. Therefore a hybrid approach has to be used when searching for GL, e.g., a combination of multiple strategies identified below:

- *Filtering Web Content Based on Page Views:* Page view is the count of views by visitors on a web page. A popular web page is assumed to be viewed by a number of visitors. Once such a count is available, an informed decision can be reached whether to include/exclude a web page. This measure has some obvious limitations. A new web page will not have a higher count while greater number of counts do not correlate with high quality content. Moreover such a count might not be available on every web page.

- *Filtering Web Content Based on User Comments:* For evaluating content in online blogs, discussion boards and bulletins, one can count the number of user comments as an indication of interest a particular post has generated. Again, one cannot entirely judge the importance of content with count of user comments as some comments might only be responses to earlier comments made by others (not relevant to the post).

- *Number of Citations:* If a certain document/report is cited extensively by other authors, it can provide a measure of the importance of such a document/report. A highly cited source may be included while a low cited source may warrant a full-text read to ascertain quality.

- *Filtering GL Based on Type:* There are certain types of GL which are of greater interest than others, such as conference proceedings are more likely to contain important evidence as compared to a company brochure. Similarly literature from certain research labs might be of high quality. Therefore the GL needs to be categorized based on types. One such categorization is based on SE SLRs and is given in Table 9.

- *Filtering GL Based on Authors:* While performing an SLR, it is sometimes obvious that few authors publish more than others. Consequently it might be of interest to look for GL from such authors (scanning their web pages and resources from their research groups).

- *Filtering GL Based on Affiliations:* Our study indicates that 67.7% of GL is contributed by universities, research institutes, labs and scientific societies. This means that it is useful to search for GL in these sources. This step can be performed as a secondary step after filtering GL based on prominent authors.

- *Filtering GL Based on Research Methodology:* Depending on the research question of an SLR, certain research methodologies will be excluded, such as one might only be interested in experimental evidence and thus surveys and case studies will be excluded.

- *Filtering GL Chronologically:* One of the advantages of GL is that new data is available quickly. Therefore sorting GL based on date can lead researchers to capture trends and allow them an insight into innovations. Research gaps can be identified quickly, setting foundations for interesting future research ideas.
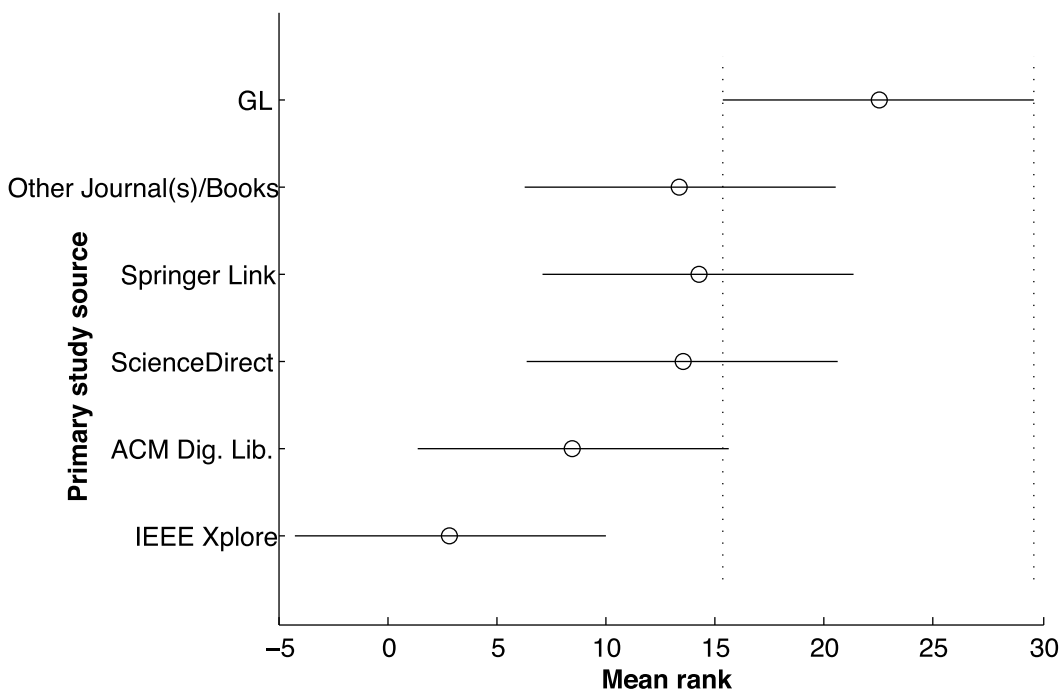
FIGURE 6. Multiple comparisons test for the different sources of primary studies.

All the strategies presented above have their own pros and cons. The recommendation is to use hybrid approach while using these strategies. An example combination of these strategies can be as follows;

1) Search the String/ Keyword.
2) Categorize by grey literature type (Conference Proceedings, Thesis, Reports etc.)
3) Categorize by no. of hits or no. of citations.

There are many different combinations which can be adopted in order to fetch quality data from Internet. It totally depends on the researcher to select a certain combination of strategies which suits his research requirements.

A. ASSESSMENT OF GL QUALITY

While inclusion of GL can help protect us from publication bias, their quality has to be assessed. GL usually do not undergo rigorous peer-review therefore their quality must be assessed against a minimum number of preset criteria. We have come up with a list of quality assessment criteria (a checklist) designed for GL, along with the motivations of including them (Table 16). The criteria are based on our experience of searching GL during this study and are by no means complete. Furthermore we have not yet evaluated the validity of the quality criteria which is planned as a future study.

V. VALIDITY THREATS AND CONCLUSION
A. VALIDITY THREATS
This study is conducted using the guidelines for performing SLRs [22], though on the scale of a systematic mapping study

TABLE 16. Quality assessment criteria.

No.	Question	Motivation
1.	Can the document source be traced?	With enough bibliographical detail, the GL should be accessible.
2.	Is the document produced by an university/research institute, lab/scientific society/international organization?	The origin of GL help ascertain certain level of quality.
3.	Is the document cited by other published papers?	Citations to the document can help reduce uncertainty about quality.
4	Has the document received any user comments?	User comments can point to importance or research contribution.
5	Do majority of comments on the document support its quality?	A high-quality document should receive more positive comments.
6	Have the authors published elsewhere?	Prominent authors in a field are more likely to have published elsewhere.
7	Can the results be reproduced?	To ensure enough methodological details are provided.

as we asked general questions (i.e., what do we know about use of GL in SE SLRs?). The search strategy was initially piloted on a small number of studies to ensure maximum coverage. The search strategy was not only limited to electronic

databases but also included searching for relevant studies in the reference lists of included papers, asking researchers about any SLRs we might have missed and using Google Scholar. A validity threat is that we did not search in electronic databases other than ACM digital library, IEEE Xplore, Science Direct and Springer Link. We intend to add more databases in the future extension of this mapping study in to a detailed SLR. We defined explicit inclusion/exclusion criteria but did not perform quality assessment because we only included SLRs following standard guidelines [22] and also because our research questions were not posed to evaluate research outcomes. Quality assessment will however be required once this mapping study is extended to an SLR where we would be interested in specific research outcomes. The data extraction in our case was lengthy but not complex. On few occasions it was not easy to find primary studies of a particular SLR. In that case, two of the researchers matched their outcomes and resolved differences. The validity of data synthesis was reached by cross-checking, i.e., the data extracted by one researcher was checked for any mistakes by the other researchers. The categorization of GL in case of conference proceedings was tricky since we did not know about the review policy of some of the conferences. We took the assumption that conference proceedings not included in the four major electronic databases are GL. We know that this is not the case with every conference proceeding in SE but this threat was minimized using authors' knowledge in SE research. However in the future SLR we intend to come up with a more detailed mechanism of categorizing conference proceedings as GL.

According to Hasteer et al. [28] and Dybå et al. [29], IEEEXplore, ACM Digital Library, Springer and Elsevier/Science Direct cover the most relevant journals, conferences and workshop proceedings within SE. Nevertheless, we acknowledge that adding more databases *(including Scopus)* will increase the validity of the study.

Grey literature is a new and emerging area in the Software Engineering field [23]. Researchers are exploring and proposing methods to utilize grey literature. Some suggest methods on utilizing quality blogs while others suggest utilizing quality online literature in research work. To the best of our knowledge, no study in SE has tried to calculate the magnitude of this grey evidence. Therefore, we have not included a separate related work section in this study, however some of the important contributions related to grey literature are mentioned earlier in Section I.

B. CONCLUSION

The below subsections will summarize and conclude the results of our study.

1) GREY LITERATURE RESULTS

Despite the known importance of GL during SLRs, we have found out that the level of grey literature evidence is 9%. Thus, most of the literature, which is included as primary studies in SLRs, is published and peer-reviewed. GL has

gained more importance in "Health and Medical Science" research because of the sensitivity of research topics about human health and life. The inclusion of grey trials is necessary to limit any publication bias in Health Science [20]. We have found out that in the field of SE, researchers undertake SLRs with overwhelming use of peer-reviewed articles. In the following section, we state our answers to previously stated research questions.

