Information and Software Technology 105 (2019) 117–149

Contents lists available at ScienceDirect

Information and Software Technology

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

CERSE - Catalog for empirical research in software engineering: A Systematic mapping study

Jefferson Seide Molléri*, Kai Petersen, Emilia Mendes

BTH - Blekinge Tekniska Högskola Sweden

ARTICLE INFO

Keywords:
Empirical research
Empirical methods
Mapping study

ABSTRACT

Context Empirical research in software engineering contributes towards developing scientific knowledge in this field, which in turn is relevant to inform decision-making in industry. A number of empirical studies have been carried out to date in software engineering, and the need for guidelines for conducting and evaluating such research has been stressed.

Objective: The main goal of this mapping study is to identify and summarize the body of knowledge on research guidelines, assessment instruments and knowledge organization systems on how to conduct and evaluate empirical research in software engineering.

Method: A systematic mapping study employing manual search and snowballing techniques was carried out to identify the suitable papers. To build up the catalog, we extracted and categorized information provided by the identified papers.

Results: The mapping study comprises a list of 341 methodological papers, classified according to research methods, research phases covered, and type of instrument provided. Later, we derived a brief explanatory review of the instruments provided for each of the research methods.

Conclusion: We provide: an aggregated body of knowledge on the state of the art relating to guidelines, assessment instruments and knowledge organization systems for carrying out empirical software engineering research; an exemplary usage scenario that can be used to guide those carrying out such studies is also provided. Finally, we discuss the catalog's implications for research practice and the needs for further research.

1. Introduction

Methodological guidance for research practice is more often provided in the form of procedures or guidelines i.e., a step-by-step description of the particular research process and its underlying actions. Assessment instruments are sometimes used to ensure that the process is followed according to particular evaluation criteria.

The challenge of selecting a suitable research method and furthermore to identify appropriate guidelines to apply such method is not new to Empirical Software Engineering (ESE) [254,318]. Researchers such as Shaw [13] highlight the lack of guidance in designing research and organizing the reporting of ESE studies.

More recently, Wohlin [17] recommended a set of commitments that researchers should be aware of when conducting their studies. One of these commitments is mainly related to using and following available guidelines when conducting research. Ultimately, this commitment raises another issue relating to how to identify the appropriate guidelines and other related methodological support.

In the field of medicine, there has been an initiative [15] to address such issue. The EQUATOR Network comprises a comprehensive catalog of guidelines for health research. We believe that a similar initiative in SE will raise researchers' awareness regarding the wider range of available guidelines for the methods they are using [17,318], hence supporting them to make informed decisions of how to design their research.

Thus, this mapping study contributes to the body of knowledge in SE by proposing a catalog of existing guidelines, assessment instruments, and knowledge organization systems (see a glossary of terms in the Appendix A.1) supporting different kinds of empirical methods. We further make the catalog available to researchers by means of the CERSE Web tool[1], which also provides searching and filtering functionalities. The Web tool implements a step-by-step strategy to identify and select the guidelines more suitable for the particular research. We intend to update the catalog periodically, so making the more recent

* Corresponding author.
E-mail address: jefferson.molleri@bth.se (J.S. Molléri).
[1] Available at: http://www.cerse.org

https://doi.org/10.1016/j.infsof.2018.08.008
Received 6 February 2017; Received in revised form 23 July 2018; Accepted 13 August 2018
Available online 28 August 2018
0950-5849/

methodological support also available in our tool.

The remainder of this paper is organized as follows: Section 2 summarizes similar, and related work; Section 3 describes the methodology we employed at searching and selecting the relevant literature for the catalog. In Section 4 we present the findings according to the research methods, data collection and data analysis methods related. Later, Section 5 describes a process for using the catalog to select a set of proper guidelines, and discuss the implications for research practice. We conclude our study in Section 6 providing some suggestion for further work.

2. Related work

2.1. Research methods selection procedures

Early stages of a research design require from researchers many decision-making steps (regarding hypothesis formulation, context definition and research methods employed). They are not trivial choices, as they impact the research process and are likely to shape research results. On the selection of a Research Method, it is essential to recognize the characteristics, advantages, and limitations of each methodology.

Creswell's work [1] is a well-established and recognized literature addressing such issue. The book addresses the whole research process and its underlying steps, further proposing a framework for research design in human and social sciences. Regarding the selection procedure, it provides support on choosing from quantitative, qualitative, and mixed methods designs and methods. In software engineering, Shull *et al.* [14] elicited practical and theoretical knowledge for conducting, reporting and using empirical research methods.

Additionally, specific literature addresses the research method selection issue in software engineering. Eastbrook *et al.* [2] describe five empirical methods available (i.e., experiment, case study, survey, ethnography and action research), identifying the types of questions best addressed by each one. Kitchenham, Linkman and Law [182] proposed DESMET, a methodology for evaluating software tools and methods that employ different research approaches combined. The guidelines support researchers at identifying the appropriated research units and methods to use.

Recently, Wohlin and Aurum [346] propose a framework to support researchers to decide upon which methodologies to employ at strategical, tactical and operational levels. Data collection and data analysis techniques are listed along with more comprehensive research units, such as case study and action research. Relevant literature to each method is also presented, some of which include actual guidelines to conduct the process.

2.2. Aggregations, classifications and secondary studies on methods in software engineering

Studies that extensively investigate a research field in search for methodological guidance are comparable and related to this work. Their importance lies in providing researchers with an opportunity to consider a comprehensive set of available guidelines for their research.

Secondary and tertiary studies aggregating guidelines and supporting literature for research emerged across different fields, such as healthcare [3,9,15] and economy [12]. These studies provide an overview of the state-of-art from a method research's perspective; however, they often focus a specific context e.g. decision-analytic modeling [3] and enterprise development [12].

As previously mentioned, a particular example of such aggregation is the EQUATOR initiative [15]. It provides a searchable database for researchers interested in conducting different studies. The catalog also presents extensions to the main guidelines, providing additional support for its particular phases or contexts e.g., controlled trials, longitudinal observations.

Aggregations resulting from secondary studies on methods are also not unknown in software engineering. Kitchenham & Brereton [4] conducted a systematic literature review (SLR) to identify published experiences of performing such reviews in the SE context. The resulting 63 studies are analyzed in relation to their contributions to each particular task of the SLR process. This work differs from ours by focusing only on the guidelines' extensions obtained through experience reports for a specific research method i.e., SLR.

Furthermore, Borges *et al.* [47] aggregate and classify supporting mechanisms adopted by empirical studies published in major ESE scientific venues (i.e., EASE, ESEM, and ESEJ). A total of 412 mechanisms were identified since 1997, addressing experiment (42%), case study (24%), survey (7%) and other 11 research methods. Although this study has a similar aim to ours, its results differ from ours in a number of ways:

i) we aimed to specifically investigate the guidelines published in the SE domain, instead of assessing the specific individual studies that have employed such guidelines;
ii) we covered a broader set of venues;
iii) we employed snowballing strategies to identify guidelines from other disciplines cited in ESE context (e.g., Information Systems);
iv) we also extracted the additional information we believe is vital for researchers when selecting the appropriate guidelines. Such information include: type and objective of the instrument, process phases addressed, the maturity of the instrument, and relation with other research methods; and
v) we aggregated a catalog of guidelines and supporting instruments for empirical research, which is available online via a Web tool.

3. Method

Our review protocol was developed according to the guidelines for conducting systematic literature reviews (SLRs) and mapping studies (MSs) in software engineering by Kitchenham *et al.* [7,195]. Those guidelines have been extensively used in SE secondary studies, and further validated through a series of papers (e.g., [27,33,54]).

All the researchers involved in this study took part in the development of the protocol. We conducted several meetings to discuss the activities of the MS process, such as specifying the research question, allocating tasks to each researcher, planning the search and selection process, data extraction and synthesis. We also ensure that all changes during the execution of this mapping study and their implications were discussed among the three authors and further updated in the protocol.

3.1. Research question

The objective of this MS is to aggregate and report evidence about existing literature on guidelines, assessment instruments and knowledge organization systems (see a glossary of terms in the Appendix A.1) for conducting and evaluating empirical studies in SE. To achieve the mapping study's main objective we formulated the following research question:

> *What are the available guidelines, assessment instruments and knowledge organization systems for empirical research in software engineering?*

Note that, it is important to categorize and group the supporting literature from multiple perspectives (namely research methods, phases of the research process, and maturity). The categorization is intended to make explicit the coverage of the research field according to the different perspectives.

In general, SLRs and MS report evidence gathered from primary studies, i.e. studies that report the results of empirical research by means of, for example, case studies, surveys and experiments. However, within the context of this study we are using the term "included

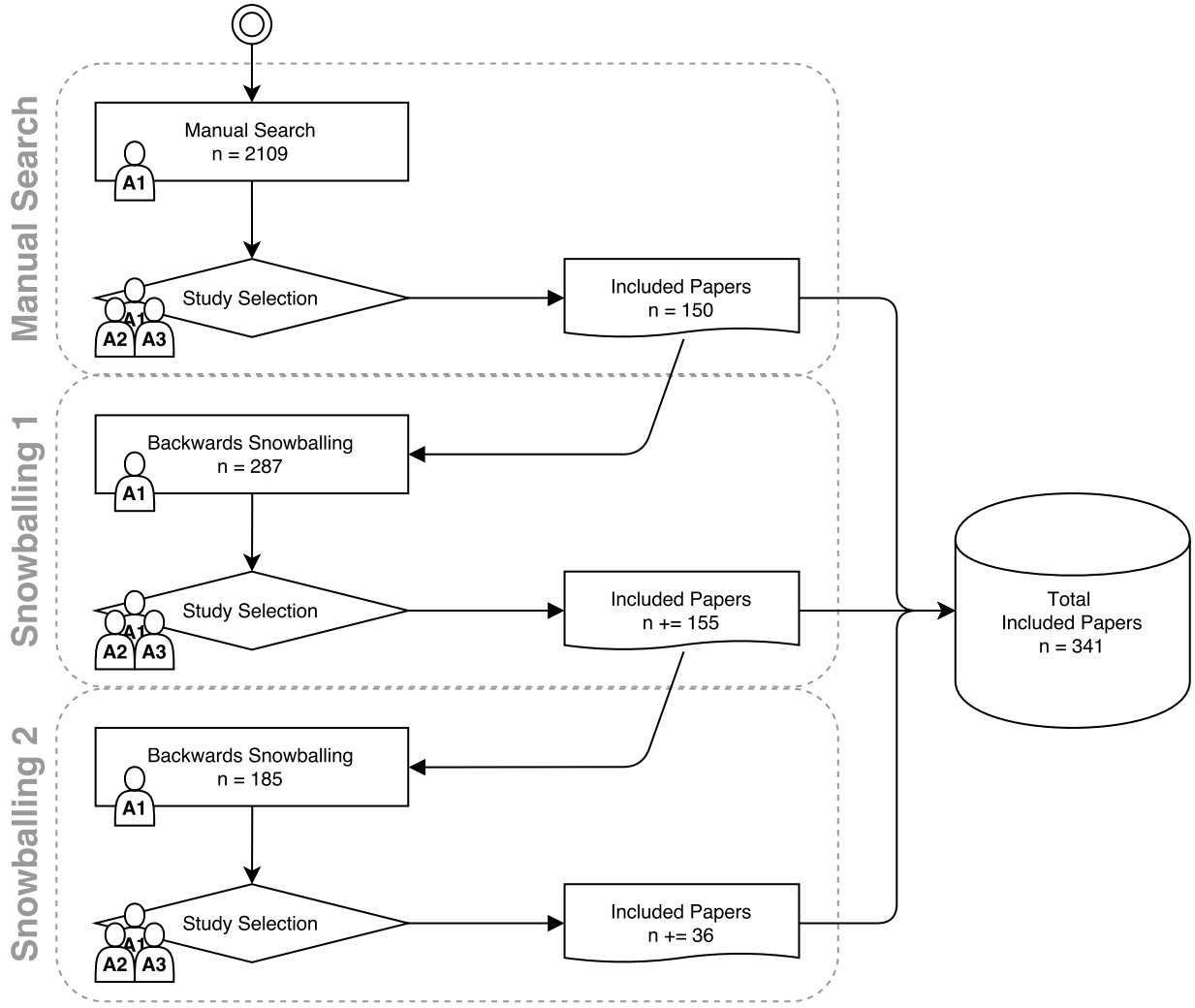


Fig. 1. Overview of the search and selection process. On each step (i.e., manual search and snowballing iterations) we identify a list of candidate studies for selection. Those candidates already excluded any irrelevant paper (i.e., according to exclusion criteria E2, E3, and E4). After each study selection stage, the included papers identified are aggregated into a joint list.

studies" instead, because our results (i.e. guidelines and other supporting literature) are publications that provide methodological support to conduct the so-called primary studies.

3.2. Search and selection process

First, we carried out a database search, as suggested by Petersen, Vakkalanka & Kuzniarz [11]. However, the results from this automated search, after using numerous different combinations of search strings, did not retrieve a set of already known papers. Therefore, we chose to carry out a manual search including a backward snowballing strategy.

We adopted a combination of search and selection strategies [256,344,355]. This mixed approach is intended to identify as much relevant evidence as possible, to ensure the repeatability, and to evaluate the search and selection process. An overview of the search and selection process is presented in Fig. 1, where A1, A2 and A3 represent the first, second and third authors, respectively.

Quasi-gold standard. Prior to conducting the search process, we identify a collection of already-known papers. This quasi-gold standard comprises guidelines to the main research methods in SE domain: action research [212], case study [185,288], experiment [261,353], interviews [295], observation [206], survey [6], systematic literature review [195] and systematic mapping [257].

Piloting a database search. Based on the quasi-gold standard, we derived a search string that we used to pilot 4 major search engines, as proposed by Kitchenham *et al.*[7]: 1) ACM Digital Library 2) EI Compendex 3) IEEE Explore 4) Scopus (which indexes the references by a series of publishers e.g., Springer and Wiley).

The results were not satisfactory due to the high number of false positives, given that not only guidelines but also studies following them were returned. We later produced alternative versions of the search string, but the results did not include all the papers within the quasi-gold standard. Moreover, the number of resulting papers was too high for a reasonable selection process (i.e., more than 100,000 papers).

Manual search. Our first search attempt i.e., a standard database search, provided a high amount of true negatives and false positives. Most of the search results comprised papers that do not present guidelines or other methodological support for conducting empirical research. Hence, the more effective alternative was to start with a manual search on the main fora where many papers with a methodological focus are published, namely:

EASE International Conference on Evaluation and Assessment in Software Engineering (1997–2014)
ESEM International Symposium on Empirical Software Engineering and Measurement (2007–2014)

ISESE International Symposium on Empirical Software Engineering (2002–2006)
IWSM-Mensura International Workshop on Software Measurement (1991–2014)
METRICS International Software Metrics Symposium (1993–2005)
EMSE Empirical Software Engineering Journal (1996–2014)

Through the snowballing process, additional venues that are not part of the starting set provided relevant papers. Some of these venues are Information and Software Technology Journal (21 included papers), ACM SIGSOFT Software Engineering Notes (20 included papers) and IEEE Transactions on Software Engineering (12 included papers).

Selection criteria. After identifying the main fora, we specified a set of selection criteria. These inclusion and exclusion criteria are meant to classify the candidate studies, and also to select those that are suitable to be included in the catalog i.e., answer the research question in Section 3.1.

- **Inclusion criteria:**
  I1 methodological articles for conducting empirical research;
  I2 checklists and instruments to assess the quality of empirical research;
  I3 ontologies, taxonomies, and artifacts to classify and characterize empirical research; and
  I4 articles reporting lessons learned and/or recommendations to improve the research process.
- **Exclusion criteria:**
  E1 nonpeer-reviewed journals, conferences, workshops, and books (i.e., doctoral symposium papers, posters, workshops, and keynotes);
  E2 papers outside to software engineering and related domains (e.g., information technology, computer science, and systems engineering);
  E3 papers not written in English; and
  E4 duplicated and not available papers.

Validating the selection criteria. We further employed a "think-aloud" for selection process on a random sample of the candidate papers to develop a shared understanding between the reviewers. The first author (from hereon A1) performed the selection documenting the reasons to include or exclude a paper while the second author (i.e., A2) acted as an observer. The reasoning was later discussed with the third author (i.e., A3) to help improve and minimize any ambiguities found in the selection criteria.