RQ1: **What is the extent of usage of GL in SLRs in SE?**
RQ1.1: **What strategies can be used to categorize GL (non-peer reviewed) and how to assess its quality?**

After investigation of 6307 primary studies during the systematic mapping, we have found out that the percentage of grey evidence is 9% in our selected SLRs. Among the total 6307 primary studies, 582 studies were classified as grey literature. While analyzing the 582 grey links, we noticed that most of the grey literature consisted of conference proceedings and technical reports (68%). The research results in these reports and proceedings are more detailed and specific than in journals and these results are available months before the official publication in traditional databases.

Our results regarding the evidence of GL in SE SLRs suggest that, on average, there is a minimal level of GL evidence (8.61%), when compared with four major electronic databases (IEEE Xplore, ACM digital library, ScienceDirect, Springer Link) and other journals/books. The comparison of GL with other sources of primary studies for the four major electronic databases is given in Figure 4.

The average percentage of primary studies source for IEEE Xplore, ACM digital library, ScienceDirect, Springer Link and other journal(s)/books is 33.97, 15.96, 14.65, 14.64 and 12.17, respectively. The results of performing a Kruskal-Wallis test to compare samples from each primary study source showed that at least one sample median is different from the others ($p = 0.004$, $\alpha = 0.05$). A multiple comparisons test (Tuckey-Kramer, $\alpha = 0.05$) showed that the primary studies from IEEE Xplore are significantly different from those belonging to GL. No other pairs of primary study sources differed significantly. This is shown in Figure 6 where the vertical dotted lines indicate differences in mean ranks of different sources, i.e., IEEE Xplore and GL have significantly different mean ranks.

We also collected three other measures of GL evidence in primary studies: frequency of GL use, frequency of GL citing and intensity of GL use. These three measures for the four major electronic databases is given in Figure 7.

We see that overall 76.09% SLRs (105 out of 138) in SE have included one or more GL studies as primary studies. Among 6307 primary studies across all SLRs, 582 are classified as GL, making the frequency of GL citing as 9.23% (the average across four databases is 8.61%). The intensity of GL use indicate that each SLR contains 5 primary studies on average (total intensity of GL use being 5.54). The ranking of GL tells us that conference papers are the most used form (43.3%) followed by technical reports (28.52%). Universities, research institutes, labs and scientific societies together make
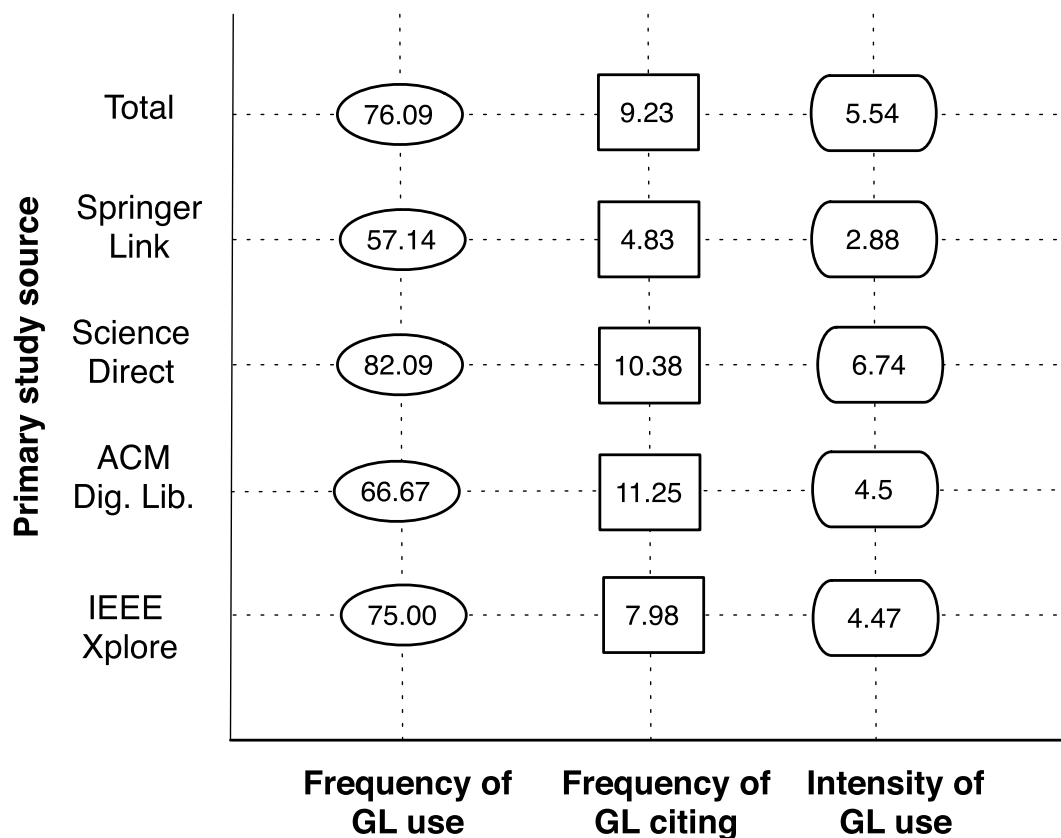
FIGURE 7. Frequency of GL use, frequency of GL citing and intensity of GL use across four databases (the numbers are percentages.

up 67.7% of GL used, indicating that these are useful sources for searching GL.

2) GOOGLE SCHOLAR RESULTS

RQ2: **Is Google Scholar alone sufficient for searching primary studies in conducting an SLR in SE?**

Searching for research literature (especially in Software Engineering) is time-consuming, and this effort increases a lot in case of an SLR. Our study aims to find a solution to this problem by answering the RQ2. A systematic mapping study is performed where in total, 138 SLRs (6307 primary studies) were extracted from various databases and searched in Google Scholar. The results from the analysis of the Google Scholar database showed that Google Scholar was able to retrieve (96%) of primary studies of SLRs. Most of the primary studies that were not found in Google Scholar belonged to grey sources. Moreover, during our research, we have seen that the literature which was not found with Google Scholar was found from simple direct Google search. Thus, it can be argued that the combination of Google Scholar and Google can increase the chances of finding maximum number of primary studies.

When we look at the results of Google Scholar, we see that Google Scholar was able to retrieve (90+%) of primary studies of SLRs. Most of the primary studies that were not found using Google Scholar were of grey sources. We found the primary studies that were not found in Google Scholar

to be heterogeneous in characteristics and therefore we could not infer much about what type of studies generally Google Scholar is not able to retrieve. During our Google Scholar analysis, we noticed that some of the primary studies that were not found in GS were retrievable through Google. There were only few primary studies that were not found in both Google Scholar and Google. All of these primary studies were conference proceedings and workshops. We found that these studies were either published before year 2000 or belonged to specific conference proceedings. So collectively, we were able to find most but not all the primary studies using combination of Google Scholar and Google.

Possible future work for the study is to bridge the gap between academia and the GL utilization process. In this study, we have suggested a preliminary quality evaluation checklist (Table 16), which can further be enhanced and utilized to access the quality of the grey literature.

APPENDIX
PRIMARY STUDIES INCLUDED IN THE SYSTEMATIC MAPPING STUDY
[27], [30], [31], [33]–[167].

REFERENCES
[1] O. Osayande and C. O. Ukpebor, "Grey literature acquisition and management: Challenges in academic libraries in Africa," Library Philosophy Pract. (e-Journal), Univ. Nebraska-Lincoln, Lincoln, Nebraska, Tech. Rep., 2012.

[2] M. H. Soule and R. P. Ryan. *Grey Literature—Technical Briefing*. IT. Accessed: Feb. 21, 2020. [Online]. Available: https://www.yumpu.com/en/document/view/11260560/grey-literature-technical-briefing-by-mason-h-soule-and-ossnet

[3] P. Pejšová, "Czech national repository of grey literature," *GreyNet*, p. 117, 2010.

[4] J. Gelfand and J. King, "Grey market science: Research libraries, grey literature, and the legitimization of scientific discourse in the Internet age," in *Proc. Socioeconomic Dimensions Electron. Workshop*, Nov. 2002, pp. 115–120.

[5] P. De Castro and S. Salinetti, "Quality of grey literature in the open access era: Privilege and responsibility," *Publishing Res. Quart.*, vol. 20, no. 1, pp. 4–12, Mar. 2004.

[6] A. Mandavilli, "Trial by Twitter," *Nature*, vol. 469, no. 1, pp. 286–287, 2011.

[7] K. Denda, "Fugitive literature in the cross hairs: An examination of bibliographic control and access," *Collection Manage.*, vol. 27, no. 2, pp. 75–86, Jun. 2002.

[8] A. F. van Raan, "Bibliometrics and Internet: Some observations and expectations," *Scientometrics*, vol. 50, no. 1, pp. 59–63, 2001.

[9] O. Osayande and C. O. Ukpebor, "Grey literature acquisition and management: Challenges in academic libraries in Africa," *Grey Literature Acquisition Manage., Challenges Acad. Libraries Africa*, 2012, pp. 1–13.

[10] M. H. Soule and R. P. Ryan. *Gray Literature, Technical Briefing*. IT. Accessed: 1999. [Online]. Available: http://www.dtic.mil/summit/tb07.html

[11] M. Vaska, J. Schöpfel, I. Fürstová, R. Polčák, J. Mach, B. Frantíková, P. Karlach, and J. Dolanský, "Grey literature repositories," Radim Bacuvcik, VeRBuM, Tech. Rep., 2010. Accessed: Feb. 21, 2020. [Online]. Available: https://core.ac.uk/download/pdf/35095112.pdf

[12] P. De Castro and S. Salinetti, "Quality of grey literature in the open access era: Privilege and responsibility," *Publishing Res. Quart.*, vol. 20, no. 1, pp. 4–12, Mar. 2004.

[13] J. Schopfel and B. Rasuli, "Are electronic theses and dissertations (still) grey literature in the digital age? A FAIR debate," *Electron. Library*, vol. 36, no. 2, pp. 208–219, Apr. 2018, doi: 10.1108/el-02-2017-0039.

[14] V. Garousi, M. Felderer, and M. V. Mäntylä, "The need for multivocal literature reviews in software engineering: Complementing systematic literature reviews with grey literature," in *Proc. 20th Int. Conf. Eval. Assessment Softw. Eng., (EASE)*, Limerick, Ireland, S. Beecham, B. A. Kitchenham, and S. G. MacDonell, Eds., Jun. 2016, pp. 26:1–26:6, doi: 10.1145/2915970.2916008.

[15] V. Garousi, M. Felderer, and M. V. Mäntylä, "Guidelines for including grey literature and conducting multivocal literature reviews in software engineering," *Inf. Softw. Technol.*, vol. 106, pp. 101–121, Feb. 2019, doi: 10.1016/j.infsof.2018.09.006.

[16] A. Lawrence, "Grey literature publishing in public policy: Production and management, costs and benefits," in *Proc. 21st Int. Conf. Electron., Expanding Perspect. Open Sci., Communities, Cultures Diversity Concepts Practices*, L. Chan and F. Loizides, Eds. Limassol, Cyprus: IOS Press, Jun. 2017, pp. 85–99, doi: 10.3233/978-1-61499-769-6-85.

[17] A. Lawrence, "Influence seekers: The production of grey literature for policy and practice," *Inf. Services Use*, vol. 37, no. 4, pp. 389–403, Jan. 2018, doi: 10.3233/isu-170857.

[18] B. Shivaram and B. Biradar, "Grey literature archiving pattern in open access (OA) repositories with special emphasis on Indian OA repositories," *Electron. Library*, vol. 37, no. 1, pp. 95–107, Feb. 2019, doi: 10.1108/el-05-2018-0100.

[19] P. Sturges, "Using grey literature in informal information services in africa," *J. Document.*, vol. 50, no. 4, pp. 273–290, Apr. 1994, doi: 10.1108/eb026934.

[20] I. Ahmed, A. J. Sutton, and R. D. Riley, "Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: A database survey," *Brit. Med. J.*, vol. 344, no. 1, pp. d7762–d7762, Jan. 2012.

[21] B. A. Kitchenham, P. Brereton, M. Turner, M. Niazi, S. G. Linkman, R. Pretorius, and D. Budgen, "The impact of limited search procedures for systematic literature reviews A participant-observer case study," in *Proc. 3rd Int. Symp. Empirical Softw. Eng. Meas. (ESEM)*. Lake Buena Vista, FL, USA: IEEE Computer Society, Oct. 2009, pp. 336–345, doi: 10.1109/ESEM.2009.5314238.

[22] B. A. Kitchenham, "Guidelines for performing systematic literature reviews in software engineering," School Comput. Sci. Math., Keele Univ., Keele, U.K., Tech. Rep. EBSE-2007-001, 2007.

[23] F. K. Kamei, "The use of grey literature review as evidence for practitioners," *SIGSOFT Softw. Eng. Notes*, vol. 44, no. 3, pp. 23–23, Nov. 2019, doi: 10.1145/3356773.3356797.

[24] K. R. Felizardo, N. Salleh, R. M. Martins, E. Mendes, S. G. Macdonell, and J. C. Maldonado, "Using visual text mining to support the study selection activity in systematic literature reviews," in *Proc. Int. Symp. Empirical Softw. Eng. Meas. (ESEM)*, Sep. 2011, pp. 77–86.

[25] M. M. Hassan, W. Afzal, B. Lindström, S. M. A. Shah, S. F. Andler, and M. Blom, "Testability and software performance: A systematic mapping study," in *Proc. 31st Annu. ACM Symp. Appl. Comput. (SAC)*. New York, NY, USA: ACM, 2016, pp. 1566–1569.

[26] D. Flemstrom, D. Sundmark, and W. Afzal, "Vertical test reuse for embedded systems: A systematic mapping study," in *Proc. 41st Euromicro Conf. Softw. Eng. Adv. Appl.*, Aug. 2015, pp. 317–324.

[27] D. S. Cruzes and T. Dybå, "Research synthesis in software engineering: A tertiary study," *Inf. Softw. Technol.*, vol. 53, no. 5, pp. 440–455, May 2011, doi: 10.1016/j.infsof.2011.01.004.

[28] N. Hasteer, A. Bansal, and B. K. Murthy, "Pragmatic assessment of research intensive areas in cloud: A systematic review," *SIGSOFT Softw. Eng. Notes*, vol. 38, no. 3, p. 1, May 2013, doi: 10.1145/2464526.2464533.

[29] T. Dyba, T. Dingsoyr, and G. K. Hanssen, "Applying systematic reviews to diverse study types: An experience report," in *Proc. 1st Int. Symp. Empirical Softw. Eng. Meas. (ESEM)*. Washington, DC, USA: IEEE Computer Society, Sep. 2007.

[30] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering—A systematic literature review," *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, Jan. 2009, doi: 10.1016/j.infsof.2008.09.009.

[31] W. Afzal, R. Torkar, and R. Feldt, "A systematic review of search-based testing for non-functional system properties," *Inf. Softw. Technol.*, vol. 51, no. 6, pp. 957–976, Jun. 2009, doi: 10.1016/j.infsof.2008.12.005.

[32] T. F. Frandsen, "The effects of open access on un-published documents: A case study of economics working papers," *J. Informetrics*, vol. 3, no. 2, pp. 124–133, Apr. 2009.

[33] S.-J. Huang, N.-H. Chiu, and Y.-J. Liu, "A comparative evaluation on the accuracies of software effort estimates from clustered data," *Inf. Softw. Technol.*, vol. 50, nos. 9–10, pp. 879–888, Aug. 2008, doi: 10.1016/j.infsof.2008.02.005.

[34] K.-J. Stol, M. A. Babar, P. Avgeriou, and B. Fitzgerald, "A comparative study of challenges in integrating open source software and inner source software," *Inf. Softw. Technol.*, vol. 53, no. 12, pp. 1319–1336, Dec. 2011, doi: 10.1016/j.infsof.2011.06.007.

[35] Ø. Hauge, C. Ayala, and R. Conradi, "Adoption of open source software in software-intensive organizations—A systematic literature review," *Inf. Softw. Technol.*, vol. 52, no. 11, pp. 1133–1154, Nov. 2010, doi: 10.1016/j.infsof.2010.05.008.

[36] A. Talaei-Khoei, P. Ray, N. Parameshwaran, and L. Lewis, "A framework for awareness maintenance," *J. Netw. Comput. Appl.*, vol. 35, no. 1, pp. 199–210, Jan. 2012, doi: 10.1016/j.jnca.2011.06.011.

[37] M. Wicks and R. Dewar, "A new research agenda for tool integration," *J. Syst. Softw.*, vol. 80, no. 9, pp. 1569–1585, Sep. 2007, doi: 10.1016/j.jss.2007.03.089.

[38] M. Escalona, J. Gutierrez, M. Mejías, G. Aragón, I. Ramos, J. Torres, and F. Domínguez, "An overview on test generation from functional requirements," *J. Syst. Softw.*, vol. 84, no. 8, pp. 1379–1393, Aug. 2011, doi: 10.1016/j.jss.2011.03.051.

[39] M. Jørgensen, "A review of studies on expert estimation of software development effort," *J. Syst. Softw.*, vol. 70, nos. 1–2, pp. 37–60, 2004, doi: 10.1016/S0164-1212(02)00156-5.

[40] L. M. Karg, M. Grottke, and A. Beckhaus, "A systematic literature review of software quality cost research," *J. Syst. Softw.*, vol. 84, no. 3, pp. 415–427, Mar. 2011, doi: 10.1016/j.jss.2010.11.904.