Piloting the selection. Further, another selected subset of papers was assessed independently by each author as a pilot. Our subset consisted only of papers published in the EASE 2014 and ESEM 2013 conferences, totaling 114 papers. Details about the choice were documented this time, but the disagreements were analyzed using different scenarios for study selection, as proposed by Petersen & Ali [256]. An inter-rater agreement level is calculated based on the outcomes, and as it was considered good (observed kappa = 0.693), the actual study selection started.

Study selection process. In this step, each paper was simultaneously assessed based on titles and abstracts by two authors i.e., A1 assessed all identified papers, while the A2 and A3 each assessed half of the studies. We applied a set of rules for deciding on whether an article was to be included or excluded, as shown in Table 1. Our rules were inclusive, meaning that minor disagreements (i.e., include-uncertain) were solved by including the paper. Further, only papers excluded at least by two authors were removed without any additional assessment. A single vote on excluding a paper (or two uncertainties) led to a disagreement solved by the other author, who had not participated in the selection of the paper.

Snowballing search 1. The results of the study selection process established a collection of potentially relevant sources for backward snowballing, according to Jalali & Wohlin [146]. The source for snowballing was substantial (150 included papers) and clearly fulfilled the criteria for heterogeneity of the starting set (i.e., different authors, publication venues and covered years).

Table 1
Decision rules to study selection. All papers were independently classified as included (I), excluded (E) or uncertain (U). By comparing the assessment of two reviews, we arrive at a decision whether include or exclude the paper. The disagreements (D) were solved by an additional decision vote whether included or excluded.

		1st review		
		I	U	E
2nd review	**I**	I	I	D
	U	I	D	D
	E	D	D	E

Our snowballing approach consisted on reading specific sections of the papers (i.e., background, related work, and methods) likely to provide methodological references. Then, we collected the candidate references, excluded the duplicates (among the ones identified in the same snowballing iteration plus the ones already included by previous iterations) and papers that otherwise were not available to retrieve (see exclusion criteria E4). The resulting 287 references were aggregated into a candidates list and assessed according to the study selection process described above. At the end of this process, 155 additional papers were included.

Snowballing search 2. A second snowballing iteration followed the same process as *snowballing search 1*. Papers included in the previous iteration were our starting set, from which we retrieved 185 candidates. After selection, 36 new papers were added to our included data set.

Analysis of included papers. Once the snowballing search and selection process was finished, we compiled a list of the included papers and analyzed their particular characteristics (e.g., year and venue of publication). Although our starting seed contained just 6 fora (5 conferences plus 1 journal), the snowballing strategy extended our reach to 75 additional venues (59 after the first iteration step plus 16 after the second iteration).

Search and selection performance. Finally, we compared the proportion of resulting candidates (diminishing returns henceforth) and quasi-gold standard (quasi-sensitivity henceforth) in the retrieved references to an optimum threshold of less than 20% and greater than 80%, respectively, as described by Zhang and Babar [355]. The threshold was not reached by the manual search, or by the first snowballing iteration (Table 2), thus we performed the snowballing process a second time.

The decrease in the number of **diminishing returns** of the second snowballing iteration implies that we were approaching a critical mass i.e., the number of retrieved references will exponentially grow until it

Table 2
Analysis of the performance of search and selection process. Diminishing returns is defined as the proportion of candidate studies in the retrieved references, whereas quasi-sensitivity is calculated by comparing the quasi-gold standard to the included papers.

	Iteration Step			
	Manual	Snowballing 1	Snowballing 2	Overall
Retrieved references	2114	744	997	3855
Candidate studies	2109	287	185	2581
Diminishing returns	99.8%	38.6%	18.6%	67%
Already-known papers	10	10	10	10
Already-known included	2	10	10	10
Quasi-sensitivity	20%	100%	100%	100%

reaches the stage beyond which no further study is selected. A high degree of saturation is desirable, as it reflects the completeness of the search process. However, a similar exponential effort would be needed to identify and select such critical mass, that is likely to provide a negligible number of candidate studies.

The **quasi-sensitivity** validated the completeness of our search strategy regarding the quasi gold-standard. During the manual search, just 2 out of the 10 already-known papers were included. At the end of the first snowballing iteration, all papers in the quasi gold-standard were identified and included. This does not imply that all the relevant papers within the search universe were included, instead, it suggests that our search performance is acceptable.

3.3. Data extraction

The data extraction stage comprises the identification and gathering of information needed to address the research question. The first author designed the data extraction instrument in accordance with the research objective and the intended categorization. Further, all the three authors discussed design issues that should be considered. After a few refinement iterations, the data extraction instrument was judged as ready for use. This instrument includes records of relevant information from the included studies, as well some additional bibliographic data, as follows:

D1 Paper title;
D2 Authors and affiliation;
D3 Year of Publication;
D4 Paper type (i.e., conference or journal) and publication venue;
D5 Name and summary of the guideline or supporting instrument targeted by the paper;
D6 Research methods to which the instrument can be applied i.e., experimentation, survey, case study;
D7 Process phases covered by the guideline or supporting instrument i.e., planning, execution, analysis and/or reporting;
D8 Aim of the supporting instrument e.g., to conduct, to classify, to evaluate or to replicate research; and
D9 Maturity score i.e., classification of research facets according to Wieringa *et al.* [343].

D6. Research methods. Aiming to consolidate the knowledge regarding research methods and their relations we investigated the previously identified related work already [47,254,346,353]. However, we found out that different studies presented differing points of view relating to research methods, producing divergences in our classification. Thus, we decided for an initial classification according to [346], updating the classes as needed during the data extraction process. We also validated the new labels in comparison to the existent classifications [2,318,346]. Further, we reviewed the entire classification, aggregating research methods by similarities described in the included papers. Further information regarding the research methods and their relations is given in Section 4.1.

D8. Aim of the supporting instrument. To better understand how the identified instruments can support the research process, we propose a classification according to their intended objectives, as follows: 1. **guidelines** for conducting research; 2. **assessment instruments** to evaluate research; and 3. **knowledge organization systems** to classify or characterize research. For details of this classification see a glossary of terms in the Appendix A.1.

D9. Maturity score. To determine the contribution of each guideline and instrument to our study, we distinguish between proposed and evaluated solutions (see a definition of these terms in the Appendix A.1). Each included study was assessed using the classification of research facets by Wieringa et al. [343], as recommended in Petersen *et al.* [11]. The classification provides a means to identify mature research, as well as instruments that need further investigation:

- Problem investigation: opinion paper (stated position).
- Solution design: philosophical paper (conceptual framework).
- Solution validation: validation research (weak empirical study).
- Solution selection: experience report (lessons learned).
- Solution implementation: solution proposal (proof-of-concept).
- Implementation evaluation: evaluation research (strong empirical study).

Pilot extraction. Prior to the data extraction, the three authors performed a pilot data extraction on a random sample of five selected studies. For each paper, all three authors separately extracted the relevant information (see Section 3.3) comprising mainly qualitative data. We were able to assess the degree of agreement related to the categorical data i.e., D6, D7, D8, and D9, as follows:

D6 Regarding the research method there were two cases of partial disagreement, in which A1 classify the paper into two research method categories (e.g., case study and experiment) and one of the second authors related to just one.
D7 A partial agreement is reported in just one case, due to a set of multiple process phases.
D8 No disagreement is reported regarding the aims of the supporting instrument. However, our sample consisted only of guidelines for conducting research.
D9 One disagreement was reported in relation to the maturity score.

Further, we compared our data collection forms and discussed the disagreements, so to reduce any possible data extraction-related validity threats. We consider that the partial disagreements are not a threat to the data collection, as the first author presented the most comprehensive set. The single disagreement of D9 is due to a divergent interpretation from one of the second authors. We scheduled periodic meetings so A1 could discuss further data extraction with the other two authors.

Validation of data extraction. The extraction process was carried out by the first author of this paper and reviewed by the second author, as suggested by Kitchenham et al. [7]. After the data extraction, a sample of 10% (i.e., 37) of the included papers was validated independently by the second and third authors. The outcomes of the validation were further discussed by all authors.

The validation identified uncertainties regarding a study's classification due to the information provided in the paper. These uncertainties are related to the multiple classifications of the included papers regarding (D6) research methods, (D7) process phases, and (D9) aims of the instrument. Several papers can be classified in more than one category as they provide support for a set of research methods (e.g., a case study using observations for data collection), research phases (e.g., guidelines for analysis and report of studies) and aims (e.g., both to conduct and evaluate the research).

3.4. Analysis and classification

The extracted data was further tabulated and arranged so to be used to answer the research question, as follows:

RQ. *What are the available guidelines, assessment instruments and knowledge organization systems for empirical research in software engineering?* The available guidelines, assessment instruments and knowledge organization systems for empirical research in SE were compiled in a reference list provided in the Appendix A.2. Further, we summarize the findings according to three perspectives:

- *Research method.* An overview of included papers related to each research method is presented and discussed in Section 4.1.
- *Phases of the research process.* An overview of the coverage of process phases (i.e., planning, execution, analysis, and reporting) by guidelines and other supporting instruments is given in

Section 4.2.
- *Evaluation.* More mature instruments are implemented and evaluated in practice rather than proposed and validated through toy examples. In Section 4.3 we discuss the instruments that achieve a better classification according to our maturity score.

3.5. Validity threats

Some potential threats that could affect the validity of this mapping study and corresponding mitigation strategies are detailed below. The threats were classified according to Maxwell's categorization of validity threats [8]:

3.5.1. Descriptive validity

To ensure the objectivity of the process, we undertook several quality measures. First, a **review protocol** was iteratively developed and updated during this study. During the planning phase, we conducted a series of meetings to ensure the protocol completeness and understandability. During the execution and analysis phase, additional meetings addressed responses to changing events. After each meeting, the protocol was updated by A1 and further shared with A2 and A3 for review.

Our **search strategy** combined both manual search and snowballing technique. A manual search was performed at some venues with a specific focus on the development and improvement of methodology in SE domain. Further, two snowballing iterations extended our search beyond those venues, still limited due to the references in the papers included in the previous iteration.

Although we cannot ensure that all the available papers were collected, we employed two actions to validate the search process:

i) before the search activity, we identified a set of already known studies that were further compared to the selected papers; and
ii) the diminishing returns on the search and selection were used as a decision criteria for the snowballing iterations.

Our **snowballing strategy** does not aim to retrieve all the references from the candidate papers, rather a subset of the references contained in the background, related work and methods sections. We assume that other sections are unlikely to cite any new methodological references. We also piloted the **selection criteria** through a think-aloud exercise with a sample of the identified papers. Further, the **study selection process** involved two of the authors, and further disagreements were mediated by the third author.

The *data extraction form* was piloted by all the authors on a sample of included papers. We compared the extraction form to ensure that consistent information was retrieved from us and discuss any divergences. Similar discussions were carried out during the actual data collection, conducted mainly by A1. The first author has prior experience with the SLR process and theoretical understanding of multiple research methods. However, a reliability threat exists whenever the researcher may be influenced by their prior experiences and the clarity of the papers being extracted. The results of data extraction were further discussed with the other two authors in an attempt to minimize any judgmental error that may have happened while categorizing the research methods, related phases, aims, and maturity score.

The systematic process of secondary studies is intended to allow readers to assess its repeatability. Further replication or extension of this review should be grounded on the details provided in the review protocol and employed guidelines. The steps we undertook to carry out this study are detailed in the methods Section 3, and the guidelines are provided mainly by Kitchenham & Chartes' [195].

3.5.2. Interpretive validity

The interpretative validity is related to the objectiveness of data representation aiming to infer conclusions. To interpret the results of this systematic mapping study, we relied on visual and tabular representations of the data. Moreover, the included studies were classified according to the research methods, process phases, aim, and maturity level. This classification allowed us to group the data and to investigate trends.

Our research question is qualitative in nature, and the categorization perspectives could be addressed by counting the number of included papers in a particular group. Ultimately, we aggregate a comprehensive list of the identified methodological studies into a catalog for empirical research.

3.5.3. Theoretical validity

To minimize the potential theoretical threats related to the uncertainty of collecting suitable data to answer the research question, we proposed a study design through careful deliberation and reasoning. Before starting the systematic mapping study, we detailed a research question addressing the review objectives. We further defined variables that could answer such questions, built a strategy to identify the related literature and proposed the instrumentation for data collection. All the three authors (i.e., A1, A2, and A3) actively participated in these steps.

3.5.4. Generalizability

Generalizability is concerned with the application of the results to different contexts and settings. Some of the identified methodological papers are also employed in other scientific domains e.g., business, and human sciences. However, they are not likely to be a comprehensive set of those domains, and therefore our results are not readily applicable to such contexts.

Furthermore, our results address multiple research methods, process phases, and maturity levels. However, the methodological support is not evenly distributed, and some gaps exist in relation to each category. We assume that this lack of equity reflects the actual state of the field. Thus, we suggest that further studies are needed to investigate the gaps and to provide support for all the research methods.

4. Results

Our systematic mapping identified 341 papers that present instruments or provide guidance to ESE research. An extensive list of the included references is provided in Appendix A.4 and is available by means of the Web tool CERSE[2]. This Web tool also allows for researchers to filter the catalog references according to the paper categories, as described in Section 3.3.

With regard to their aims, most of the instruments (236 out of 341 included studies) describe guidelines, frameworks or experiences on how to conduct research. Assessment instruments, such as checklists and evaluation processes, are presented in 74 papers. Finally, circa of 42% (41 papers) provide structures to classify and categorize the research. A number guidelines provide additional support to research-related activities, such as teaching and replication.

The included papers cover a broad spectrum of research methods (e.g., case study, experiment, SLR) and research phases (e.g., planning, execution, and analysis). The instruments are also classified according to their maturity. The categorization provides a structured map of the identified guidelines and supporting instruments, and also highlights existing gaps in the field.

Further, we detail our findings according to three distinct perspectives of categorization:

4.1 Research methods supported by the instrument;
4.2 Phases of the research process covered; and
4.3 Maturity with regarding the evaluation of the instrument.

[2] http://www.cerse.org

4.1. Research methods

In this section, we summarize the coverage of the research methods identified in our study. References to all identified studies according to the classification (i.e., research methods and aims), along with some highlights, are available in Appendix A.2. A subset of these references can also be generated when selecting guidelines and other supporting literature using the CERSE Web tool[3]. The heuristics for guideline selection are detailed in Section 5.1.

Time-trend. Since 1980, several guidelines have been published, first addressing experimental studies, later on case studies, action research and other methodologies. An overview of the temporal trends of the identified literature concerning the different research methods is presented in Fig. 2.

The time-trend plot shows when and how of10 a methodological support for each particular method in SE has been published. Experiment has been the research method with the highest number of methodological support, with 155 references (i.e., 43.4% of the total papers collected) since the first publication in 1980 [80]. Case study research comes second, with 76 references identified, and with the first guideline published in 1987 [41]. In relation to SLRs, we identified 70 papers published since 1997.

One expects that the time-trend results are aligned to the needs for a specific SE-related methodology; however, the frequency in which such "primary studies" are conducted are not aimed by this study. It is as well likely that

(a) primary research is conducted and reported after the methodological support in the context of SE is published; or
(b) guidelines and supporting instruments emerges after primary studies have been published.

Multiple classification. A large part of the instruments target at more than one research method, often employing additional methods for data collection or data analysis. An example of such combinations is illustrated by the Venn diagram presented in Fig. 3.

The Venn diagram shows the relations between case study research (CSR) and some other research methods often employed in such research. CSR often uses different research units to achieve data triangulation, an approach aimed at obtaining stronger evidence. From 72 guideline papers addressing CSR, more than half also discuss additional methods e.g., experiment, survey, and observation.

4.1.1. Case study research

Case study research employs multiple methods of data collection to investigate a phenomenon in its natural settings [41]. The methodology is well suited for many SE research topics, as it addresses a contemporary case in depth. It aims to understand the particular case and create the basis for further research on the topic [288].

It is often associated with an element of triangulation i.e., employing different strategies (e.g., data collection instruments or sources) on the same target. Combining the results from a set of strategies can provide stronger evidence to support the findings that a single point of view.

Its flexibility implies a lack of boundaries as compared to other methodologies, and thus it is sometimes hard to classify case study research [288]. Both qualitative and quantitative approaches have been documented in the literature, and sometimes a mixed approach combines them both.

Summary of findings. We identified 76 papers supporting case study research in SE. Most of them presenting guidelines to conduct the process (58), but also instruments to classify (13) or evaluate the research (11).