[41] G. S. Walia and J. C. Carver, "A systematic literature review to identify and classify software requirement errors," *Inf. Softw. Technol.*, vol. 51, no. 7, pp. 1087–1109, Jul. 2009, doi: 10.1016/j.infsof.2009.01.004.

[42] P. A. Da M. S. Neto, I. D. C. Machado, J. D. Mcgregor, E. S. De Almeida, and S. R. De L. Meira, "A systematic mapping study of software product lines testing," *Inf. Softw. Technol.*, vol. 53, no. 5, pp. 407–423, May 2011, doi: 10.1016/j.infsof.2010.12.003.

[43] F. Elberzhager, J. Münch, and V. T. N. Nha, "A systematic mapping study on the combination of static and dynamic quality assurance techniques," *Inf. Softw. Technol.*, vol. 54, no. 1, pp. 1–15, Jan. 2012, doi: 10.1016/j.infsof.2011.06.003.

[44] G. Holl, P. Grünbacher, and R. Rabiser, "A systematic review and an expert survey on capabilities supporting multi product lines," *Inf. Softw. Technol.*, vol. 54, no. 8, pp. 828–852, Aug. 2012, doi: 10.1016/j.infsof.2012.02.002.

[45] M. S. Ali, M. Ali Babar, L. Chen, and K.-J. Stol, "A systematic review of comparative evidence of aspect-oriented programming," *Inf. Softw. Technol.*, vol. 52, no. 9, pp. 871–887, Sep. 2010, doi: 10.1016/j.infsof.2010.05.003.

[46] M. Khurum and T. Gorschek, "A systematic review of domain analysis solutions for product lines," *J. Syst. Softw.*, vol. 82, no. 12, pp. 1982–2003, Dec. 2009, doi: 10.1016/j.jss.2009.06.048.

[47] L. Chen and M. A. Babar, "A systematic review of evaluation of variability management approaches in software product lines," *Inf. Softw. Technol.*, vol. 53, no. 4, pp. 344–362, Apr. 2011, doi: 10.1016/j.infsof.2010.12.006.

[48] M. Höst and A. Orucevic-Alagic, "A systematic review of research on open source software in commercial software product development," *Inf. Softw. Technol.*, vol. 53, no. 6, pp. 616–624, Jun. 2011, doi: 10.1016/j.infsof.2010.12.009.

[49] D. Mellado, C. Blanco, L. E. Sánchez, and E. Fernández-Medina, "A systematic review of security requirements engineering," *Comput. Standards Inter.*, vol. 32, no. 4, pp. 153–165, Jun. 2010, doi: 10.1016/j.csi.2010.01.006.

[50] H. P. Breivold, I. Crnkovic, and M. Larsson, "A systematic review of software architecture evolution research," *Inf. Softw. Technol.*, vol. 54, no. 1, pp. 16–40, Jan. 2012, doi: 10.1016/j.infsof.2011.06.002.

[51] F. J. Lucas, F. Molina, and A. Toval, "A systematic review of UML model consistency management," *Inf. Softw. Technol.*, vol. 51, no. 12, pp. 1631–1645, Dec. 2009, doi: 10.1016/j.infsof.2009.04.009.

[52] E. Engström, P. Runeson, and M. Skoglund, "A systematic review on regression test selection techniques," *Inf. Softw. Technol.*, vol. 52, no. 1, pp. 14–30, Jan. 2010, doi: 10.1016/j.infsof.2009.07.001.

[53] M. Svahnberg, T. Gorschek, R. Feldt, R. Torkar, S. B. Saleem, and M. U. Shafique, "A systematic review on strategic release planning models," *Inf. Softw. Technol.*, vol. 52, no. 3, pp. 237–248, Mar. 2010, doi: 10.1016/j.infsof.2009.11.006.

[54] D. Benavides, S. Segura, and A. Ruiz-Cortés, "Automated analysis of feature models 20 years later: A literature review," *Inf. Syst.*, vol. 35, no. 6, pp. 615–636, Sep. 2010, doi: 10.1016/j.is.2010.01.001.

[55] S. U. Khan, M. Niazi, and R. Ahmad, "Barriers in the selection of offshore software development outsourcing vendors: An exploratory study using a systematic literature review," *Inf. Softw. Technol.*, vol. 53, no. 7, pp. 693–706, Jul. 2011, doi: 10.1016/j.infsof.2010.08.003.

[56] C. Blanco, J. Lasheras, E. Fernández-Medina, R. Valencia-García, and A. Toval, "Basis for an integrated security ontology according to a systematic review of existing proposals," *Comput. Standards Inter.*, vol. 33, no. 4, pp. 372–388, Jun. 2011, doi: 10.1016/j.csi.2010.12.002.

[57] B. J. Williams and J. C. Carver, "Characterizing software architecture changes: A systematic review," *Inf. Softw. Technol.*, vol. 52, no. 1, pp. 31–51, Jan. 2010, doi: 10.1016/j.infsof.2009.07.002.

[58] M. Turner, B. Kitchenham, P. Brereton, S. Charters, and D. Budgen, "Does the technology acceptance model predict actual use? A systematic literature review," *Inf. Softw. Technol.*, vol. 52, no. 5, pp. 463–479, May 2010, doi: 10.1016/j.infsof.2009.11.005.

[59] T. Dybå and T. Dingsøyr, "Empirical studies of agile software development: A systematic review," *Inf. Softw. Technol.*, vol. 50, nos. 9–10, pp. 833–859, Aug. 2008, doi: 10.1016/j.infsof.2008.01.006.

[60] A. S. Jadhav and R. M. Sonar, "Evaluating and selecting software packages: A review," *Inf. Softw. Technol.*, vol. 51, no. 3, pp. 555–563, Mar. 2009, doi: 10.1016/j.infsof.2008.09.003.

[61] M. Shepperd and S. Macdonell, "Evaluating prediction systems in software project estimation," *Inf. Softw. Technol.*, vol. 54, no. 8, pp. 820–827, Aug. 2012, doi: 10.1016/j.infsof.2011.12.008.

[62] M. Staples and M. Niazi, "Experiences using systematic review guidelines," *J. Syst. Softw.*, vol. 80, no. 9, pp. 1425–1437, Sep. 2007, doi: 10.1016/j.jss.2006.09.046.

[63] A. L. Mesquida, A. Mas, E. Amengual, and J. A. Calvo-Manzano, "IT service management process improvement based on ISO/IEC 15504: A systematic review," *Inf. Softw. Technol.*, vol. 54, no. 3, pp. 239–247, Mar. 2012, doi: 10.1016/j.infsof.2011.11.002.

[64] F. O. Bjørnson and T. Dingsøyr, "Knowledge management in software engineering: A systematic review of studied concepts, findings and research methods used," *Inf. Softw. Technol.*, vol. 50, no. 11, pp. 1055–1068, Oct. 2008, doi: 10.1016/j.infsof.2008.03.006.

[65] 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," *J. Syst. Softw.*, vol. 80, no. 4, pp. 571–583, Apr. 2007, doi: 10.1016/j.jss.2006.07.009.

[66] K. Petersen, "Measuring and predicting software productivity: A systematic map and review," *Inf. Softw. Technol.*, vol. 53, no. 4, pp. 317–343, Apr. 2011, doi: 10.1016/j.infsof.2010.12.001.

[67] H. Sharp, N. Baddoo, S. Beecham, T. Hall, and H. Robinson, "Models of motivation in software engineering," *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 219–233, Jan. 2009, doi: 10.1016/j.infsof.2008.05.009.

[68] S. Beecham, N. Baddoo, T. Hall, H. Robinson, and H. Sharp, "Motivation in software engineering: A systematic literature review," *Inf. Softw. Technol.*, vol. 50, nos. 9–10, pp. 860–878, Aug. 2008, doi: 10.1016/j.infsof.2007.09.004.

[69] W. Afzal and R. Torkar, "On the application of genetic programming for software engineering predictive modeling: A systematic review," *Expert Syst. Appl.*, vol. 38, no. 9, pp. 11984–11997, Sep. 2011, doi: 10.1016/j.eswa.2011.03.041.

[70] J. Valaski, A. Malucelli, and S. Reinehr, "Ontologies application in organizational learning: A literature review," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7555–7561, Jun. 2012, doi: 10.1016/j.eswa.2012.01.075.

[71] S. Lane and I. Richardson, "Process models for service-based applications: A systematic literature review," *Inf. Softw. Technol.*, vol. 53, no. 5, pp. 424–439, May 2011, doi: 10.1016/j.infsof.2010.12.005.

[72] R. Prikladnicki and J. L. N. Audy, "Process models in the practice of distributed software development: A systematic review of the literature," *Inf. Softw. Technol.*, vol. 52, no. 8, pp. 779–791, Aug. 2010, doi: 10.1016/j.infsof.2010.03.009.

[73] A. M. Magdaleno, C. M. L. Werner, and R. M. D. Araujo, "Reconciling software development models: A quasi-systematic review," *J. Syst. Softw.*, vol. 85, no. 2, pp. 351–369, Feb. 2012, doi: 10.1016/j.jss.2011.08.028.

[74] V. Alves, N. Niu, C. Alves, and G. Valença, "Requirements engineering for software product lines: A systematic literature review," *Inf. Softw. Technol.*, vol. 52, no. 8, pp. 806–820, Aug. 2010, doi: 10.1016/j.infsof.2010.03.014.

[75] F. Q. Da Silva, A. L. Santos, S. Soares, A. C. C. FranÇa, C. V. Monteiro, and F. F. Maciel, "Six years of systematic literature reviews in software engineering: An updated tertiary study," *Inf. Softw. Technol.*, vol. 53, no. 9, pp. 899–913, Sep. 2011, doi: 10.1016/j.infsof.2011.04.004.

[76] S. Grimstad, M. Jørgensen, and K. Moløkken-østvold, "Software effort estimation terminology: The tower of Babel," *Inf. Softw. Technol.*, vol. 48, no. 4, pp. 302–310, Apr. 2006, doi: 10.1016/j.infsof.2005.04.004.

[77] A. Ampatzoglou and I. Stamelos, "Software engineering research for computer games: A systematic review," *Inf. Softw. Technol.*, vol. 52, no. 9, pp. 888–901, Sep. 2010, doi: 10.1016/j.infsof.2010.05.004.

[78] E. Engström and P. Runeson, "Software product line testing—A systematic mapping study," *Inf. Softw. Technol.*, vol. 53, no. 1, pp. 2–13, Jan. 2011, doi: 10.1016/j.infsof.2010.05.011.

[79] S. Barney, K. Petersen, M. Svahnberg, A. Aurum, and H. Barney, "Software quality trade-offs: A systematic map," *Inf. Softw. Technol.*, vol. 54, no. 7, pp. 651–662, Jul. 2012, doi: 10.1016/j.infsof.2012.01.008.

[80] S. G. Macdonell, "Software source code sizing using fuzzy logic modeling," *Inf. Softw. Technol.*, vol. 45, no. 7, pp. 389–404, May 2003, doi: 10.1016/s0950-5849(03)00011-9.

[81] J. Wen, S. Li, Z. Lin, Y. Hu, and C. Huang, "Systematic literature review of machine learning based software development effort estimation models," *Inf. Softw. Technol.*, vol. 54, no. 1, pp. 41–59, Jan. 2012, doi: 10.1016/j.infsof.2011.09.002.

[82] B. A. Kitchenham, R. Pretorius, D. Budgen, P. Brereton, M. Turner, M. Niazi, and S. G. Linkman, "Systematic literature reviews in software engineering—A tertiary study," *Inf. Softw. Technol.*, vol. 52, no. 8, pp. 792–805, 2010. [Online]. Available: https://doi.org/10.1016/j.infsof.2010.03.006

[83] M. Staples and M. Niazi, "Systematic review of organizational motivations for adopting CMM-based SPI," *Inf. Softw. Technol.*, vol. 50, nos. 7–8, pp. 605–620, Jun. 2008, doi: 10.1016/j.infsof.2007.07.003.

[84] M. Palacios, J. García-Fanjul, and J. Tuya, "Testing in service oriented architectures with dynamic binding: A mapping study," *Inf. Softw. Technol.*, vol. 53, no. 3, pp. 171–189, Mar. 2011, doi: 10.1016/j.infsof.2010.11.014.

[85] J. E. Hannay, T. Dybå, E. Arisholm, and D. I. Sjøberg, "The effectiveness of pair programming: A meta-analysis," *Inf. Softw. Technol.*, vol. 51, no. 7, pp. 1110–1122, Jul. 2009, doi: 10.1016/j.infsof.2009.02.001.

[86] J. Portillo-Rodríguez, A. Vizcaíno, M. Piattini, and S. Beecham, "Tools used in Global Software Engineering: A systematic mapping review," *Inf. Softw. Technol.*, vol. 54, no. 7, pp. 663–685, Jul. 2012, doi: 10.1016/j.infsof.2012.02.006.

[87] E. Loukis, K. Pazalos, and A. Salagara, "Transforming e-services evaluation data into business analytics using value models," *Electron. Commerce Res. Appl.*, vol. 11, no. 2, pp. 129–141, Mar. 2012, doi: 10.1016/j.elerap.2011.12.004.

[88] A. Fernandez, E. Insfran, and S. Abrahão, "Usability evaluation methods for the Web: A systematic mapping study," *Inf. Softw. Technol.*, vol. 53, no. 8, pp. 789–817, Aug. 2011, doi: 10.1016/j.infsof.2011.02.007.

[89] C. Catal and B. Diri, "A systematic review of software fault prediction studies," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7346–7354, May 2009, doi: 10.1016/j.eswa.2008.10.027.

[90] S. M. Mitchell and C. B. Seaman, "A comparison of software cost, duration, and quality for waterfall vs. iterative and incremental development: A systematic review," in *Proc. 3rd Int. Symp. Empirical Softw. Eng. Meas.*, Lake Buena Vista, FL, USA: IEEE Computer Society, Oct. 2009, pp. 511–515, doi: 10.1109/esem.2009.5314228.

[91] S. Jalali and C. Wohlin, "Agile practices in global software engineering—A systematic map," in *Proc. 5th IEEE Int. Conf. Global Softw. Eng.* Princeton, NJ, USA: IEEE Computer Society, Aug. 2010, pp. 45–54, doi: 10.1109/icgse.2010.14.

[92] E. Barreiros, A. Almeida, J. Saraiva, and S. Soares, "A systematic mapping study on software engineering testbeds," in *Proc. Int. Symp. Empirical Softw. Eng. Meas.*, Banff, AB, Canada: IEEE Computer Society, Sep. 2011, pp. 107–116, doi: 10.1109/esem.2011.19.

[93] C. Blanco, J. Lasheras, R. Valencia-Garc, E. Fern, A. Toval, and M. Piattini, "A systematic review and comparison of security ontologies," in *Proc. 3rd Int. Conf. Availability, Rel. Secur.* Barcelona, Spain: IEEE Computer Society, Mar. 2008, pp. 813–820, doi: 10.1109/ares.2008.33.

[94] B. Kitchenham, P. Brereton, M. Turner, M. Niazi, S. Linkman, R. Pretorius, and D. Budgen, "The impact of limited search procedures for systematic literature reviews—A participant-observer case study," in *Proc. 3rd Int. Symp. Empirical Softw. Eng. Meas.* Lake Buena Vista, FL, USA: IEEE Computer Society, Oct. 2009, pp. 336–345, doi: 10.1109/esem.2009.5314238.

[95] M. Riaz, E. Mendes, and E. Tempero, "A systematic review of software maintainability prediction and metrics," in *Proc. 3rd Int. Symp. Empirical Softw. Eng. Meas.* Lake Buena Vista, FL, USA: IEEE Computer Society, Oct. 2009, pp. 367–377, doi: 10.1109/esem.2009.5314233.

[96] H. P. Breivold, M. A. Chauhan, and M. A. Babar, "A systematic review of studies of open source software evolution," in *Proc. Asia–Pacific Softw. Eng. Conf.*, J. Han and T. D. Thu, Eds., Sydney, NSW, Australia: IEEE Computer Society, Nov./Dec. 2010, pp. 356–365, doi: 10.1109/apsec.2010.48.

[97] S. Ali, L. C. Briand, H. Hemmati, and R. K. Panesar-Walawege, "A systematic review of the application and empirical investigation of search-based test case generation," *IEEE Trans. Softw. Eng.*, vol. 36, no. 6, pp. 742–762, Nov. 2010, doi: 10.1109/tse.2009.52.

[98] T. Tahir, G. Rasool, and C. Gencel, "A systematic literature review on software measurement programs," *Inf. Softw. Technol.*, vol. 73, pp. 101–121, May 2016, doi: 10.1016/j.infsof.2016.01.014.

[99] B. Cornelissen, A. Zaidman, A. Van Deursen, L. Moonen, and R. Koschke, "A systematic survey of program comprehension through dynamic analysis," *IEEE Trans. Softw. Eng.*, vol. 35, no. 5, pp. 684–702, Sep. 2009, doi: 10.1109/tse.2009.28.

[100] P. Karpati, G. Sindre, and A. L. Opdahl, "Characterising and analysing security requirements modelling initiatives," in *Proc. 6th Int. Conf. Availability, Rel. Secur.* Vienna, Austria: IEEE Computer Society, Aug. 2011, pp. 710–715, doi: 10.1109/ares.2011.113.

[101] S. G. Macdonell and M. J. Shepperd, "Comparing local and global software effort estimation models—Reflections on a systematic review," in *Proc. 1st Int. Symp. Empirical Softw. Eng. Meas. (ESEM)*, Madrid, Spain: IEEE Computer Society, Sep. 2007, pp. 401–409, doi: 10.1109/esem.2007.45.