[3] http://www.cerse.org

4.1.2. Action research

Action research aims to study a problem in the field by proposing, applying and evaluating a solution to solve it. The methodology is often described by the following phases: 1) diagnosing, 2) action planning, 3) action taking, 4) evaluating, and 5) specifying learning.

The researcher is an active part of the methodology, as he/she investigate the problem and propose action in response to it. Because of that, special measures to ensure the objectivity should be taken. Finally, knowledge is aggregated by collecting insights about the relationship between the intervention and the investigated phenomenon [346].

Action research is usually classified according to its focus and objective [10], namely, such as *i* Action research focusing on change and reflection; *ii* Action science trying to resolve conflicts between espoused and applied theories; *iii* Participatory action research emphasizing participant collaboration; and *iv* Action learning for programmed instruction and experiential learning.

Summary of findings. A total of 24 papers were identified as supporting action research in SE and related fields.

4.1.3. Design science research

While design science (DS) methodology shares the problem-solving paradigm of action research, it focuses on building and evaluating original artifacts (e.g., a model or a framework). These products are designed to meet the requirements of a particular problem and can be added to the shared knowledge base of researchers [337].

Summary of Findings Only 2 guidelines to conduct DS research were identified in our review. The first one [337] focuses on methodological guidelines considering nested knowledge and practice, two intrinsical facets of DS methodology. The other [334] presents a conceptual framework and seven guidelines for design science in IS research.

4.1.4. Interviews and focus groups

Interviews are often used in ESE to investigate the participants' perception regarding the research topic. It usually involves a meeting between the researcher (a.k.a. interviewer) and the participant (a.k.a. interviewee) wherein the interviewer ask questions to collect in-depth qualitative data [141,346]. Thus, interviews are a qualitative method, also defined as one of the field study methods i.e., featuring data collection outside a laboratory setting.

Interviews are classified according to the questions used [346], as such: *i) Structured:* mainly close-ended questions i.e., limiting the interviewee to a list of possible answer choices; *ii) Unstructured:* typically open-ended questions beginning with "what" and "how" to capture interviewee's experience; and *iii) Semi-Structured:* combine both close- and open-ended questions.

Focus groups (also known as focused interviews or focus group interviews) allow the researcher to interact with a group of participants in a manner similar to the interviews [208]. Planning and execution phases are similar to unstructured interviews, but extra care should be taken to compensate its weakness e.g., secrecy and limited comprehension.

Summary of findings. We identified 15 papers presenting instruments to support interviews in SE. However, just two of those addressed interviews alone [141,241]. Alternatively, the focus groups method is sometimes considered to be related to unstructured interviews [208].

4.1.5. Observation, ethnography and think-aloud

Participant observation is an empirical method to collect qualitative data in a less intrusive way than an interview. It consists in systematically inspect the participant to capture behaviors and interactions that might not be noticed otherwise [295,346]. Along with the interview, observation is also considered as a field study.

According to the level of awareness of the observer, the research can be classified as [346]: *i) Overt:* the participant is aware of the process; and *ii) Covert:* the participant is not aware of the observation.

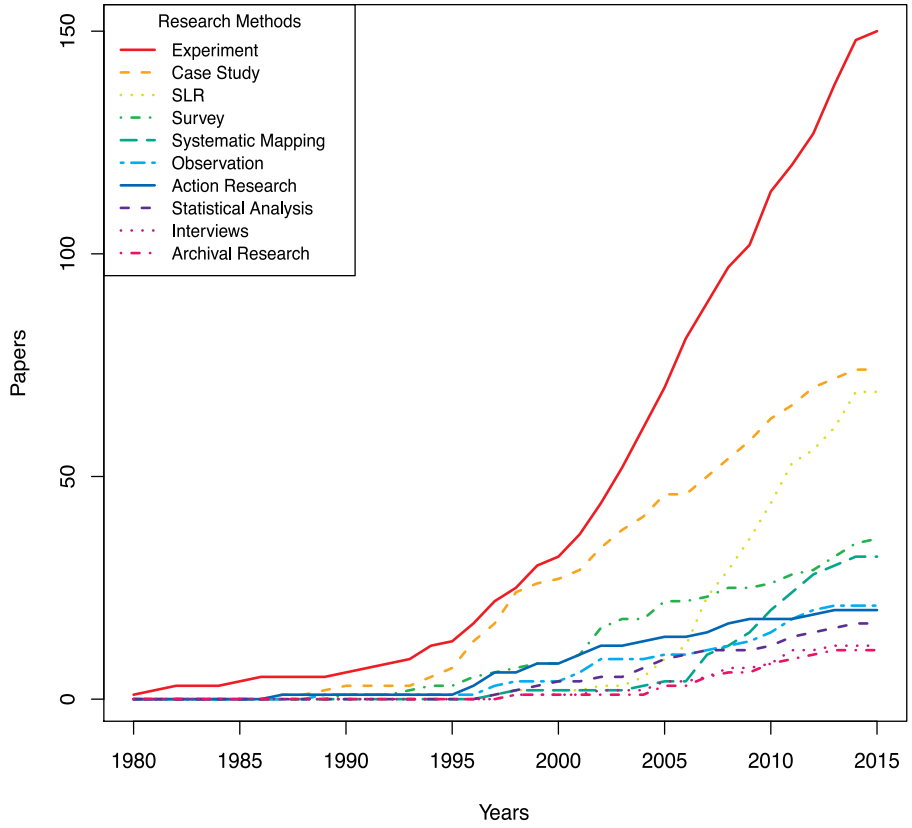


Fig. 2. Time-trend of identified studies according to the topmost 10 cited research methods. The lines show the cumulative sum of methodological papers over the years i.e., each new publication pushes the related line up.

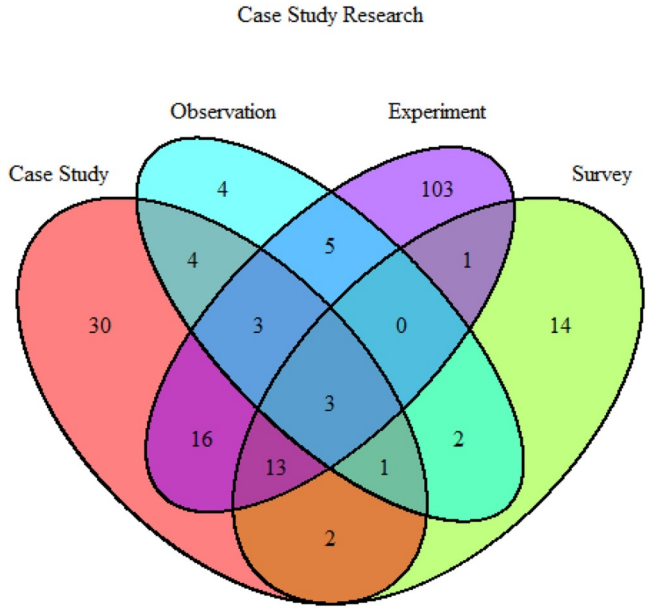


Fig. 3. Venn diagram shows the number of papers covering a set of research methods related to case study research. The diagram is divided into 15 separate sets, each of which presents the count of papers related to a particular research method or the intersection between them i.e., targeting more than one research method.

Ethnography is a methodology similar to the participant observation that requires the researchers to immerse themselves in the investigated environment [283]. The think-aloud protocol involves participants saying whatever comes into their mind (i.e., thinking aloud) as they are observed by researchers.

Summary of findings. Our mapping study identified 28 papers reporting guides and advice to the observation process. Similarly to the interviews, a significant portion of them relates observations to other research methods, such as case studies and action studies.

4.1.6. Archival research

Archival research methods investigate the SE field by looking at their output and by-products e.g., source code, documentation, and reports [215]. It comprises activities such as locating the data, systematically collecting the data, analyzing and interpreting the data [346].

It is non-intrusive and requires little commitment from the subjects. However, real data may sometimes have missing components and/or be difficult to interpret [215]. Moreover, archival research can be a valuable data source for triangulation in both case study and action research.

Depending on the source of data, archival research can be classified as qualitative or quantitative. Qualitative data sources include e.g., meeting minutes, and software documentation, while quantitative data come from e.g., project databases, software repositories, change logs, and fault reports.

Archival research is often performed in industrial settings, but may also be applied over academic papers. According to the decision-making structure presented in [346], systematic literature reviews are archival research applied to academic literature. However, we treated SLRs separately (Section 4.1.7), as the guidelines for secondary studies contain specific steps related to study identification, selection, etc.

Postmorten reviews investigate existing data from concluded projects to gather insights for future cases. These light studies employ retrospective reflection often allied with support from individuals that participated in the project.

Summary of findings. 13 articles related to archival research were collected in our review. None of the papers is focused on archival

research alone, and the advice is often found in field study methodology.

4.1.7. Systematic literature review

Systematic Literature Review (SLR) aims to collect and summarize the available literature regarding a particular investigated topic of the phenomenon. It employs a rigorous and systematic methodology to retrieve a wide range of pieces of research, aggregating their results to draw more general conclusions than individual studies in isolation [195]. The systematic process is often described in three main phases, and its particular activities or stages:

- *Planning:* specify a research question and develop a research protocol, which details the process used to conduct and report the review;
- *Conducting:* conduct a search strategy to retrieve candidate studies that are further selected (to remove irrelevant studies) and assessed (to ensure the quality of evidence); later, data relevant to answer the research question are extracted, aggregated and synthesized;
- *Reporting:* write and validate a report given the target audience and appropriated format.

SLRs are often described as secondary studies, a general category that also includes systematic mapping studies (see Section 4.1.8). SLRs provide a more in-depth analysis of the selected studies than mapping studies, but often require more effort [257].

Summary of findings. A total of 70 papers provide guidance and support to SLR, 30 of them are also related to systematic mapping studies.

4.1.8. Systematic mapping study

Systematic mappings are secondary studies that aggregate the available literature in a more comprehensive and often visual summary. The differences from other secondary studies (i.e., SLRs) also reflect on the process employed, as mapping studies often require less effort. The quality assessment activity is not required, and both study selection and data extraction only consider high-level characteristics of the identified papers [257].

Summary of findings. We identified 35 papers focused systematic mapping studies, most of them also addressing SRLs. Only 5 papers provide specific guidance to systematic mapping alone.

4.1.9. Survey Research and Delphi Method

Surveys aim to collect and summarize evidence from a large representative sample of a population of interest. Outcomes from survey research are then combined to identify patterns with the aim of generalizing to the overall population. Identified patterns can be even compared to different populations and evaluated over time.

According to the research purpose, survey research can be[266]: *i) exploratory:* whether aiming to identify preliminary concepts, or discover new dimensions from the population of interest; *ii) descriptive:* whether investigating the opinions or phenomenon occurring in the population or a subset of the population; or *iii) explanatory:* whether testing theories and causal relations.

Regarding the data collected, survey research can also employ a qualitative or quantitative process, or even a mixed research process when collecting simultaneously qualitative and quantitative data.

Delphi method is a survey-based communication technique to obtain the consensus of a group of experts [247]. Through several rounds, the group discusses a problem and answers questionnaires related to it. An anonymous summary of the responses is provided at the end of each round, thus promoting further discussion and consensus.

Summary of findings. In total, 39 papers supporting the survey process and its activities were identified. Some of them are focused on specific aspects of survey research, as others also cover it along with other field study methods. e.g., observations, interviews, and case study research.

4.1.10. Experiments and quasi-Experiments

Experimental research is a vast though well-explored field on ESE often used to compare two or more alternative treatments (e.g., methods, techniques or tools). Participants are assigned and subject to a treatment condition, and the effects of such are recorded. Further, the data between the treatments are analyzed and compared through statistical methods.

Experimental design is often classified according to its characteristics, such as: *i) laboratory or field setting:* whether conducted in a controlled or real-world environment; *ii) controlled or non-controlled:* whether or not employ controls i.e., individuals not subject to any treatment; *iii) randomized or non-randomized:* whether randomly assign the participants to the treatment or not. *iv) laboratory or quasi-experiments:* randomized controlled trials (RCT) are conducted in a laboratory setting, whereas quasi-experiment designs (a.k.a. natural experiments) are carried out in a field setting.

Summary of findings. Most of the literature identified in our review is related to experiments i.e., 155 out of 341 papers. Very often identified papers relate experiments to case study research, and to observation, survey and simulation methods.

4.1.11. Simulation

Simulation-based studies (SBS) use the model of a real-world entity (e.g., a process or product) to evaluate alternative solutions under investigation. The model can reproduce the components of an actual entity in a virtual state, with varying degrees of accuracy [346]. SBS share several characteristics with the controlled methods for technology evaluation, such as experimental designs [353].

According to the model, simulations can be characterized as: *i) static or dynamic:* the system is time dependent or not; *ii) discrete or continuous:* the investigated events are countable or continuous in time, and *iii) stochastic or deterministic:* the simulation depends on probabilistic behavior or not.

Moreover, the employed research logic also depends on the model: *i) Deductive:* generated from theory and compared to actual observations, and *ii) Inductive:* derivative from gathered data and used to examine the relationships between its components.

Summary of findings. Our review identified only 5 papers related to simulation, most of which are also relevant for experimental research.

4.1.12. Thematic analysis

Also known as thematic synthesis, it is another qualitative data analysis technique used in SE to provide understanding about an investigated topic. It identifies the recurring concepts from multiple data sources exploring the relationships between them to create a model of high-order themes [77].

Although some differences in the literature [77,346], the process can be described by the following phases: *i*) initial reading/familiarizing of data/text; *ii*) identifying initial codes/segments of text; *iii*) labeling/identifying themes; *iv*) reviewing themes/reduce overlaps; *iv*) defining the results.

Although very similar to thematic analysis, content analysis is focused on counting the frequency of concepts within data. It requires well-defined categories and strict rules, and often employ computational tools to aid the process.

Thematic analysis can be classified according to the process to identify the themes as *i) semantic:* when based on explicit meaning in the data; and *ii) latent:* based on underlying ideas within the data that is theorized. The semantic approach tends to be more positivist whereas the latent approach is often interpretivist.

Summary of findings. From 5 papers identified, only one is focused on supporting thematic analysis. Other papers are concern overview various data analysis techniques (e.g., [78,129]) or how to combine various data collection and data analysis methods (e.g., [108,224]).

4.1.13. Grounded theory

The Grounded Theory Method (GTM) is intended to construct a theory from a systematic process of data analysis. It has been used to understand how participants deal with problems concerned to them, based on data gathered from opinions and behaviors [19].

GTM is an openended inquiry process that involves collecting, coding and categorizing qualitative data (e.g., interview transcripts, literature review, or archival data) until a saturation level is reached. Potential relationships between the codes are further interpreted to build theories and interrelated hypotheses. The theories can then be tested by a further study [346].

Two different versions of GTM exist, according to the data coding and interpretation: *i) Glaserian:* identifying participants perspectives data at an abstract level, and later conceptualizes them to find patterns, and *ii) Straussian:* develop categories according to the literature or the researchers experience simultaneously to the coding process.

Summary of findings. 8 identified papers presented support to grounded theory process. Half of them are focused on grounded theory solely, while as four other include other field study or data analysis methods.

4.1.14. Hermeneutics

A qualitative data analysis approach focused on the interpretation of meanings. The method requires interpreting small parts of the data that leads to a understanding of the whole. It is especially suitable for human behavior studies, including information systems and software engineering research [346].

Hermeneutics most fundamental principle is called circle of understanding (a.k.a. hermeneutics circle). The process uses a dialogic approach to interpreting new knowledge connected to prior beliefs. In that sense, a global understanding of the context can lead to an improved understanding of the parts, and vice versa [206]. Hermeneutics is often characterized by philosophical perspectives applied i.e., conservatism, pragmatism, criticism, the radicalism [62].

Summary of findings. We identified just two studies presenting an overview of hermeneutics. They detail the concepts and principles of the approach, discussing its philosophical aspects. Both papers (i.e., [62,206]) outline the application through interpretive studies on the information systems field.

4.1.15. Statistical analysis and meta-Analysis

Statistical methodologies have been used for data analysis in a multitude of fields. The statistical approach relies on the application of mathematical techniques (e.g., linear algebra, differential equations, and probability theory) to interpret quantitative data regarding a population or a model to be studied.

Although quantitative data (e.g., collected in an experiment or survey-based research) is required at statistical analysis; qualitative data, such as concepts and themes, can be converted to quantitative form by e.g., counting the frequency of appearance.