[102] S. Beecham, J. Noll, I. Richardson, and N. Ali, "Crafting a global teaming model for architectural knowledge," in *Proc. 5th IEEE Int. Conf. Global Softw. Eng.*, Princeton, NJ, USA: IEEE Computer Society, Aug. 2010, pp. 55–63, doi: 10.1109/ICGSE.2010.15.

[103] B. A. Kitchenham, E. Mendes, and G. H. Travassos, "Cross versus within-company cost estimation studies: A systematic review," *IEEE Trans. Softw. Eng.*, vol. 33, no. 5, pp. 316–329, May 2007, doi: 10.1109/tse.2007.1001.

[104] A. Davis, O. Dieste, A. Hickey, N. Juristo, and A. Moreno, "Effectiveness of requirements elicitation techniques: Empirical results derived from a systematic review," in *Proc. 14th IEEE Int. Requirements Eng. Conf. (RE)*, Paul, MN, USA: IEEE Computer Society, Sep. 2006, pp. 176–185, doi: 10.1109/re.2006.17.

[105] N. Salleh, E. Mendes, and J. Grundy, "Empirical studies of pair programming for CS/SE teaching in higher education: A systematic literature review," *IEEE Trans. Softw. Eng.*, vol. 37, no. 4, pp. 509–525, Jul. 2011, doi: 10.1109/TSE.2010.59.

[106] P. Sfetsos and I. Stamelos, "Empirical studies on quality in agile practices: A systematic literature review," in *Proc. 7th Int. Conf. Qual. Inf. Commun. Technol.*, F. B. e Abreu, J. P. Faria, and R. J. Machado, Eds. Porto, Portugal: IEEE Computer Society, Sep./Oct. 2010, pp. 44–53, doi: 10.1109/QUATIC.2010.17.

[107] M. Unterkalmsteiner, T. Gorschek, A. K. M. M. Islam, C. K. Cheng, R. B. Permadi, and R. Feldt, "Evaluation and measurement of software process improvement—A systematic literature review," *IEEE Trans. Softw. Eng.*, vol. 38, no. 2, pp. 398–424, Mar./Apr. 2012, doi: 10.1109/TSE.2011.26.

[108] J. M. Moreno-Rivera and E. Navarro, "Evaluation of SPL approaches for Webgis development: Sigtel, a case study," in *Proc. 44th Hawaii Int. Int. Conf. Syst. Sci. (HICSS)*, Kauai, HI, USA: IEEE Computer Society, Jan. 2011, pp. 1–10, doi: 10.1109/HICSS.2011.195.

[109] A. Causevic, D. Sundmark, and S. Punnekkat, "Factors limiting industrial adoption of test driven development: A systematic review," in *4th IEEE Int. Conf. Softw. Test., Verification Validation (ICST)*. Berlin, Germany: IEEE Computer Society, Mar. 2011, pp. 337–346, doi: 10.1109/ICST.2011.19.

[110] R. Berntsson-Svensson, M. Höst, and B. Regnell, "Managing quality requirements: A systematic review," in *Proc. 36th EUROMICRO Conf. Softw. Eng. Adv. Appl. (SEAA)*, Lille, France: IEEE Computer Society, Sep. 2010, pp. 261–268, doi: 10.1109/SEAA.2010.55.

[111] R. Y. Huang and J. Symonds, "Mobile marketing evolution: Systematic literature review on multi-channel communication and multi-characteristics campaign," in *Proc. 12th IEEE Int. Enterprise Distrib. Object Comput. Conf. (EDOCW)*, Auckland, New Zealand: IEEE Computer Society, Sep. 2009, pp. 157–165, doi: 10.1109/EDOCW.2009.5332001.

[112] P. Brereton, M. Turner, and R. Kaur, "Pair programming as a teaching tool: A student review of empirical studies," in *Proc. 22nd Conf. Softw. Eng. Edu. Training (CSEET)*. Hyderabad, India: IEEE Computer Society, Feb. 2009, pp. 240–247, doi: 10.1109/CSEET.2009.11.

[113] N. Condori-Fernández, M. Daneva, K. Sikkel, and A. Herrmann, "Practical relevance of experiments in comprehensibility of requirements specifications," in *Proc. 1st Int. Workshop Empirical Requirements Eng. (EmpiRE)*. Trento, Italy: IEEE Computer Society, Aug. 2011, pp. 21–28, doi: 10.1109/EmpiRE.2011.6046251.

[114] M. J. Monasor, A. Vizcaino, M. Piattini, and I. Caballero, "Preparing students and engineers for global software development: A systematic review," in *Proc. 5th Int. IEEE Int. Conf. Global Softw. Eng.* Princeton, NJ, USA: IEEE Computer Society, Aug. 2010, pp. 177–186, doi: 10.1109/icgse.2010.28.

[115] G. K. Hanssen, D. Smite, and N. B. Moe, "Signs of agile trends in global software engineering research: A tertiary study," in *Proc. IEEE 6th Int. Conf. Global Softw. Eng. Workshop*. Helsinki, Finland: IEEE Computer Society, Aug. 2011, pp. 17–23, doi: 10.1109/icgse-w.2011.12.

[116] H. Zhang, B. A. Kitchenham, and D. Pfahl, "Software process simulation modeling: Facts, trends and directions," in *Proc. 15th Asia–Pacific Softw. Eng. Conf. (APSEC)*. Beijing, China: IEEE Computer Society, Dec. 2008, pp. 59–66, doi: 10.1109/APSEC.2008.50.

[117] S. Noponen, J. Salonen, H. Sihvonen, and T. A. Kurki, "Systematic literature review of virtual role," in *Proc. 6th Int. Conf. Internet Technol. Secured Trans. (ICITST)*, Abu Dhabi, UAE, Dec. 2011, pp. 738–743. [Online]. Available: http://ieeexplore.ieee.org/document/6148430/

[118] L. Major, T. Kyriacou, and O. Brereton, "Systematic literature review: Teaching novices programming using robots," *IET Softw.*, vol. 6, no. 6, pp. 502–513, 2012, doi: 10.1049/iet-sen.2011.0125.

[119] O. Dieste and N. Juristo, "Systematic review and aggregation of empirical studies on elicitation techniques," *IEEE Trans. Softw. Eng.*, vol. 37, no. 2, pp. 283–304, Mar. 2011, doi: 10.1109/TSE.2010.33.

[120] S. Kollanus, "Test-driven development—Still a promising approach?" in *Proc. 7th Int. Conf. Qual. Inf. Commun. Technol. Qual. Inf. Commun. Technol. (QUATIC)*, F. B. e Abreu, J. P. Faria, and R. J. Machado, Eds. Porto, Portugal: IEEE Computer Society, Sep./Oct. 2010, pp. 403–408, doi: 10.1109/QUATIC.2010.73.

[121] D. Liu, Q. Wang, and J. Xiao, "The role of software process simulation modeling in software risk management: A systematic review," in *Proc. 3rd Int. Symp. Empirical Softw. Eng. Meas.* Lake Buena Vista, FL, USA: IEEE Computer Society, Oct. 2009, pp. 302–311, doi: 10.1109/esem.2009.5315982.

[122] M. Kalinowski, G. H. Travassos, and D. N. Card, "Towards a defect prevention based process improvement approach," in *Proc. 34th Euromicro Conf. Softw. Eng. Adv. Appl.* Parma, Italy: IEEE Computer Society, Sep. 2008, pp. 199–206, doi: 10.1109/seaa.2008.47.

[123] T. S. da Silva, A. Martin, F. Maurer, and M. S. Silveira, "User-centered design and agile methods: A systematic review," in *Proc. Agile Conf. (AGILE)*. Salt Lake City, UT, USA: IEEE Computer Society, Aug. 2011, pp. 77–86, doi: 10.1109/AGILE.2011.24.

[124] E. Hossain, M. A. Babar, and H. Paik, "Using scrum in global software development: A systematic literature review," in *Proc. 4th IEEE Int. Conf. Global Softw. Eng. (ICGSE)*. Limerick, Ireland: IEEE Computer Society, Jul. 2009, pp. 175–184, doi: 10.1109/ICGSE.2009.25.

[125] A. Maglyas, U. Nikula, and K. Smolander, "What do we know about software product management?—A systematic mapping study," in *Proc. 5th Int. Workshop Softw. Product Manage. (IWSPM)*, M. Kauppinen and K. Wnuk, Eds. Trento, Italy: IEEE Computer Society, Aug. 2011, pp. 26–35, doi: 10.1109/IWSPM.2011.6046201.

[126] M. Jorgensen and M. Shepperd, "A systematic review of software development cost estimation studies," *IEEE Trans. Softw. Eng.*, vol. 33, no. 1, pp. 33–53, Jan. 2007, doi: 10.1109/TSE.2007.256943.