On the one hand, *descriptive statistics* is meant to summarize data by describing, aggregating and presenting association between the variables. On the other hand, *inferential statistics* includes hypotheses testing, regression analysis, and estimation through data mining techniques [346]. Moreover, according to the distribution of data, statistical methods can be *i) parametric:* when making assumptions on the distribution of data (e.g., normal distributed); and *ii) non-parametric:* no requirements on probabilistic distribution, but still bounded by scale, amount of groups investigated, etc.

Meta-analysis is a statistical procedure that combines and summarizes the results of multiple individual studies.

Summary of findings. Our review identified 17 papers related to the application of statistical analysis in SE studies. Most of them (11) are related to experimental data collection, and just a few of them (2) also address survey-based research.

Table 3
Number and percentage of studies according to research methods and covered phases.

Method	Planning	Execution	Analysis	Reporting	All Phases
Case Study	60	47	40	33	18
Action Research	17	16	10	11	6
Design Science	2	2	2	0	0
Interviews	10	13	10	6	4
Focus Group	1	1	1	0	0
Observation	14	14	11	9	5
Ethnography	10	13	9	5	2
Think-aloud	1	2	1	0	0
Archival Research	4	8	9	5	3
Postmortem Review	2	2	2	1	1
Systematic Literature Review	34	54	33	23	16
Systematic Mapping	14	29	13	10	6
Survey	32	28	18	14	8
Delphi Method	1	1	1	0	0
Experiment	89	85	80	59	19
Quasi-Experiment	10	8	6	4	3
Simulation	5	1	0	2	0
Thematic Analysis	2	3	5	1	1
Content Analysis	2	3	3	0	0
Grounded Theory	3	4	7	1	1
Hermeneutics	2	2	2	0	0
Statistical Analysis	7	5	13	5	1
Meta-Analysis	2	3	10	1	0
Overall empirical research	27	17	9	19	4
Totals[a]	54.9%	60.2%	47.6%	31.9%	14%

[a] Note that several studies cover more than one method and/or activity, thus summing up to more than 100%.

4.2. Research Process' phases covered

Table 3shows the number of identified studies associated with a research method, as discussed in Section 4.1. Identified studies do not equally cover all the research phases, such as: *i* Planning; *ii* Execution; *iii* Analysis; and *iv* Reporting. These phases represent the conventional life cycle of typical empirical research; however, more specific discrete activities could be required according to the method used (e.g., to write a review protocol during the SLR planning).

Most of the papers address multiple phases, sometimes providing support for the entire research process. The number of papers that cover all the four phases is presented in the last column. 14% of the identified studies cover all the phases of the related process. Most of the papers address the execution (60.2%), planning (54.9%) and analysis (47.6%). Even few papers (31.9%) provide guidance to report the research.

Although those methodological papers are often not covering all the research phases, they can be particularly useful for researchers willing to conduct more complex studies i.e., combining research methods. On the one hand, some of those papers explain how different methods can be combined, using the results from previous studies to more in-depth research. On the other hand, others discuss the definitions of how research methods are distinct, helping researchers in selecting a more suitable setting for an intended research goal. A comprehensive list of studies addressing more than one research method is given in Appendix A.3.

4.3. Validated and evaluated guidelines

Some of the guidelines, assessment instruments, and knowledge organization systems have been validated (67 papers) or evaluated within its actual context (37 papers). Herein validation and evaluation research (see a glossary of terms in Appendix A.1) mean distinct degrees of maturity of an instrument according to empirical investigations employed to assess them.

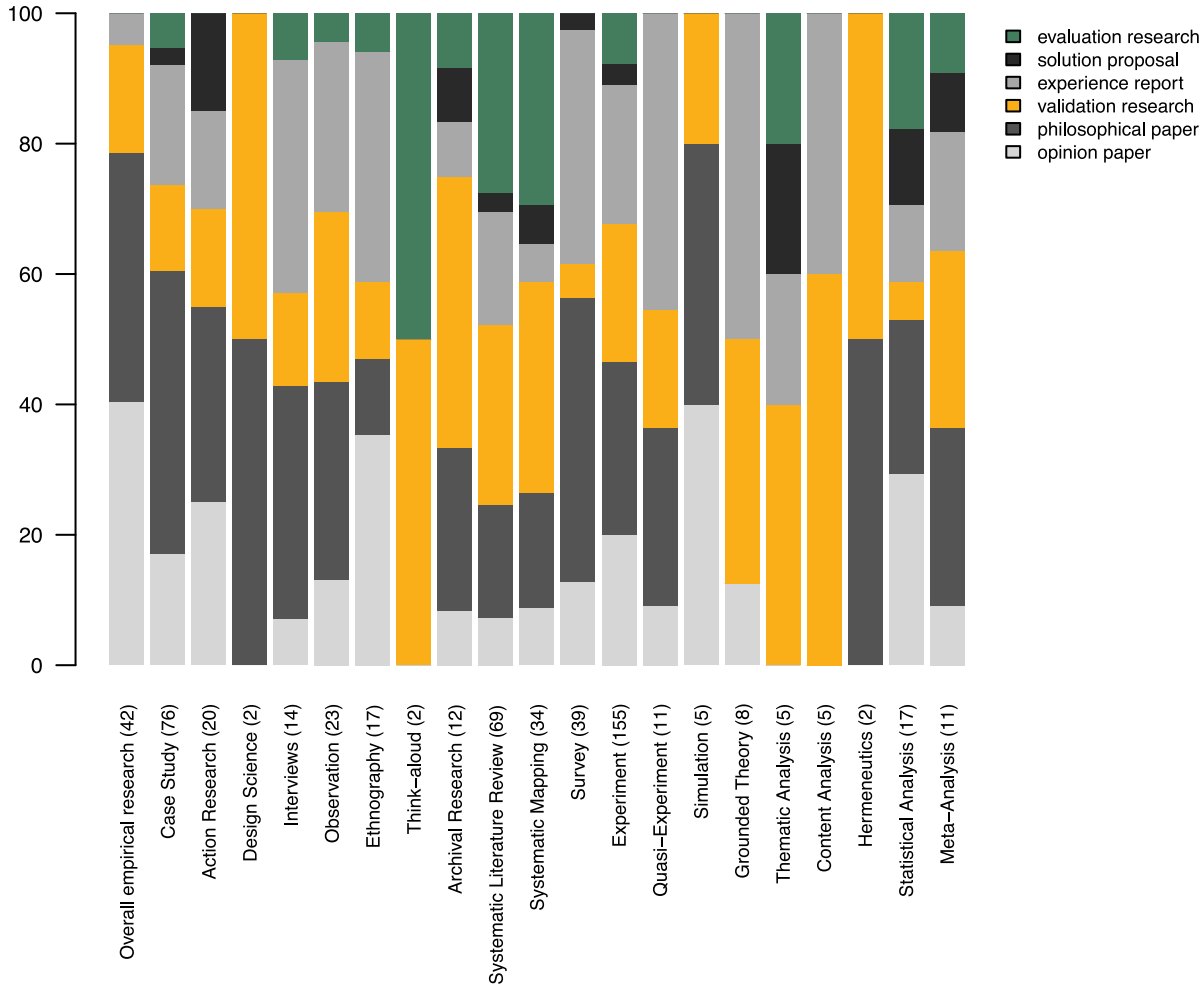


Fig. 4. 100% stacked column chart of identified papers according to maturity level. Each column is split into colored bar segments representing the relative contributions (i.e., in percentage) to the total bar. The total number of papers identified for each column is shown between parenthesis (e.g., there are a total of 76 case study related papers). Validated and evaluated papers are highlighted in orange and green, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Validation and evaluation are not equally distributed among all the research methods, as shown in Fig. 4. SLR and systematic mapping studies are the ones presenting more evaluation and validation (i.e., considering the total amount of papers), but also experiment and case study research have been evaluated in the SE domain. Several research methods are targeted of validation studies, but no evaluations, such as action research, grounded theory and content analysis. Finally, some methods (e.g., think-aloud, content analysis and hermeneutics) presented a relatively higher number of validated or evaluated papers, but this is mainly a result of few identified studies. Some of the highlighted references are given below:

Overall empirical research. Most of the literature aiming at the overall empirical research (i.e., rather than a particular research method) is in the early stages of maturity (i.e., opinion and philosophical papers). Due to the high level of abstraction of such philosophical papers, it is hard to evaluate their support for research and practice. Three distinct KOS [47,138,154] were validated through their application in a series of reported studies.

Case Study Research. three instruments supporting CSR were evaluated: *i*) DESMET has been subject to some evaluation exercises [182], even using the methodology as the evaluation instrument itself [173]; *ii*) a longitudinal case study methodology is both presented and evaluated by Laurie McLeod and MacDonell [224], and *iii*) a unified checklist for overall empirical research [338] has been evaluated by Condori-Fernandez *et al.* [72] for conducting case studies and experiments. Additional papers (e.g., [116,288]) also presented validation studies.

Action Research. We found no evidence of any evaluations. However, three papers addressing AR were validated, two of which are KOS [293,349] and another one presenting a comprehensive set of guidelines and an assessment checklist for case study and action research [288].

Interviews and Observations. No paper addressing those methods alone presented evaluations. However, guidelines addressing more general research methods, such as case study [224,288] and experiment [100], while employing interviews and observations were proposed and evaluated through empirical.

Secondary Studies. Kitchenham and Charters' SLR guidelines [174,195] were evaluated by comparison with other domains [54], by interviewing practitioners [27] and by a systematic review process itself [33]. The EBSE approach [196] was also validated [155,245,272,273]) and evaluated [274] by applying it with students.

Several activities of secondary studies (i.e., both SLRs and mapping studies) has been evaluated, consist of a plethora of best practices reported. The most investigated topic is the search strategy (e.g., [101,146,179,192,344,358]), covering the differences between manual search (and snowballing techniques) and database search (sometimes through automated mechanisms). Other topics include research question formulation [82], study selection [20,57,58,326], quality assessment [191], data extraction [326], data analysis [97], software tool

support [221,222], and repeatability [178].

Survey Research. No evaluation of guidelines and assessment instruments for survey research were identified. Just two papers addressing survey research were validated, both also covering case study and experiment research [116,349].

Experiments. According to Dieste *et al.* [98], the quality of experiments is related to the degree of use of methods that reduce bias. Several assessment instruments addressing different criteria for experiment were both proposed and evaluated [72,158,167,173,175,176,182,186]. Another metric for experiment's assessment is the replicability; which relates to two specifically evaluated guidelines [290,322].

Simulation. Identified literature regarding simulation studies showed no evaluation. However, an assessment instrument for simulation and experiments [353] was validated. The instrument aimed to improve simulation-based studies understandability, replicability, generalization, and validity.

Thematic Analysis and Grounded Theory. We found a single evaluated paper addressing the thematic synthesis in SE research [224]. None of the identified papers presented an evaluation of the GTM process in the SE context. However, two studies addressing different data analysis methods, such as thematic analysis and grounded theory (i.e., [78,129]) discuss the difficulties in applying such technique in SE studies.

Statistical Analysis. Kirsopp and Shepperd [172] evaluated the lack of accuracy of prediction statistics for effort estimation with very small samples. Moreover, both a guideline paper [224] and an assessment instrument [158] addressing experimental studies employing statistical analysis were also validated.

5. Discussion

The systematic mapping detailed herein provided evidence used to build a catalog of guidelines for empirical research in software engineering. This extensive set of references is structured according to the target research method and type of instrument described. To summarize the included studies, we also collected additional information related to the identified instruments, such as which research methods are covered and which phases are targeted, its main objectives (e.g., conduct, evaluate or classify the research) and the maturity level of the instrument.

5.1. Heuristics for guideline selection

The results of this mapping study are aimed to support researchers in identifying and selecting guidelines and other supporting literature (i.e., assessment instruments and knowledge organization systems) suitable for their research. The catalog presented herein is the first step towards achieving such aim. It provides a comprehensive set of methodological studies, classified according to their related research methods.

Researchers willing to use it effectively still need to retrieve a subset of the catalog and to form an opinion about which of guidelines are most suitable for their research. Therefore, we proposed a practical approach on how to find a subset based on the characteristics of the proposed research, as illustrated in Fig. 5, and described below: **1. Define a research proposal.** A research proposal should be founded on underlying theories, implemented through one or more related studies, aiming application or derivation of new theories. Before proceeding to the next phase, we suggest outlining the individual studies that compose the research proposal. Researchers willing to read about writing good research proposals can benefit from reference books e.g., [1,14].

2. Identify the research units. Those units include the research methods intended for each individual study. Note that the choice of a particular methodology affects the data collection and data analysis techniques. The decision point based structure proposed by Wohlin and Aurum [346] could support this particular step.

3. Select the suitable references. Search in the catalog (Appendix A.2) or the CERSE Web tool (http://www.cerse.org) for the section related to each of the research methods intended for use. Often, an extensive list of available guidelines is presented. It should be necessary to filter the available guidelines according to specific needs. Note the following heuristic approach to better choose the guidelines:

3.a. *Guidelines for conducting the research:* Prefer the guideline papers that address the whole process. If there is no such guideline, build a set of papers that covers all the intended phases.
3.b. *Context of the research:* As some research methods present several available guidelines (e.g., case study, survey, experiment), we suggest researchers to screen the list keeping in mind the particular context of the proposed research. Moreover, the discussion in Section 4.1 include some highlighted references that could be valuable.
3.c. *Multiple research methods:* In case that your project employ a set related studies, also browse the guidelines addressing those research methods together (Appendix A.3).
3.d. *Mature guidelines:* Give priority to evaluated (or otherwise validated) guidelines (as shown in Section 4.3).
3.e. *Assessment instruments:* Add to your references the assessment instruments or checklists to assure that the planning and reporting meets the quality requirements. Note that some guidelines already include quality assessment steps.
3.f. *Knowledge organization systems:* If necessary for planning or reporting, also read KOS to support the methodology categorization.
3.g. *Informed decision:* Finally, it is essential to assess the resulting references in order to decide which guidelines to use. Consider the limitations, strengths, and challenges of each instrument with regard to the proposed research. The highlights provided in Appendix A.2 are intended to support this step.

4. Aggregate the supporting literature. Combine the references into a joint list that could be easily consulted. Make sure that the guidelines are followed during the project, and any change is reported.

5. Reflect on the experience. Finally, it is desirable to conduct a critical analysis on the use of the selected guidelines and other supporting instruments. Such reflexion could be included in the study report, thus informing fellow researchers the lessons learned during the process.

Novice researchers are encouraged to follow the five-step strategy, making use of supporting references listed on the right hand side of Fig. 5 to design a research proposal. More experienced researchers (i.e., already used to the initial steps) would most likely only benefit from steps 3 and 4. Furthermore, researchers still need to make an informed decision on the relevant guidelines to use, as the subset is likely to contain overlapping or contextual-related guidelines.

5.2. Usage scenario for guideline selection

Based upon the heuristics abovementioned, we can derive a usage scenario rooted on an exemplar of ESE research proposal originally described in [346]. The research scenario is as follows:

Example. A Student wants to conduct research on technical debt, as this phenomenon has not been addressed enough in academic

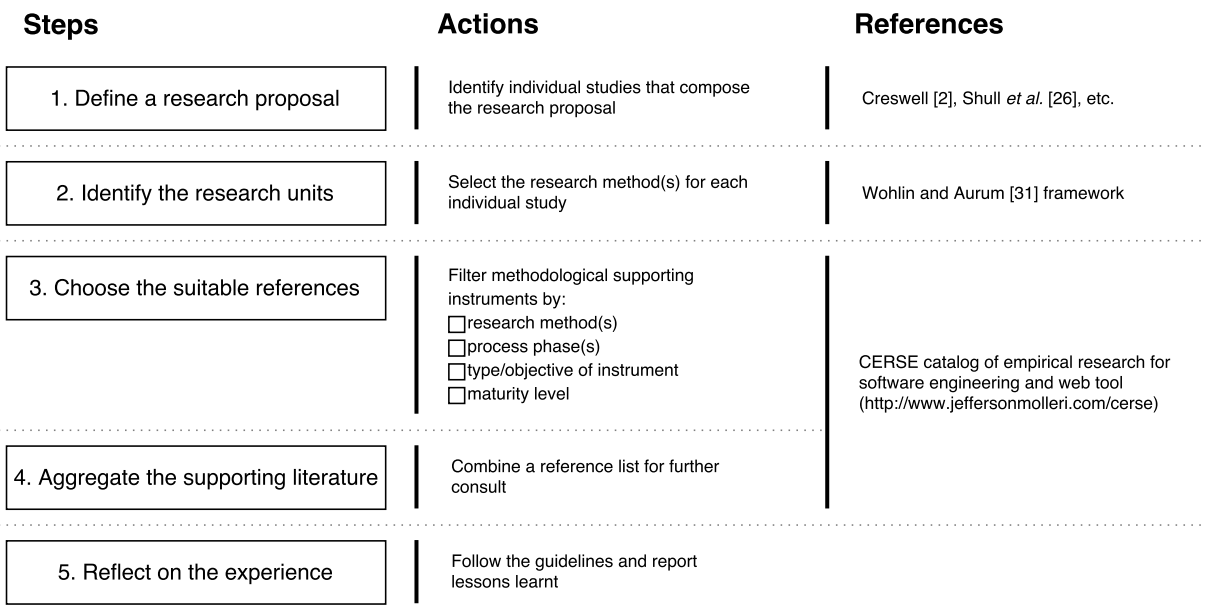


Fig. 5. Heuristics for guideline selection. Each step requires specific actions and can be linked to particular literature. The initial steps (i.e., 1 and 2) are addressed by related work, whereas the intermediate steps (3 and 4) are supported by the catalog of empirical research methods.

literature, even though practitioners have emphasized its importance. (...) The student believes that both academia and practitioners will benefit from having a framework that defines what constitutes technical debt, why it occurs and its detrimental implications in projects. Practitioners will be involved to evaluate the framework.

1. Define a research proposal. In the excerpt above, the Student describes the research idea, the intended contributions and the target group. In the first step of the proposed heuristics for guideline selection, the Student should also identify the individual studies needed to achieve the aims of the research. In this context it is important to explicitly highlight the methods to be used in the different studies. We already highlighted the important information needed to identify the research units.

Example. (...) Student conducts *a systematic literature (M1)*, using *thematic analysis (M2)* to identify to what extent technical debt is covered in academic literature. The outcome of the systematic literature is a framework that illustrates reasons and outcomes of technical debt. Then the student carries out *a multi-vocal literature review (M3)* to revise the framework. The *multi-vocal literature review (M3)* considers gray literature authored by practitioners, which is considered as a data-point of a *case study (M4)*. Both the systematic literature review and the multi-vocal literature review apply a qualitative research approach, (...) and use *thematic analysis as data analysis method (M2)*. The resulting framework is *evaluated by interviewing several practitioners (M5)*. The interviews are transcribed and *thematic analysis (M2)* is applied. Then, experts evaluate the framework and reviews are integrated to research findings.

2. Identify the research units. The excerpt above already mentioned several methods in the research proposal. Overall, the student conducts three different research units: (RU1) a preliminary literature review, followed by (RU2) a case study, and further (RU3) an interview study. According to the proposal, the following methods should be looked up in CERSE:

Table 4
Candidate references for the exemplar scenario.

Method	**Reference tables** (Appendix A.2)	Qty.
(M1) SLR	Section A.2.11	69
(M2) Thematic Analysis	Section A.2.18	5
(M3) Archival Research	Section A.2.9	12
(M4) Case Study	Section A.2.1	76
(M5) Interviews	Section A.2.4	14
Total		**152**[a]

[a] Note that some references cover more than one method, as shown in Appendix A.3.

- **RU1:** makes use of the methods *systematic literature reviews (M1)* and *thematic analysis (M2)*.
- **RU2:** makes use of the methods *"multi-vocal" literature review (M3)*, which uses the publications of practitioners as data for a qualitative data analysis (M2). The process for data collection described in the example can be considered as a *archival research method (M3)* following the definition by [346] which classifies the multi-vocal literature reviews as such. In summary, **RU2** employs the methods *case study (M4)* research, *archival research (M3)* and *thematic analysis (M2)*.
- **RU3:** uses *interviews (M5)* and also *thematic analysis (M2)*.

After having identified the relevant methods for the research units, the student should consult the catalog (see Appendix A.2 or the CERSE Web tool (http://www.cerse.org) and select the research methods that fit each individual unit.

3. Select the suitable references. Now, as a starting point, the student searches the catalog for five methods identified from the research proposal, and 152 references of supporting instruments are found. Table 4 maps the relevant parts of CERSE that should be consulted.

The Student should filter the results further in order to narrow down a set that is more appropriate to handle. The heuristic for the selection

Table 5
Selection criteria for the exemplar scenario.

Step	Criteria	Rationale
3.a.	Type of Instrument = *Guidelines* Process phases = *All*	The student's main aim is to identify instruments that help conduct the research according to the five identified methods. It is also important for the Student to identify guidelines that address all the phases of the process.
3.b.	Context of the research	Instruments that provide guidance to how to carry out practitioner-oriented research in the context of technical debt are preferred. Moreover, the research is mainly exploratory and qualitative by nature.
3.c.	Multiple research methods	Aiming to find guidelines that provide more complete methodological support to conduct the entire research proposal, the Student selects mainly guidelines that combine support to more than one method for the same unit of research.
3.d.	Maturity = *evaluated, validated*	If two or more guidelines addressing the same method are available, the Student should select one that has been evaluated or validated.
3.e 3.f.	Type of Instrument != *Ass. instruments, KOSs*	At this stage of the research, the Student is not interested in supporting instruments other than guidelines. Assessment instruments can be valuable in the future to assess if the research was conducted according to quality standards.
3.g.	Decision-making	Finally, the Student should critically read and assess the candidates to make an informed decision as to which guidelines to select. The Student also makes use of the highlights in Appendix A.2 to support this process.

of the relevant guidelines (see Section 5.1, step 3.a-3.g) are used to guide the selection. Table 5 details the selection criteria and the rationale for each step.

4. Aggregate the supporting literature. Later, the Student selects the included candidates, and aggregates them into a list (illustrated in Table 6) of supporting literature, as follows:

RU1: Considering the support to all the phases of the research process (*Heuristic 3.a*), the student could benefit from the guidelines for performing SLRs [195]. This paper is the basis for the student's literature review. Moreover, by checking the references employing both the SLR and thematic synthesis, the student could identify two additional studies [108,129] discussing a combination of those methods (*Heuristic 3.c*).

RU2: Case study research is covered by a long list of guidelines, thus making harder for the Student to decide. The guidelines for conducting and reporting case study research in SE [288] is a well established reference (*Heuristic 3.g.*). The paper presents recommendations for all the phases of the research process (*Heuristic 3.a.*) and was also validated (*Heuristic 3.d.*). Further, archival research as part of case study research is also supported by Runeson and Höst [288] (*Heuristic 3.c.*).

Alternatively, the DESMET approach [187] is mainly developed to promote a case study approach to evaluate SE methods and tools. However, as the evaluations proposed by DESMET are mainly quantitative, they do not match the context of the research, therefore excluded from the candidates list (*Heuristic 3.b.*).

RU3: Interviews with practitioners are planned to evaluate the proposed framework. Advice on how to conduct semi-structured interviews are given by Hove and Anda [141] (*Heuristic 3.a.*). Moreover, thematic analysis is also used to analyze the data collected during interviews. In that sense, additional support for interviews is provided by guidelines already in the student's list (Runeson and Höst [288], *Heuristic 3.a.*).

5. Moving forward. The Student should now become familiar with the selected guidelines and make sure that they are followed during the research process. Any changes to the process should be reported in order to allow its replication or validation. After conducting the research, the Student should also report the experiences with the chosen methods and selected guidelines, providing insights on their usage to the research community.

6. Conclusions

In this paper, we detailed existing guidelines, assessment instruments and knowledge organization systems to empirical research in SE. Via a systematic mapping 341 references from papers proposing such methodological support were gathered and selected. Some relevant findings were described in detail in Section 4.

6.1. Contributions of the mapping study

The mapping study's main contribution is to provide a list of guidelines, assessment instruments, and KOS supporting different research methods in ESE. The catalog is further provided in Appendix A.2. Furthermore, we summarize the results according to three different perspectives:

Research method. A comprehensive list of covered research methods and some highlighted references are further discussed in Section 4.1. Further information on papers covering more than one method is given in Appendix A.2. We also note that some of the methods are inter-related to another one, sharing some similar characteristics.

Phases of the research process. Most of the identified papers provide methodological support for more than one phase of the process. However, just a few of them (15%) covers all the phases together, as shown in Section 4.2. We suggest that researchers willing to employ such guidance carefully consider to aggregate a set of papers to bridge such gaps.

Evaluated instruments. Evaluation studies are important to assess the completeness and accuracy of such methodological supporting

Table 6
List of supporting literature for the exemplar scenario.

Res. Unit	Reference	Selection criteria	Method(s)
RU1	Guidelines for performing Systematic Literature Reviews in Software Engineering [195]	3.a, 3.g	M1
	Applying Systematic Reviews to Diverse Study Types: An Experience Report [108]	3.c	M1 + M2
	Survey on Research Synthesis in Software Engineering [129]	3.c, 3.d	M1 + M2
RU2	Guidelines for conducting and reporting case study research in software engineering [288]	3.a-3.d, 3.g	M3 + M4 + M5
RU3	Experiences from Conducting Semi-Structured Interviews in Empirical Software Engineering Research [141]	3.a, 3.g	M5
RU1-3	Recommended Steps for Thematic Synthesis in Software Engineering[77]	3.a, 3.g	M2

literature in ESE research. Our findings show that more mature research (i.e., evaluated guidelines and assessment instruments) are 10% of the identified studies. Moreover, there is a contrast in relation to the extension in which the papers addressing different research methods have been evaluated (also validated) in SE domain.

6.2. Implications for research

Software engineering researchers are the main recipients of this work. Foremost, the catalog provides a comprehensive collection of the research methods employed in ESE research. Although similar inventories have been provided in some other papers (e.g., [47,346,353,354]), our actual list is mapped to the guidelines, assessment instruments and knowledge organization systems reported in the SE literature.

Opportunities for Research. We encourage further research aiming the gaps identified in the Section 6.1 that are worthy of further investigation:

- *Research methods are not well-supported by specific guidelines for the software engineering domain e.g., design science, focus groups, think-aloud protocol, Delphi method.* By reflecting on how the guidelines from different domains could address SE research problems and conducting studies to validate such approaches, the community can improve the available guidelines with experience reports and lessons learned.
- *Guidelines and supporting instruments often do not cover all the research phases together.* We should be able to address this gap by aggregating and comparing the relevant guidelines and instruments related to a specific research method or context.
- *Supporting instruments are scarcely evaluated.* Results from evaluations can increase the reliability on using that methodological guidance, and identify improvement needs. By improving the guidelines, the research community can also increase the quality of empirical studies conducted using such guidelines as basis [318]. We also noted the need to conduct evaluation studies, especially on the research methods still in the early stages of maturity.

Learning Aspects. Moreover, the catalog can be employed by students and novice researchers learning how to conduct empirical studies. By mapping the catalog references to the particular units of a research proposal, researchers can select a subset of the guidelines fitting their intended research processes. A strategy to this is described in Section 5.1 and exemplified through a usage scenario in Section 5.2. The heuristics alone are not intended to guide novices through the study design phase, but instead can be employed together with other pedagogical instruments (e.g., [13,346]) for learning or teaching purposes.

6.3. Future work

As future work, we plan to update the catalog periodically by applying the search strategy on the most recent journal editions and conferences proceedings, as well as also conducting further snowballing iterations. The updates should be made available through the Web tool's database. We also aim to conduct an empirical investigation to validate the proposed catalog and the related heuristics for guidelines selection.

In addition to that, we intent to conduct further studies aiming the relation between the methodological support and the empirical research in SE. In particular, we aim to investigate whether the availability of a specific SE-related methodology is aligned to the quantity and quality of primary studies using such method.

Contributors

All authors contributed to conceiving the idea and planning the research. J.M. conducted the search process and collect data from the candidate papers. All authors took part in the study selection process. K.P. and E.M. also verified the data extraction and synthesis of included studies. Finally, all authors discussed the findings and contributed to the final manuscript.

Acknowledgements

The work of Jefferson Seide Molléri is supported by the Science Without Borders program, funded by CNPq (National Council for Scientific and Technological Development - Brazil).

Appendix A

A1. Glossary

A1.1. Maturity levels / score

Evaluation study. An empirical research that investigates the implementation of a guideline or supporting instrument in real life setting e.g., with actual researchers and/or using the research projects conducted by them. Evaluated research is the highest stage of maturity according to the classification of research facets [343].
Validation study. An empirical study aiming to validate a guideline or supporting instrument in a laboratory setting, or otherwise using toy examples. Validation research is weaker than evaluation studies, according to the classification of research facets [343].

A1.2. Classification of supporting instruments

Assessment instrument. Methodological tools to ensure that the research process is followed according to particular evaluation criteria. Examples of such tools include checklists, rubrics and assessment metrics [107,318]. Some assessment instruments rather aim at the reported study or the guidelines supporting the research.
Guidelines. A step-by-step structure that details the process of conducting research according to a particular methodology [17,318]. It is often described as regarding the process phases covered and their underlying actions. In this work, we also relate the guidelines to frameworks and tutorials that provide guidance to how the research should be executed. Some guidelines also address how to teach research, and how to replicate the research.
Knowledge organization system. A semantic structure that formally describes the research process via vocabulary control and knowledge organization [16]. Examples of KOs are taxonomies, ontologies and classification schemes.

A2. Catalog of empirical research

A2.1. Case study

	Guidelines	Ass. Instruments	KOS
Planning	[38,51,70,188,197,220,260,341,352]		[70,104,117,260,341]
Execution	[50,268,304]		
Analysis	[79,113,246,270,327]	[113]	
Reporting	[259,262,342]	[173]	
Planning + Execution	[151,185,190,213,215,234,269,278,292,332]		
Planning + Analysis	[297]		[353]
Planning + Execution + Analysis	[29,62,63,118,189,199,203,206,309,310,317,338]	[338]	
Planning + Execution + Reporting	[64]	[349]	
Execution + Analysis + Reporting		[258]	
Analysis + Reporting	[204]	[296]	
All Phases	[23,41,68,83,115,116,139,149,202,205,224,249,255,280,288,313,318]	[41,139,182,202,288]	[41,116]

Highlights

1. *Guidelines for case study research*: Several guidelines are intended to guide researchers through all phases of case study research. They often include recommended practices to report and review such studies. Those instruments include guidelines and tutorials for conducting CSR in SE (e.g., [139,255,288]) and IS [41].
   - *Quantitative case studies:* The DESMET series [5] propose a methodology for evaluating software engineering methods and tools through quantitative data collection and feature analysis. Particular advice includes selecting host projects, finding proper metrics, minimizing confounding factors, monitoring data collection, and quantitative analysis.
   - *Mixed case studies:* A unified checklist [72,338] proposes a research design based on the relationship between case studies and experimental research methods.
   - *Triangulation:* The use of multiple data sources as a means to identify more and more trustful findings is discussed by Bratthall and Jørgensen [50].
   - *Replication:* Some lessons learned on the replication of case study research are given in [234,278]. Several levels of replication are categorized and discussed by Ohlsson and Runeson [246].
2. *Assessment instruments:* Several criteria can be used to evaluate CSR, such as design and report phases [72,288], validity threats [113,258] and generalization of claims [296].
3. *KOS:* The main characteristics of case studies and experiments are presented by [260]. Some additional schemes propose the categorization based on the differences between positivist and interpretative approaches [104].

A2.2. Action research

	Guidelines	Ass. Instruments	KOS
Planning	[38,70]		[70]
Reporting		[293]	[84]
Planning + Execution	[26,212,289]		[215]
Planning + Execution + Analysis	[40,223,238]		
Planning + Execution + Reporting	[211]		[349]
Execution + Analysis + Reporting		[258]	
All Phases	[39,41,115,288,313,318]	[41,288]	[41]

Highlights

1. *Guidelines for action research in IS:* Our review identified no instrument to conduct action research according to particular software engineering needs. Most of the guidelines cited in SE domain come from information systems (e.g., [39,84,115,212,238]).
   - *Action and case studies:* Action research is sometimes defined as action case studies, as it shares several characteristics with case studies. Some papers present guidelines along with case study research (e.g., [41,70]). One particular article discusses the contrasts between them [38].
2. *Assessment instruments:* A set of five principles and associated metrics to assess the quality (rigor and relevance) of canonical action research practice is proposed by Davison, Martinsons & Kock [84].
3. *KOS:* A literature survey identified 16 papers reporting action research in SE journals and classified them according to the study adherence, type, and control structures [293].