[127] M. Razavian and P. Lago, "A frame of reference for SOA migration," in *Proc. Towards Service-Based Internet, 3rd Eur. Conf.* (Lecture Notes in Computer Science), vol. 6481, E. D. Nitto and R. Yahyapour, Eds. Ghent, Belgium: Springer, Dec. 2010, pp. 150–162, doi: 10.1007/978-3-642-17694-4_13.

[128] A. Y. Teka, N. Condori-Fernández, and B. Sapkota, "A systematic literature review on service description methods," in *Proc. 18th Int. Work. Conf., Requirements Eng., Found. Softw. Qual. (REFSQ)* (Lecture Notes in Computer Science), vol. 7195, B. Regnell and D. E. Damian, Eds. Essen, Germany: Springer, Mar. 2012, pp. 239–255, doi: 10.1007/978-3-642-28714-5_22.

[129] G. Loniewski, E. Insfrán, and S. Abrah ao, "A systematic review of the use of requirements engineering techniques in model-driven development," in *Proc. 13th Int. Conf. Model Driven Eng. Lang. Syst. (MODELS)* (Lecture Notes in Computer Science), vol. 6395, D. C. Petriu, N. Rouquette, and Ø. Haugen, Eds. Oslo, Norway: Springer, Oct. 2010, pp. 213–227, doi: 10.1007/978-3-642-16129-2_16.

[130] T. Yue, L. C. Briand, and Y. Labiche, "A systematic review of transformation approaches between user requirements and analysis models," *Requirements Eng.*, vol. 16, no. 2, pp. 75–99, Jun. 2011, doi: 10.1007/s00766-010-0111-y.

[131] Y. A. Khan, M. O. Elish, and M. El-Attar, "A systematic review on the impact of CK metrics on the functional correctness of object-oriented classes," in *Proc. 12th Int. Conf., Comput. Sci. Appl. (ICCSA)* (Lecture Notes in Computer Science), vol. 7336, B. Murgante, O. Gervasi, S. Misra, N. Nedjah, A. M. A. C. Rocha, D. Taniar, and B. O. Apduhan, Eds. Salvador de Bahia, Brazil: Springer, Jun. 2012, pp. 258–273, doi: 10.1007/978-3-642-31128-4_19.

[132] I. Steinmacher, A. P. Chaves, and M. A. Gerosa, "Awareness support in global software development: A systematic review based on the 3C collaboration model," in *Proc. 16th Int. Conf. Collaboration Technol. (CRIWG)* (Lecture Notes in Computer Science), vol. 6257, G. L. Kolfschoten, T. Herrmann, and S. G. Lukosch, Eds. Maastricht, The Netherlands: Springer, Sep. 2010, pp. 185–201, doi: 10.1007/978-3-642-15714-1_15.

[133] R. L. Vivian, E. H. M. Huzita, G. C. L. Leal, and A. P. C. Steinmacher, "Context-awareness on software artifacts in distributed software development: A systematic review," in *Proc. 17th Int. Conf. Collaboration Technol. (CRIWG)* (Lecture Notes in Computer Science), vol. 6969, A. S. Vivacqua, C. Gutwin, and M. R. S. Borges, Eds. Paraty, Brazil: Springer, Oct. 2011, pp. 30–44, doi: 10.1007/978-3-642-23801-7_3.

[134] S. Kollanus, "Critical issues on test-driven development," in *Proc. 12th Int. Conf. Product-Focused Softw. Process Improvement (PROFES)* (Lecture Notes in Business Information Processing), vol. 6759, D. Caivano, M. Oivo, M. T. Baldassarre, and G. Visaggio, Eds. Torre Canne, Italy: Springer, Jun. 2011, pp. 322–336, doi: 10.1007/978-3-642-21843-9_25.

[135] D. Blanes, E. Insfrán, and S. Abrahão, "Requirements engineering in the development of multi-agent systems: A systematic review," in *Proc. 10th Int. Conf. Intell. Data Eng. Automated Learn. (IDEAL)* (Lecture Notes in Computer Science), vol. 5788, E. Corchado and H. Yin, Eds. Burgos, Spain: Springer, Sep. 2009, pp. 510–517, doi: 10.1007/978-3-642-04394-9_62.

[136] M. Ivarsson and T. Gorschek, "Technology transfer decision support in requirements engineering research: A systematic review of REj," *Requirements Eng.*, vol. 14, no. 3, pp. 155–175, Jul. 2009, doi: 10.1007/s00766-009-0080-1.

[137] C. Moraga, M. A. Moraga, C. Calero, and A. Caro, "Towards the discovery of data quality attributes for Web portals," in *Proc. 9th Int. Conf. Web Eng. (ICWE)* (Lecture Notes in Computer Science), vol. 5648, M. Gaedke, M. Grossniklaus, and O. Díaz, Eds. Donostia-San Sebastian, Spain: Springer, Jun. 2009, pp. 251–259, doi: 10.1007/978-3-642-02818-2_20.

[138] S. D. S. Reis and R. O. Prates, "Applicability of the semiotic inspection method: A systematic literature review," in *Proc. 10th Brazilian Symp. Hum. Factors Comput. Syst., 5th Latin Amer. Conf. Human Comput. Interact. (IHC, CLIHC)*, A. S. Gomes, C. S. de Souza, and J. A. Sánchez, Eds. Porto de Galinhas, Brazil: Brazilian Computer Society, Oct. 2011, pp. 177–186. [Online]. Available: http://dl.acm.org/citation.cfm?id=2254468

[139] A. Arcuri and L. C. Briand, "A practical guide for using statistical tests to assess randomized algorithms in software engineering," in *Proc. 33rd Int. Conf. Softw. Eng. (ICSE)*, R. N. Taylor, H. C. Gall, and N. Medvidovic, Eds. Honolulu, HI, USA: ACM, May 2011, pp. 1–10, doi: 10.1145/1985793.1985795.

[140] R. Rasmussen, "Electronic whiteboards in emergency medicine: A systematic review," in *Proc. ACM Int. Health Inform. Symp. (IHI)*, G. Luo, J. Liu, and C. C. Yang, Eds. Miami, FL, USA: ACM, Jan. 2012, pp. 483–492, doi: 10.1145/2110363.2110418.

[141] E. Engström, M. Skoglund, and P. Runeson, "Empirical evaluations of regression test selection techniques: A systematic review," in *Proc. 2nd Int. Symp. Empirical Softw. Eng. Meas. (ESEM)*, H. D. Rombach, S. G. Elbaum, and J. Münch, Eds. Kaiserslautern, Germany: ACM, Oct. 2008, pp. 22–31, doi: 10.1145/1414004.1414011.

[142] R. C. de Boer and R. Farenhorst, "In search of 'architectural knowledge,'" in *Proc. 3rd Int. Workshop Sharing Reusing Architectural Knowl. (SHARK)*, P. Avgeriou, P. Lago, and P. Kruchten, Eds. Leipzig, Germany: ACM, May 2008, pp. 71–78, doi: 10.1145/1370062.1370080.

[143] S. do R. S. de Souza, M. A. S. Brito, R. A. Silva, P. S. L. de Souza, and E. Zaluska, "Research in concurrent software testing: A systematic review," in *Proc. 9th Workshop Parallel Distrib. Syst., Test., Anal., Debugging, (PADTAD)*, J. Lourenço and E. Farchi, Eds. Toronto, ON, Canada: ACM, Jul. 2011, pp. 1–5, doi: 10.1145/2002962.2002964.

[144] D. S. Cruzes and T. Dybå, "Synthesizing evidence in software engineering research," in *Proc. ACM-IEEE Int. Symp. Empirical Softw. Eng. Meas. (ESEM)*, G. Succi, M. Morisio, and N. Nagappan, Eds. Bolzano, Italy: ACM, 2010, doi: 10.1145/1852786.1852788.

[145] R. L. Glass, V. Ramesh, and I. Vessey, "An analysis of research in computing disciplines," *Commun. ACM*, vol. 47, no. 6, pp. 89–94, Jun. 2004, doi: 10.1145/990680.990686.

[146] P. Mohagheghi, V. Dehlen, and T. Neple, "Definitions and approaches to model quality in model-based software development—A review of literature," *Inf. Softw. Technol.*, vol. 51, no. 12, pp. 1646–1669, Dec. 2009, doi: 10.1016/j.infsof.2009.04.004.

[147] J. Pardillo and C. Cachero, "Domain-specific language modelling with UML profiles by decoupling abstract and concrete syntaxes," *J. Syst. Softw.*, vol. 83, no. 12, pp. 2591–2606, Dec. 2010, doi: 10.1016/j.jss.2010.08.019.

[148] S. U. Khan, M. Niazi, and R. Ahmad, "Factors influencing clients in the selection of offshore software outsourcing vendors: An exploratory study using a systematic literature review," *J. Syst. Softw.*, vol. 84, no. 4, pp. 686–699, Apr. 2011, doi: 10.1016/j.jss.2010.12.010.