A2.3. Design science

	Guidelines	Ass. Instruments	KOS
Planning + Execution + Analysis	[334,337]		

Highlights

1. *Very few guidelines for SE research:* Although the studies cover the main process phases (including the evaluation of the design artifact) there is no specific support on how to report such studies.

A2.4. Interviews

	Guidelines	Ass. Instruments	KOS
Planning + Execution	[241,252,269]		[215]
Planning + Execution + Analysis	[111,309]		
Execution + Analysis	[214,295]		
Execution + Analysis + Reporting		[258]	
All Phases	[141,224,288,313]	[288]	

Highlights

1. *Guidelines for interviews:* A comprehensive guide and a set of advice to plan and conduct semi-structured interviews is given in [141]. Its activities include: *i*) scheduling; *ii*) collecting of background information; *iii*) preparing interview guides; *iv*) discussion/meetings; *v*) summary writing; and *vi*) transcribing. The guidelines aggregate existing knowledge from other disciplines and complement with authors' own interviewing experiences.
   - *Field studies:* Some papers address interviews as part of field study techniques, thus providing guidelines for conducting interviews along with observations and surveys [214,252].
   - *Ethical issues:* The application of interviews often face ethical dilemmas (e.g., consent, confidentiality, and beneficence). Singer and Vinson [313] identify ethical issues relevant to empirical studies, exemplified by real SE empirical studies.
   - *Globally distributed:* Interviews can become more complex when interviewer and interviewee are not meeting face-to-face. Papers discussing such challenges and solutions are related to field studies [252] and qualitative methods [269,309].
2. *KOS:* A taxonomy of field studies is given in the following [214,215], discussing the advantages and disadvantages of each method (i.e., interviews are classified as direct and inquisitive, are also highly interactive but time-consuming).

A2.5. Focus group

	Guidelines	Ass.Instruments	KOS
Planning + Execution + Analysis	[208]		

Highlights

1. *Guidelines for focus groups:* This method involves well-planned discussions to gather the perceptions of a group of participants [208]. The researcher acts as a mediator, facilitating and focusing the discussion on the topic.

A2.6. Observation

	Guidelines	Ass. Instruments	KOS
Planning	[233,325,329]		[329]
Execution	[251,285]		
Analysis	[113]	[113]	
Reporting	[61]	[339]	
Planning + Execution	[213,252]		[215]
Planning + Analysis			[353]
Planning + Execution + Analysis	[309]		
Planning + Execution + Reporting			[349]
Execution + Analysis	[133,214,295]		
Analysis + Reporting	[100]		
All Phases	[202,224,288,313]	[202,288]	

Highlights

1. *Guidelines for observational studies:* None of the identified studies presented a guide to observations alone. Most of the guidelines incorporate it to case study research e.g., [224,288,313].
   - *Reporting observational studies:* A set of guidelines for reporting the outcomes of observations in the form of lessons learned is provided by Budgen and Zhang [61]. Additional guidance is proposed for observations along case studies [288] and experiments [339].
   - *Qualitative data analysis:* It is often employed together with observations, such as grounded theory (e.g., [113,133]) and thematic analysis (e.g., [224]).
   - *Think-aloud protocol:* It is mainly used for testing user interfaces, in which the method was subject to an evaluation [243]. An alternative called dialog-based protocol can be used in pair programming tasks, as proposed and validated by Xu and Rajlich [351].
2. *Assessment instruments:* A checklist for assessing observation and experiment research reports is given by Wieringa *et al.* [339].
3. *KOS:* Field study techniques can be categorized according to the human intervention degree [215] (i.e., observation is classified as direct or inquisitive technique).

A2.7. Ethnography

	Guidelines	Ass. Instruments	KOS
Planning	[70,233]		[70]
Execution	[74,169,251,268,285]		
Planning + Execution	[289]		
Planning + Execution + Analysis	[108,206,279,309]		
Planning + Analysis + Reporting	[284]		
Execution + Analysis + Reporting	[282]		
Analysis + Reporting	[333]		
All Phases	[240,283]		

Highlights

1. *Guidelines for ethnography:* An overview of ethnography studies in SE is given by Rönkkö [283] and some guidelines are provided in the following references [240,251,279].

A2.8. Think-Aloud

	Guidelines	Ass. Instruments	KOS
Planning + Execution	[243]		
Execution + Analysis	[351]		

Highlights

1. *Guidelines for think-aloud protocol:* It is mainly used for testing user interfaces, in which the method was subject to an evaluation [243]. An alternative called dialog-based protocol can be used in pair programming tasks, as proposed and validated by Xu and Rajlich [351].

A2.9. Archival research

	Guidelines	Ass. Instruments	KOS
Execution	[74,304]		
Analysis	[75]	[216]	
Planning + Execution			[215]
Execution + Analysis	[214]	[127]	
Analysis + Reporting	[100,306]		
All Phases	[224,280,288]	[288]	

Highlights

1. *Guidelines for archival research:* We found no step-by-step guide supporting the process in SE research. The need for such guidelines is supported by several examples of SE-related archives that can be addressed by such studies (e.g., software repositories, project artifacts, and models).
   - *Analysis of work artifacts:* Several data collection techniques for archival research (e.g., electronic databases of work personal, tool logs, and documentation) are discussed by Lethbridge, Sim, and Singer [214,215]. Their advantages and disadvantages are discussed, and some advice on how to keep the records are given.
   - *Replicating data mining:* Studies based on data retrieved from development repositories are especially suitable for reproduction. Best practices, tools, and recommendations to replicate empirical studies on mining software repositories (MSR) are given by Gonzalez-Barahona and Robles [127] and Robles and GermÃ!'n [280].
   - *Data collection for content analysis:* Archival research can provide a systematical approach to collect data for content analysis techniques.

Papers such as [74,100,216] address such joint approach.
2. *Assessment instruments:* A method for assessing the reproducibility of studies based on data retrieved from development repositories is proposed by Gonzalez–Barahona and Robles [127].
3. *KOS:* Also archival research figure the taxonomy of field study techniques [215] (i.e., classified as indirect and unobtrusive).

A2.10. Postmortem review

	Guidelines	Ass. Instruments	KOS
Planning + Execution + Analysis	[102]		
All Phases	[149]		

Highlights

1. *Guidelines for postmortem reviews:* Three different methods for carrying out such reviews are provided in [102]. Another work [149] provide lessons learned on aggregating different studies, including postmortem reviews, controlled experiments and case studies.

A2.11. Systematic literature review

	Guidelines	Ass. Instruments	KOS
Planning	[329,341,356]		[329,341]
Execution	[20,60,99,101,123,178,192,256,32-0,326,328,330,344,357,358]	[28,58,71,81,99,107,144,1-46,179,191,192,357]	
Analysis	[75,76,97,129,277,345]		
Reporting	[57,89,335]	[55,57]	[89]
Planning + Execution	[347,355]		
Planning + Analysis			[353]
Planning + Reporting			[299,354]
Planning + Execution + Analysis	[33,67,108,155,221,272,276]	[33,136,222,272]	
Planning + Execution + Reporting		[82]	
Execution + Analysis		[219]	
All Phases	[44,52,54,86,143,174,195,196,209,-228,245,273,274,318,324]	[27,195,273,274]	[86]

Highlights

1. *Guidelines for systematic literature reviews:* The most influential SLR guidelines in SE were proposed by Kitchenham and Charters [174,195]. Another template based on SLR protocols developed in medical area is described by Biolchini *et al.* [44,86,228].
   - *Supporting the intrinsic activities:* Secondary studies follow a systematic process that requires several phases and discrete activities. Specific support for such phases include: *i*) search strategies e.g., [99,101,192,320,344,355,358]; *ii*) primary study selection e.g., [20,256,328]; *iii*) quality assessment e.g., [144,192]; *iv*) data extraction e.g., [330]; *v*) data synthesis e.g., [75,97,129]; *vi*) reporting the review e.g., [277,335,345]; and *vii*) replication e.g., [89,178].
   - *Evidence-based software engineering (EBSE):* SE research might benefit from an approach for integrating the results from secondary studies with practitioners needs and values. Such approach is proposed by Kitchenham, Dybå and Jørgensen [196]. Some guidelines address the pedagogical aspects of using EBSE by students [155,245,272,274].
   - *Tools to support SLRs:* such as SLuRp, StArt, SLR-Tool and SLRTOOL are identified and further analyzed by Marshall, Brereton and Kitchenham [221]. A set of features that such tools should possess has been developed and used as the criteria to evaluate the candidate tools [222].
2. *Assessment instruments:* The systematic process employed in secondary studies require particular validation for some of its phases, such as: *i*) defining the research question e.g., [82]; *ii*) search strategies e.g., [28,146,179,192,355]; *iii*) quality assessment e.g., [191]; and *iv*) reporting the review e.g., [55,57]. The particular characteristics and benefits of mapping studies have been assessed in a series of studies e.g., [59,177,193,194].
3. *KOS:* A scientific research ontology provide support to describe the knowledge regarding SLRs studies [86].

A2.12. Systematic mapping

	Guidelines	Ass. Instruments	KOS
Planning	[341]		[341]
Execution	[99,101,123,192,256,320,326,344,-357,358]	[28,71,81,99,146,177,179,-192,357]	
Analysis	[345]		
Planning + Execution	[355]		
Planning + Analysis			[353]
Planning + Reporting			[299,354]
Planning + Execution + Analysis	[33,257]	[33]	

Planning + Execution + Reporting		[82]
Execution + Analysis	[193]	[59]
Analysis + Reporting		[194]
All Phases	[52,143,174,195,273,318]	[195,273]

Highlights

1. *Guidelines for systematic mapping studies:* Even though general guidance for secondary studies is given in Section 4.1.7, a comparative study investigated the differences between SLR and mapping studies. Based on the results, a specific set of guidelines for systematic maps is proposed by Petersen *et al.* [257]. Benefits and challenges of such studies have also been investigated by Kitchenham, Bugden & Brereton [193].
2. *Assessment instruments:* Some measures of quality (e.g., completeness, reliability, effectiveness) for systematic mapping studies were proposed and validated in the given papers [59,177,194].

A2.13. Survey

	Guidelines	Ass. Instruments	KOS
Planning	[183,197,200,263,264]		[117]
Execution	[268]		
Analysis	[91,184]	[198]	
Reporting	[262]		
Planning + Execution	[73,94,153,190,201,252,269]	[201]	
Planning + Reporting			[187,299]
Planning + Execution + Analysis	[199,309]	[29]	
Planning + Execution + Reporting	[64]		[349]
Execution + Analysis	[69,214]		
Execution + Analysis + Reporting		[258]	
All Phases	[68,115,116,170,202,266,313,318]	[202]	[116,266]

Highlights

1. *Guidelines for survey research:* Two general guidelines provide support for the whole process. The principles of survey research series [6] is a starting point for novice researchers to understand the process and its main challenges. Another seven-stage framework [170] present similar guidance based on lessons learned and include additional resources, such as the activities breakdown and a cover letter. Furthermore, a set of recommendations to replicate survey-based research is given by Cater-Steel *et al.* [69].
   - *Field studies:* Some papers discuss guidance for the survey process along with other field study methods e.g., [115,214,252,313,349].
   - *Qualitative and quantitative surveys:* As mixed research, surveys usually collect both qualitative and quantitative data. However, several guidelines favor one type over another, such as [64,268,269,309] emphasize the perspectives of the subjects, whereas [91,187,189,199] address quantified data collection and analysis.
   - *On-line surveys:* Recommendations for on-line self-administrated surveys are given by Punter *et al.* [271]. The paper covers issues such as tracking responses in real-time, motivating respondents to return, and identifying drop-out questions i.e., where people stop filling out the questionnaire.
   - *Sampling strategies:* A proper sampling strategy is a key to generalizing the findings from a survey questionnaire. Such issues are discussed in part five of principles of survey research series [183]. Moreover, there is a set of papers [92–94] investigating proposing improvements for large-scale sampling in surveys.
2. *Assessment instruments:* The survey questionnaire is often the target of evaluation regarding validity and reliability aspects [170,201]. A comprehensive checklist provides support for identifying relevant threats to several research methods, including surveys [258].
3. *KOS:* A framework for classifying and examining survey research is proposed by Pinsonneault & Kraemer [86]. Some other categorization instruments cover several empirical research methods, distinguishing surveys' particular characteristics [116,299,349].

A2.14. Delphi method

	Guidelines	Ass. Instruments	KOS
Planning + Execution + Analysis	[247]		

Highlights

1. *Guidelines for delphi method:* Guidelines for selecting appropriate experts and ensure the validity of the study in IS are given by Okoli and Pawlowski [247].

A2.15. Experiment

	Guidelines	Ass. Instruments	KOS
Planning	[156,157,181,197,226,233,253,260-,302,325,329,341]	[159,250]	[88,117,159,171,260,329,-341]
Execution	[36,37,128,135,140,167,268,304,3-31]	[65,167,168,186]	[140]
Analysis	[22,91,95,113,121,160,231,232,23-5,246,265,287,291]	[95,113,131,198]	
Reporting	[30,66,89,90,126,145,150,152,262,-275,294,305,311,342]	[173,175,176,323,339]	[30,89,90,126,319]
Planning + Execution	[24,25,43,56,92,94,112,151,190,20-7,242,244,301,308,315,316]	[87,308]	
Planning + Analysis	[290,340]		[353]
Planning + Reporting	[300,303]	[72,300]	[187,299,300,354]
Planning + Execution + Analysis	[32,34,35,46,53,106,114,118,122,1-47,161,162,180,189,199,218,237,3-10,317,336,338]	[29,98,338]	
Planning + Execution + Reporting	[64]		[349]
Planning + Analysis + Reporting		[163]	[132]
Execution + Analysis	[48,80,125,225,230,236,295,322]	[48,127,322]	
Execution + Reporting	[307]		[307]
Execution + Analysis + Reporting		[158,258]	
Analysis + Reporting	[100,105,239,348]	[109,165,296]	
All Phases	[21,68,116,124,148,149,166,196,2-02,210,217,227,229,261,280,313,3-18]	[21,124,148,182,202]	[116,120]

Highlights

1. *Guidelines for experiments:* A set of activities for performing experiment studies in SE are discussed in the following references: [32,34,53,148,261]. Additional guidance include meta-analysis techniques (e.g., [118,229]) and replicating experiments (e.g., [217,227]). Some guidelines focus on the pedagogical aspects of teaching or/and learning how to conduct experiment studies [74,336].
   - *Human-based experiments:* SE often study human subjects i.e., practitioners, students, researchers. Several papers cover sampling and data collection strategies (e.g., [92,268]), human subjects-related bias (e.g., [237]), ethical issues (e.g., [24,313]), knowledge transfer (e.g., [301,305]), and replication (e.g., [218]).
   - *Technology-based experiments:* Are also known as computational experiments and benchmarks i.e., studies comparing the performance of computer systems. A series of experience reports provide guidance in performing such studies, such as [46,106,210,310]. Additional papers covers the reporting (e.g., [145]) and replication (e.g., [127]) aspects.
2. *Assessment instruments:* Several studies present evaluation of experimental research, addressing aspects such as continuous feedback [168], external and internal validity [65,98,131,250], and experimental reporting [165,339].
3. *KOS:* Categorization of experiments also cover different features, such as context [171], theory usage [132], subjects, tasks and environments [120,140,319], and type of experimental replications [30,88,126,307].

A2.16. Quasi-Experiment

	Guidelines	Ass. Instruments	KOS
Planning	[329]	[159]	[159,329]
Analysis	[91]		
Planning + Execution	[56,94]		
Planning + Execution + Analysis	[35,180]		
Planning + Execution + Reporting	[64]		
All Phases	[68,124,166]	[124]	

Highlights

1. *Guidelines for quasi-experiments:* A.k.a. natural experiments are non-randomized experiments often conducted in a field setting. An overview of how such studies are designed and analyzed, the threats to validity and results are reported, is presented by Kampenes *et al.* [166]. Moreover, lessons learned on designing, conducting and analyzing data from quasi-experiments are reported in the following references [56,180].

A2.17. Simulation

	Guidelines	Ass. Instruments	KOS
Planning	[341]		[117,341]
Planning + Execution		[87]	
Planning + Reporting	[248]		[354]

Highlights

1. *Guidelines for simulation studies:* A secondary study characterizing the SBS initiatives in SE resulted in a set of guidelines [87]. The guidelines aim to improve such studies' understandability, replicability, generalization and validity.
2. *Assessment instruments:* A framework is proposed to assess the acceptability of the components in a simulation study (e.g., parametric model, real-world and simulated data, simulation results, experiment specification, and methodologies or techniques in use) [248].

A2.18. Thematic analysis

	Guidelines	Ass. Instruments	KOS
Analysis	[129]	[78]	[78]
Planning + Execution + Analysis	[108]		
Execution + Analysis	[77]		
All Phases	[224]		

Highlights

1. *Guidelines for thematic analysis:* Our systematic process identified a proposal for thematic synthesis process in SE through a checklist [77]. The process and outcome associated with each step are described and illustrated with examples from the literature.
2. *KOS:* An overview of the data analysis methods and their related characteristics contributes to a better understanding of the challenges in synthesizing SE research [78].

A2.19. Content analysis

	Guidelines	Ass. Instruments	KOS
Execution	[74]		
Analysis		[78,216]	[78]
Planning + Execution	[269]		
Planning + Execution + Analysis	[108]		

Highlights

1. *Guidelines for content analysis:* Examples on the application of qualitative content analysis are presented by the following references [78,216,269].

A2.20. Grounded theory

	Guidelines	Ass. Instruments	KOS
Analysis	[45,113,129]	[78,113]	[78]
Planning + Execution	[142]		
Planning + Execution + Analysis	[19]		
Execution + Analysis	[133]		
All Phases	[18]		

Highlights

1. *Guidelines for grounded theory:* A model including fifteen recommendations is presented by Adolph, Hall and Kruchten [19] based on their experience using classical (i.e., Glaserian) grounded theory in the SE context. It is intended to help researchers interpret the canons of grounded theory in a manner that is relevant to software engineers.
   - *Lessons learned:* Most of the identified papers are experience reports on conducting grounded theory e.g., [18,19,133,142]. These studies discuss insights when conducting such research, in addition to examples of its application in SE research.

A2.21. Hermeneutics

	Guidelines	Ass. Instruments	KOS
Planning + Execution + Analysis	[62,206]		

Highlights

1. *Very few guidelines for hermeneutics:* We identified just two studies (i.e. [62,206]) presenting an overview of the method. Both papers outline the application through interpretive studies on the information systems field.

A2.22. Statistical analysis

	Guidelines	Ass. Instruments	KOS
Planning	[156,159]		[159]
Analysis	[96,121,137,172,232,286]	[96]	
Planning + Execution	[56]		
Planning + Analysis	[340]		
Planning + Reporting			[299]
Planning + Analysis + Reporting		[163]	
Execution + Analysis	[69,230]		
Execution + Analysis + Reporting		[158]	
Analysis + Reporting		[109]	
All Phases	[224]		

Highlights

1. *Guidelines for statistical analysis:* Statistical analysis includes an extensive set of techniques, and just some of them are addressed by guidelines in the particular context of SE research. As an example, a 10-step framework for statistical prediction of software characteristics is proposed by Rosenberg [286].
   - *Descriptive statistics:* Visual techniques for statistical analysis in empirical SE studies are proposed by Garcia *et al.* [121]. A case study illustrates how such visualizations may improve interpretation of the experimental data.
   - *Inferential statistics:* Context and generalization problems potentially resulting from statistical hypothesis testing approaches are discussed by Jørgensen and Sjøberg [156]. The paper attempt to formulate how we should conduct empirical software engineering studies.
2. *Assessment instruments:* Statistical significance, statistical power, effect size and magnitude plays important roles in hypotheses testing and therefore should be carefully handled. Some papers investigate the assessment of such pivotal characteristics in ESE studies e.g., [56,109,137,163,232].

A2.23. Meta-Analysis

	Guidelines	Ass. Instruments	KOS
Planning			[159]
Analysis	[91,95–97,103,129,137]	[95,96]	
Planning + Reporting			[299]
Planning + Execution + Analysis	[118]		
Execution + Analysis	[219,230]		

Highlights

1. *Meta-analysis:* Meta-analytic techniques applied to SE studies are illustrated by Hayes [137]. Moreover, Dieste *et al.* [97] provide a set of rules to help to choose the proper aggregation technique. Finally, for applying such techniques, a heterogeneity analysis of the studies' results is required; Dieste *et al.* [95] validates how accurately the statistical methods detect heterogeneity.

A2.24. Overall empirical research

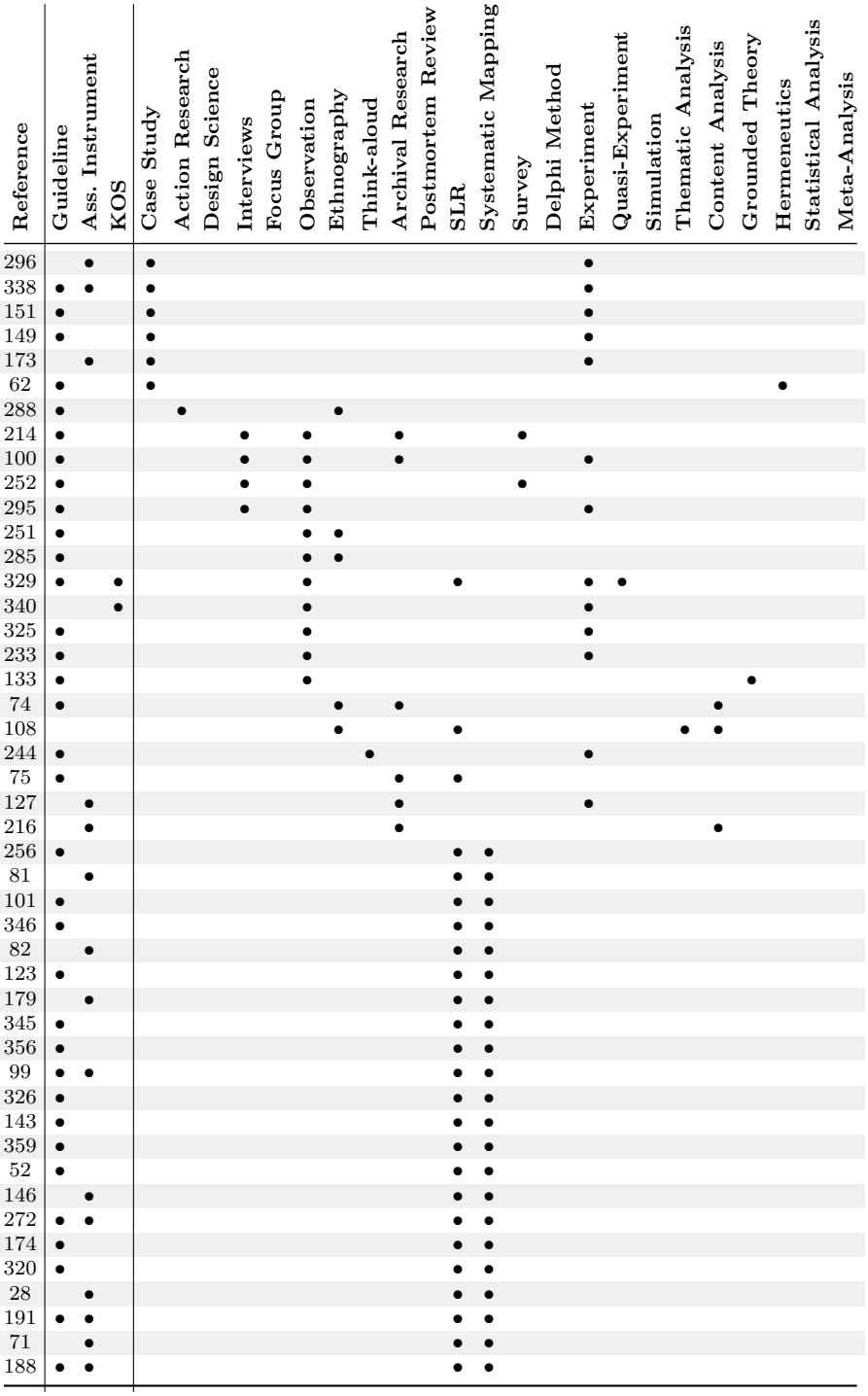
	Guidelines	Ass. Instruments	KOS
Planning	[42,156,233,312,346,350]	[298]	[47,117,346]
Execution	[135,314]	[154]	[49,154,343]
Analysis	[113,231]	[113]	
Reporting	[31,259,267,275]		[134,138,281,321]
Planning + Execution	[24,110,130,254,308]	[308]	
Planning + Analysis			[353]
Planning + Reporting	[300]	[300]	[85,299,300,354]
Planning + Execution + Analysis	[164]		
Planning + Execution + Reporting	[64,119]		
Planning + Analysis + Reporting	[284]		
All Phases	[68,202,209,318]	[202]	

Highlights

1. *Guidelines for empirical research*: Some papers aim to help at identifying the appropriate methods and discuss the implications of this choice to the furure research e.g., [85,117,135,202,346].
   - *Describing a context:* In order to decide whether a piece of evidence is helpful for a given situation, a complete and accurate description of the context which the study was conducted is needed. Papers supporting the description and assessment of the research context include the following references [110,259].
2. *KOS:* Ontologies, taxonomies and glossaries help researchers in describe the research, as also to better communicate its findings. Some literature presenting such tools include [117,138,281,353]. Moreover, Happel and Seedorf [134] present a framework to classify and apply ontologies.
3. *Assessment instruments:* A proper validity analysis and proposal of strategies to mitigate the impact of threats are essential to ensure the quality of empirical research. Some example of such literature is [113,135]. Additionally, ethical issues of SE empirical research are discussed by the following [24,130,312].

A3. Papers covering multiple research methods

Reference	Guideline	Ass. Instrument	KOS	Case Study	Action Research	Design Science	Interviews	Focus Group	Observation	Ethnography	Think-aloud	Archival Research	Postmortem Review	SLR	Systematic Mapping	Survey	Delphi Method	Experiment	Quasi-Experiment	Simulation	Thematic Analysis	Content Analysis	Grounded Theory	Hermeneutics	Statistical Analysis	Meta-Analysis
289	●	●		●	●		●		●			●														
215			●	●	●		●		●			●														
313	●			●	●		●		●							●		●								
258		●		●	●		●									●		●								
350			●	●	●				●							●		●								
70	●		●	●	●					●																
318	●			●	●									●	●	●		●								
115	●			●	●											●										
41	●	●	●	●	●																					
38	●			●	●																					
309	●			●			●		●	●						●										
224	●			●			●		●			●									●				●	
270	●			●			●									●						●				
269	●			●					●	●								●								
354			●	●					●					●	●			●								
196	●	●		●					●							●		●								
113		●	●	●					●									●					●			
303	●			●					●									●								
213	●			●					●																	
206	●			●						●														●		
280	●			●								●						●								
304	●			●								●						●								
299			●	●										●	●	●		●							●	●
342	●		●	●										●	●			●		●						
355			●	●										●	●			●		●						
64	●			●												●		●	●							
68	●			●												●		●	●							
117			●	●												●		●		●						
200	●			●												●		●								
262	●			●												●		●								
198	●			●												●		●								
201	●			●												●		●								
29		●		●												●		●								
116	●		●	●												●		●								
197			●	●												●		●								
193	●			●												●		●								
118	●			●														●								●
260	●		●	●														●								
317	●			●														●								
72		●		●														●								
310	●			●														●								
246	●			●														●								
343	●			●														●								
182	●			●														●								


Reference
Guideline
Ass. Instrument
KOS
Case Study
Action Research
Design Science
Interviews
Focus Group
Observation
Ethnography
Think-aloud
Archival Research
Postmortem Review
SLR
Systematic Mapping
Survey
Delphi Method
Experiment
Quasi-Experiment
Simulation
Thematic Analysis
Content Analysis
Grounded Theory
Hermeneutics
Statistical Analysis
Meta-Analysis
296
338
151
149
173
62
288
214
100
252
295
251
285
329
340
325
233
133
74
108
244
75
127
216
256
81
101
346
82
123
179
345
356
99
326
143
359
52
146
272
174
320
28
191
71
188

Reference	Guideline	Ass. Instrument	KOS	Case Study	Action Research	Design Science	Interviews	Focus Group	Observation	Ethnography	Think-aloud	Archival Research	Postmortem Review	SLR	Systematic Mapping	Survey	Delphi Method	Experiment	Quasi-Experiment	Simulation	Thematic Analysis	Content Analysis	Grounded Theory	Hermeneutics	Statistical Analysis	Meta-Analysis
33	●	●												●	●											
358	●	●												●	●											
89	●		●											●				●								
192	●													●				●								
129	●													●							●		●			●
97	●													●												●
219		●												●												●
92	●															●		●								
202		●														●		●								
91	●															●			●							
94	●															●			●							
69	●															●									●	
56	●																	●	●						●	
159		●	●															●	●						●	
124	●	●																●	●							
35	●																	●	●							
87		●																●		●						
230	●																	●							●	●
163		●																●							●	
156	●																	●							●	
232	●																	●							●	
120	●																	●							●	
158		●																●							●	
109		●																●							●	
341	●																	●							●	
95	●	●																●								●
78		●	●																		●	●	●			
96	●	●																							●	●
137	●																								●	●

A.4 CERSE References

References

[1] J.W. Creswell, Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, Sage publications, 2013.

[2] S. Easterbrook, J. Singer, M.-A. Storey, D. Damian, Selecting Empirical Methods for Software Engineering Research, Springer London, London, pp. 285–311. 10.1007/978-1-84800-044-5_11.

[3] M.M. Jalil, Practical guidelines for conducting research, Donor Committee for Enterprise Development (2013).

[4] B. Kitchenham, P. Brereton, A systematic review of systematic review process research in software engineering, Inf. Softw. Technol. 55 (12) (2013) 2049–2075.

[5] B.A. Kitchenham, S. Linkman, D. Law, Evaluating software engineering methods and tool. parts 1 to 12, ACM SIGSOFT Softw. Eng. Notes (Starting December 1995).

[6] B.A. Kitchenham, S.L. Pfleeger, Principles of survey research. parts 1 to 6, ACM SIGSOFT Softw. Eng. Notes 26–28 (Starting November 2001).

[7] B.A. Kitchenham, D. Budgen, P. Brereton, Evidence-Based software engineering and systematic reviews, 4 CRC Press, 2015.

[8] J. Maxwell, Understanding and validity in qualitative research, Harvard Educ. Rev. 62 (3) (1992) 279–301.

[9] D. Moher, L. Weeks, M. Ocampo, D. Seely, M. Sampson, D.G. Altman, K.F. Schulz, D. Miller, I. Simera, J. Grimshaw, et al., Describing reporting guidelines for health research: a systematic review, J. Clin. Epidemiol. 64 (7) (2011) 718–742.

[10] M. O'Reilly, N. Kiyimba, Advanced Qualitative Research: A Guide to Using Theory, Sage, 2015.

[11] K. Petersen, S. Vakkalanka, L. Kuzniarz, Guidelines for conducting systematic mapping studies in software engineering: an update, Inf. Softw. Technol. 64 (2015) 1–18.

[12] M.C.P. Ramos, P. Barton, S. Jowett, A.J. Sutton, A systematic review of research guidelines in decision-analytic modeling, Value Health 18 (4) (2015) 512–529.

[13] M. Shaw, Writing good software engineering research papers: Minitutorial, Proceedings of the 25th International Conference on Software Engineering, ICSE '03, IEEE Computer Society, Washington, DC, USA, 2003, pp. 726–736.

[14] F. Shull, J. Singer, D.I. Sjøberg, Guide to Advanced Empirical Software Engineering, 93 Springer, 2008.

[15] I. Simera, D. Moher, J. Hoey, K.F. Schulz, D.G. Altman, A catalogue of reporting guidelines for health research, Eur. J. Clin. Invest. 40 (1) (2010) 35–53.

[16] D. Tudhope, M. Lykke Nielsen, Introduction to knowledge organization systems and services, New Rev. Hypermedia Multimedia 12 (1) (2006) 3–9.

[17] C. Wohlin, Is there a future for empirical software engineering? Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, ACM, 2016, p. 1.

[18] S. Adolph, W. Hall, P. Kruchten, A methodological leg to stand on: lessons learned using grounded theory to study software development, Proceedings of the 2008 conference of the center for advanced studies on collaborative research: meeting of minds, ACM, 2008, p. 13.

[19] S. Adolph, W. Hall, P. Kruchten, Using grounded theory to study the experience of software development, Empirical Softw. Eng. 16 (4) (2011) 487–513.