[149] J. Nicolás and A. Toval, "On the generation of requirements specifications from software engineering models: A systematic literature review," *Inf. Softw. Technol.*, vol. 51, no. 9, pp. 1291–1307, Sep. 2009, doi: 10.1016/j.infsof.2009.04.001.

[150] J. C. De A. Biolchini, P. G. Mian, A. C. C. Natali, T. U. Conte, and G. H. Travassos, "Scientific research ontology to support systematic review in software engineering," *Adv. Eng. Informat.*, vol. 21, no. 2, pp. 133–151, Apr. 2007, doi: 10.1016/j.aei.2006.11.006.

[151] S. Kakarla, S. Momotaz, and A. S. Namin, "An evaluation of mutation and data-flow testing: A meta-analysis," in *Proc. 4th IEEE Int. Conf. Softw. Test., Verification Validation (ICST)*, Berlin, Germany, Mar. 2011, pp. 366–375, doi: 10.1109/ICSTW.2011.51.

[152] H. F. Landim, A. B. Albuquerque, and T. C. Macedo, "Procedures and conditions that influence on the efficiency of some agile practices," in *Proc. 7th Int. Conf. Qual. Inf. Commun. Technol. (QUATIC)*, Porto, Portugal, F. B. e Abreu, J. P. Faria, and R. J. Machado, Eds. Sep./Oct. 2010, pp. 385–390, doi: 10.1109/QUATIC.2010.70.

[153] E. Hossain, M. A. Babar, H. Paik, and J. M. Verner, "Risk identification and mitigation processes for using scrum in global software development: A conceptual framework," in *Proc. 16th Asia–Pacific Softw. Eng. Conf. (APSEC)*, Batu Ferringhi, Malaysia, S. Sulaiman, and N. M. M. Noor, Eds., Dec. 2009, pp. 457–464, doi: 10.1109/APSEC.2009.56.

[154] M. T. Sletholt, J. E. Hannay, D. Pfahl, and H. P. Langtangen, "What do we know about scientific software development's agile practices?" *Comput. Sci. Eng.*, vol. 14, no. 2, pp. 24–37, Mar. 2012, doi: 10.1109/mcse.2011.113.

[155] D. Wahyudin, R. Ramler, and S. Biffl, "A framework for defect prediction in specific software project contexts," in *Proc. Softw. Eng. Techn.-3rd IFIP TC 2 Central East Eur. Conf. CEE-SET*, Brno, Czech Republic, vol. 4980, in Lecture Notes in Computer Science, Z. Huzar, R. Kocí, B. Meyer, B. Walter, and J. Zendulka, Eds. Springer, 2008, pp. 261–274, doi: 10.1007/978-3-642-22386-0_20.

[156] Q. Gu and P. Lago, "Service identification methods: A systematic literature review," in *Proc. 3rd Eur. Conf., ServiceWave Towards a Service-Based Internet*, Ghent, Belgium, vol. 6481, in Lecture Notes in Computer Science, E. D. Nitto and R. Yahyapour, Eds. Cham, Switzerland: Springer, Dec. 2010, pp. 37–50, doi: 10.1007/978-3-642-17694-4_4.

[157] U. Hyrkkänen and S. Nenonen, "The virtual workplace of a mobile employee—How does Vischer's model function in identifying physical, functional and psychosocial fit?" in *Proc. 14th Int. Conf. Hum.-Comput. Interact. Towards Mobile Intell. Interact. Environments (HCI)*, Orlando, FL, USA, vol. 6763, in Lecture Notes in Computer Science, J. A. Jacko, Ed. Cham, Switzerland: Springer, Jul. 2011, pp. 69–75, doi: 10.1007/978-3-642-21616-9_8.

[158] L. Chen, M. A. Babar, and N. Ali, "Variability management in software product lines: A systematic review," in *Proc. 13th Int. Conf. Softw. Product Lines (SPLC)* (International Conference Proceeding Series), vol. 446, D. Muthig and J. D. McGregor, Eds. San Francisco, CA, USA: ACM, Aug. 2009, pp. 81–90. [Online]. Available: https://dl.acm.org/citation.cfm?id=1753247

[159] S. Heckman and L. Williams, "A systematic literature review of actionable alert identification techniques for automated static code analysis," *Inf. Softw. Technol.*, vol. 53, no. 4, pp. 363–387, Apr. 2011, doi: 10.1016/j.infsof.2010.12.007.

[160] M. Dyer, M. Shepperd, and C. Wohlin, "Systematic reviews in evidence-based software technology and software engineering," *Inf. Softw. Technol.*, vol. 47, no. 1, p. 1, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0950584904001636

[161] J. F. Bastos, P. A. Da M. S. Neto, E. S. De Almeida, and S. R. De L. Meira, "Adopting software product lines: A systematic mapping study," in *Proc. 15th Annu. Conf. Eval. Assessment Softw. Eng. (EASE)*, 2011, pp. 11–20.

[162] C. Monteiro, D. F. Arcoverde, F. Q. B. Da Silva, and H. S. Ferreira, "Software support for the Fuzzy Front End stage of the innovation process: A systematic literature review," in *Proc. IEEE Int. Conf. Manage. Innovation Technol.*, Jun. 2010, pp. 426–431.

[163] K.-J. Stol, M. A. Babar, B. Russo, and B. Fitzgerald, "The use of empirical methods in open source software research: Facts, trends and future directions," in *Proc. ICSE Workshop Emerging Trends Free/Libre/Open Source Softw. Res. Develop.*, May 2009, pp. 19–24.

[164] M. Sulayman and E. Mendes, "An extended systematic review of software process improvement in small and medium web companies," in *Proc. 15th Annu. Conf. Eval. Assessment Softw. Eng. (EASE)*, 2011, pp. 134–143.

[165] K. A. Sedek, S. Sulaiman, and M. A. Omar, "A systematic literature review of interoperable architecture for e-government portals," in *Proc. Malaysian Conf. Softw. Eng.*, Dec. 2011, pp. 82–87.

[166] S. P. Shashank, P. Chakka, and D. V. Kumar, "A systematic literature survey of integration testing in component-based software engineering," in *Proc. Int. Conf. Comput. Commun. Technol. (ICCCT)*, Sep. 2010, pp. 562–568.

[167] A. Franca, T. Gouveia, P. Santos, C. Santana, and F. Da Silva, "Motivation in software engineering: A systematic review update," in *Proc. 15th Annu. Conf. Eval. Assessment Softw. Eng. (EASE)*, 2011, pp.154–163.

AFFAN YASIN received the Bachelor of Science degree in computer science (BSCS) from the National University of Computer, and Emerging Sciences (NUCES-FAST), Lahore, Pakistan, and the Master of Science in Software Engineering (MSSE) degree from Blekinge Tekniska Högskola (BTH), Karlskrona, Sweden. He is currently pursuing the Ph.D. degree with Tsinghua University, Beijing, China, under the supervision of Prof. J. Wang and Assoc. Prof. L. Liu. His areas of interests include empirical research and development within software engineering, game-based learning, social engineering, serious game, and requirements engineering.

RUBIA FATIMA received the master's degree in information technology (MSIT) from Bahauddin Zakariya University (BZU), Multan, Pakistan. She is currently pursuing the Ph.D. degree with Tsinghua University, Beijing, China, under the supervision of Prof. J. Wang and Assoc. Prof. L. Liu. Her areas of interests include empirical research and development within software engineering, game-based learning, social engineering, serious game, and requirements engineering.

LIJIE WEN received the B.S. and Ph.D. degrees in computer science and technology from Tsinghua University, Beijing, China, in 2000 and 2007, respectively. He is currently an Associate Professor with the School of Software, Tsinghua University. He has published more than 110 conference papers and journal articles, which have been cited more than 2500 times by Google Scholar. His research interests are focused on process mining, process data management (For example, log completeness, trace clustering, process similarity, process indexing, and retrieval), and lifecycle management of workflow for big data analysis.

WASIF AFZAL received the Ph.D. degree in software engineering from the Blekinge Institute of Technology. He is currently a Senior Lecturer with the Software Testing Laboratory, Mälardalen University. His research interests include software testing, empirical software engineering, and decision-support tools for software verification and validation.

MUHAMMAD AZHAR received the master's degree in computer science from Sejong University, South Korea, in 2014. He is currently pursuing the Ph.D. degree with the College of Computer Science and Software Engineering, Big Data Institute, Shenzhen University, Shenzhen, China. His research focuses on data mining and machine learning.

RICHARD TORKAR is currently serving as a Professor with the Software Engineering Division at Chalmers and the University of Gothenburg, Sweden. He is also the Head of Software Engineering Division, Department of Computer Science and Engineering. Besides this, he is a Senator at the Faculty Senate, Chalmers University of Technology. He is ranked sixth in Europe (tenth in the world) in the category of most active Consolidated Software Engineering Researchers in Top-Quality Journals 2010–2017, and on the list of most Impactful Consolidated Software Engineering Researchers in general.

● ● ●