[20] N.B. Ali, K. Petersen, Evaluating strategies for study selection in systematic literature studies, Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, ACM, 2014, p. 45.

[21] J.P.F. Almqvist, Replication of Controlled Experiments in Empirical Software Engineering - A Survey, Faculty of Science, Lund University, Sweden, 2006 Master's thesis.

[22] A. Anderlin-Neto, T. Conte, Threats to validity and their control actions–results of a systematic literature review, Technical Report, Federal University of Amazonas, TR-USES-2014, 2014.

[23] C. Andersson, P. Runeson, A spiral process model for case studies on software quality monitoringmethod and metrics, Softw. Process 12 (2) (2007) 125–140.

[24] A.A. Andrews, A.S. Pradhan, Ethical issues in empirical software engineering: the limits of policy, Empirical Softw. Eng. 6 (2) (2001) 105–110.

[25] E. Arisholm, D.I. Sjøberg, G.J. Carelius, Y. Lindsjørn, A web-based support environment for software engineering experiments, Nordic J. Comput. 9 (3) (2002) 231–247.

[26] D. Avison, R. Baskerville, M. Myers, Controlling action research projects, Inf.

Technol. People 14 (1) (2001) 28–45.
[27] M.A. Babar, H. Zhang, Systematic literature reviews in software engineering: Preliminary results from interviews with researchers, Proceedings of the 2009 3rd International Symposium on Empirical Software Engineering and Measurement, IEEE Computer Society, 2009, pp. 346–355.
[28] J. Bailey, C. Zhang, D. Budgen, M. Turner, S. Charters, Search engine overlaps: Do they agree or disagree? Realising Evidence-Based Software Engineering, 2007. REBSE'07. Second International Workshop on, IEEE, 2007. 2–2
[29] M.T. Baldassarre, D. Caivano, G. Visaggio, Empirical studies for innovation dissemination: ten years of experience, 17th International Conference on Evaluation and Assessment in Software Engineering, EASE '13, Porto de Galinhas, Brazil, April 14–16, 2013, (2013), pp. 144–152.
[30] M.T. Baldassarre, J. Carver, O. Dieste, N. Juristo, Replication types: towards a shared taxonomy, Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, ACM, 2014, p. 18.
[31] V. Basili, R. Tesoriero, P. Costa, M. Lindvall, I. Rus, F. Shull, M. Zelkowitz, Building an experience base for software engineering: A report on the first cebase eworkshop, International Conference on Product Focused Software Process Improvement, Springer, 2001, pp. 110–125.
[32] V.R. Basili, The experimental paradigm in software engineering, Experimental Software Engineering Issues: Critical Assessment and Future Directions, Springer, 1993, pp. 1–12.
[33] V.R. Basili, The role of experimentation in software engineering: past, current, and future, Proceedings of the 18th international conference on Software engineering, IEEE Computer Society, 1996, pp. 442–449.
[34] V.R. Basili, R.W. Selby, D.H. Hutchens, Experimentation in software engineering, IEEE Trans. Software Eng. (7) (1986) 733–743.
[35] V.R. Basili, F. Shull, F. Lanubile, Building knowledge through families of experiments, IEEE Trans. Software Eng. 25 (4) (1999) 456–473.
[36] V.R. Basili, D.M. Weiss, A methodology for collecting valid software engineering data, IEEE Trans. Softw. Eng. 10 (6) (1984) 728–738.
[37] V.R. Basili, M.V. Zelkowitz, D.I. Sjøberg, P. Johnson, A.J. Cowling, Protocols in the use of empirical software engineering artifacts, Empirical Softw. Eng. 12 (1) (2007) 107–119.
[38] R.L. Baskerville, Distinguishing action research from participative case studies, J. Syst. Inf. Technol. 1 (1) (1997) 24–43.
[39] R.L. Baskerville, Investigating information systems with action research, Commun. AIS 2 (3es) (1999) 4.
[40] R.L. Baskerville, A.T. Wood-Harper, A critical perspective on action research as a method for information systems research, Enacting Research Methods in Information Systems: Volume 2, Springer, 2016, pp. 169–190.
[41] I. Benbasat, D.K. Goldstein, M. Mead, The case research strategy in studies of information systems, MIS Q. (1987) 369–386.
[42] I. Benbasat, R.W. Zmud, Empirical research in information systems: the practice of relevance, MIS Q. (1999) 3–16.
[43] S. Biffl, M. Kalinowski, F. Ekaputra, A.A. Neto, T. Conte, D. Winkler, Towards a semantic knowledge base on threats to validity and control actions in controlled experiments, Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, ACM, 2014, p. 49.
[44] J. Biolchini, P.G. Mian, A.C.C. Natali, G.H. Travassos, Systematic review in software engineering, System Engineering and Computer Science Department COPPE/UFRJ, Technical Report ES 679 (05) (2005) 45.
[45] D.F. Birks, W. Fernandez, N. Levina, S. Nasirin, Grounded theory method in information systems research: its nature, diversity and opportunities, Eur. J. Inf. Syst. 22 (1) (2013) 1–8.
[46] B. Boehm, J. Bhuta, D. Garlan, E. Gradman, L. Huang, A. Lam, R. Madachy, N. Medvidovic, K. Meyer, S. Meyers, et al., Using empirical testbeds to accelerate technology maturity and transition: the scrover experience, Empirical Software Engineering, 2004. ISESE'04. Proceedings. 2004 International Symposium on, IEEE, 2004, pp. 117–126.
[47] A. Borges, W. Ferreira, E. Barreiros, A. Almeida, L. Fonseca, E. Teixeira, D. Silva, A. Alencar, S. Soares, Support mechanisms to conduct empirical studies in software engineering, Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, ACM, 2014, p. 50.
[48] M.F. Bosu, S.G. MacDonell, Data quality in empirical software engineering: a targeted review, Proceedings of the 17th International Conference on Evaluation and Assessment in Software Engineering, ACM, 2013, pp. 171–176.
[49] M.F. Bosu, S.G. MacDonell, A taxonomy of data quality challenges in empirical software engineering, 2013 22nd Australian Software Engineering Conference, IEEE, 2013, pp. 97–106.
[50] L. Bratthall, M. Jørgensen, Can you trust a single data source exploratory software engineering case study? Empirical Softw. Eng. 7 (1) (2002) 9–26.
[51] P. Brereton, B. Kitchenham, D. Budgen, Z. Li, Using a protocol template for case study planning, Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering. University of Bari, Italy, Citeseer, 2008.
[52] P. Brereton, B.A. Kitchenham, D. Budgen, M. Turner, M. Khalil, Lessons from applying the systematic literature review process within the software engineering domain, J. Syst. Softw. 80 (4) (2007) 571–583.
[53] A. Brooks, J. Daly, J. Miller, M. Roper, M. Wood, Replication of experimental results in software engineering, International Software Engineering Research Network (ISERN) Technical Report ISERN-96-10, University of Strathclyde (1996).
[54] D. Budgen, J. Bailey, M. Turner, B. Kitchenham, P. Brereton, S. Charters, Cross-domain investigation of empirical practices, IET Softw. 3 (5) (2009) 410–421.
[55] D. Budgen, B. Kitchenham, P. Brereton, The case for knowledge translation, 2013 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, IEEE, 2013, pp. 263–266.
[56] D. Budgen, B. Kitchenham, S. Charters, S. Gibbs, A. Pohthong, J. Keung, P. Brereton, Lessons from conducting a distributed quasi-experiment, 2013 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, IEEE, 2013, pp. 143–152.
[57] D. Budgen, B.A. Kitchenham, S.M. Charters, M. Turner, P. Brereton, S.G. Linkman, Presenting software engineering results using structured abstracts: a randomised experiment, Empirical Softw. Eng. 13 (4) (2008) 435–468.
[58] D. Budgen, B.A. Kitchenham, S.M. Charters, M. Turner, P. Brereton, S.G. Linkman, Presenting software engineering results using structured abstracts: a randomised experiment, Empirical Softw. Eng. 13 (4) (2008) 435–468.
[59] D. Budgen, M. Turner, P. Brereton, B. Kitchenham, Using mapping studies in software engineering, Proceedings of PPIG 2008: 20th Annual Meeting of the Psychology of Programming Interest Group (2008), 8 Lancaster University, 2008, pp. 195–204.
[60] D. Budgen, M. Turner, S. Charters, J. Bailey, B. Kitchenham, P. Brereton, Lessons from a cross-domain investigation of empirical practices, 12th International Conference on Evaluation and Assessment in Software Engineering, EASE 2008, 2008 (2008), pp. 1–12.
[61] D. Budgen, C. Zhang, Preliminary reporting guidelines for experience papers, Proceedings of the 13th International Conference on Evaluation and Assessment in Software Engineering, EASE 2009, 2009 (2009), pp. 1–10.
[62] T. Butler, Towards a hermeneutic method for interpretive research in information systems, J. Inf. Technol. 13 (4) (1998) 285–300.
[63] J.M. Carroll, P.A. Swatman, Structured-case: a methodological framework for building theory in information systems research, Eur. J. Inf. Syst. 9 (4) (2000) 235–242.
[64] J. Carver, L. Jaccheri, S. Morasca, F. Shull, Issues in using students in empirical studies in software engineering education, Software Metrics Symposium, 2003. Proceedings. Ninth International, IEEE, 2003, pp. 239–249.
[65] J. Carver, J. VanVoorhis, V. Basili, Understanding the impact of assumptions on experimental validity, Empirical Software Engineering, 2004. ISESE'04. Proceedings. 2004 International Symposium on, IEEE, 2004, pp. 251–260.
[66] J.C. Carver, Towards reporting guidelines for experimental replications: A proposal, 1st International Workshop on Replication in Empirical Software Engineering, Citeseer, 2010.
[67] J.C. Carver, E. Hassler, E. Hernandes, N.A. Kraft, Identifying barriers to the systematic literature review process, 2013 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, IEEE, 2013, pp. 203–212.
[68] J.C. Carver, L. Jaccheri, S. Morasca, F. Shull, A checklist for integrating student empirical studies with research and teaching goals, Emp. Softw. Eng. 15 (1) (2010) 35–59.
[69] A. Cater-Steel, M. Toleman, T. Rout, Addressing the challenges of replications of surveys in software engineering research, 2005 International Symposium on Empirical Software Engineering, 2005. IEEE, 2005, pp. 10–pp.
[70] A.L. Cavaye, Case study research: a multi-faceted research approach for is, Inf. Syst. J. 6 (3) (1996) 227–242.
[71] L. Chen, M.A. Babar, H. Zhang, Towards an evidence-based understanding of electronic data sources, 14th International Conference on Evaluation and Assessment in Software Engineering, EASE 2010, (2010).
[72] N. Condori-Fernández, R. Wieringa, M. Daneva, B. Mutschler, O. Pastor, An experimental evaluation of a unified checklist for designing and reporting empirical research in Software Engineering, Technical Report, Centre for Telematics and Information Technology, University of Twente. TR-CTIT-12-12, 2012.
[73] R. Conradi, J. Li, O.P.N. Slyngstad, V.B. Kampenes, C. Bunse, M. Morisio, M. Torchiano, Reflections on conducting an international survey of software engineering, 2005 International Symposium on Empirical Software Engineering, 2005. IEEE, 2005, pp. 10–pp.
[74] D. Cruzes, M. Mendonca, V. Basili, F. Shull, M. Jino, Extracting information from experimental software engineering papers, Chilean Society of Computer Science, 2007. SCCC'07. XXVI International Conference of the, IEEE, 2007, pp. 105–114.
[75] D. Cruzes, M. Mendonca, V. Basili, F. Shull, M. Jino, Using context distance measurement to analyze results across studies, Empirical Software Engineering and Measurement, 2007. ESEM 2007. First International Symposium on, IEEE, 2007, pp. 235–244.
[76] D.S. Cruzes, T. Dybå, Synthesizing evidence in software engineering research, Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, ACM, 2010, p. 1.
[77] D.S. Cruzes, T. Dyba, Recommended steps for thematic synthesis in software engineering, 2011 International Symposium on Empirical Software Engineering and Measurement, IEEE, 2011, pp. 275–284.
[78] D.S. Cruzes, T. Dybå, Research synthesis in software engineering: a tertiary study, Inf. Softw. Technol. 53 (5) (2011) 440–455.
[79] D.S. Cruzes, T. Dyba, P. Runeson, M. Host, Case studies synthesis: Brief experience and challenges for the future, 2011 International Symposium on Empirical Software Engineering and Measurement, IEEE, 2011, pp. 343–346.
[80] B. Curtis, Measurement and experimentation in software engineering, Proc. IEEE 68 (9) (1980) 1144–1157.
[81] F.Q. Da Silva, A.L. Santos, S. Soares, A.C.C. França, C.V. Monteiro, F.F. Maciel, Six years of systematic literature reviews in software engineering: an updated tertiary study, Inf. Softw. Technol. 53 (9) (2011) 899–913.
[82] F.Q. da Silva, A.L. Santos, S.C. Soares, A.C.C. França, C.V. Monteiro, A critical appraisal of systematic reviews in software engineering from the perspective of the research questions asked in the reviews, Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, ACM, 2010, p. 33.
[83] P. Darke, G. Shanks, M. Broadbent, Successfully completing case study research:

combining rigour, relevance and pragmatism, Inf. Syst. J. 8 (4) (1998) 273–289.
[84] R. Davison, M.G. Martinsons, N. Kock, Principles of canonical action research, Inf. Syst. J. 14 (1) (2004) 65–86.
[85] R. Dawson, P. Bones, B.J. Oates, P. Brereton, M. Azuma, M.L. Jackson, Empirical methodologies in software engineering, Software Technology and Engineering Practice, 2003. Eleventh Annual International Workshop on, IEEE, 2003, pp. 52–58.
[86] J.C. de Almeida Biolchini, P.G. Mian, A.C.C. Natali, T.U. Conte, G.H. Travassos, Scientific research ontology to support systematic review in software engineering, Adv. Eng. Inf. 21 (2) (2007) 133–151.
[87] B.B.N. de França, G.H. Travassos, Reporting guidelines for simulation-based studies in software engineering, Evaluation & Assessment in Software Engineering (EASE 2012), 16th International Conference on, IET, 2012, pp. 156–160.
[88] C.V. de Magalhães, F.Q. da Silva, Towards a taxonomy of replications in empirical software engineering research: A research proposal, Replication in Empirical Software Engineering Research (RESER), 2013 3rd International Workshop on, IEEE, 2013, pp. 50–55.
[89] C.V. de Magalhães, F.Q. da Silva, R.E. Santos, Investigations about replication of empirical studies in software engineering: preliminary findings from a mapping study, Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, ACM, 2014, p. 37.
[90] C.V. de Magalhães, F.Q. da Silva, R.E. Santos, Investigations about replication of empirical studies in software engineering: preliminary findings from a mapping study, Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, ACM, 2014, p. 37.
[91] R.M. de Mello, G.H. Travassos, An ecological perspective towards the evolution of quantitative studies in software engineering, Proceedings of the 17th International Conference on Evaluation and Assessment in Software Engineering, ACM, 2013, pp. 216–219.
[92] R.M. de Mello, P.C. da Silva, P. Runeson, G.H. Travassos, Towards a framework to support large scale sampling in software engineering surveys, Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, ACM, 2014, p. 48.
[93] R.M. de Mello, P.C. da Silva, G.H. Travassos, Sampling improvement in software engineering surveys, Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, ACM, 2014, p. 13.
[94] R.M. de Mello, P.C. Da Silva, G.H. Travassos, Investigating probabilistic sampling approaches for large-scale surveys in software engineering, Journal of Software Engineering Research and Development 3 (1) (2015) 1.
[95] O. Dieste, E. Fern, R. García-Martínez, N. Juristo, et al., The risk of using the q heterogeneity estimator for software engineering experiments, 2011 International Symposium on Empirical Software Engineering and Measurement, IEEE, 2011, pp. 68–76.
[96] O. Dieste, E. Fernandez, R. Garcia, N. Juristo, Hidden evidence behind useless replications, Proceedings of the International Workshop on Replication in Empirical Software Engineering Research, (2010).
[97] O. Dieste, E. Fernández, R.G. Martínez, N. Juristo, Comparative analysis of meta-analysis methods: when to use which? Evaluation & Assessment in Software Engineering (EASE 2011), 15th Annual Conference on, IET, 2011, pp. 36–45.
[98] O. Dieste, A. Grim, N. Juristo, H. Saxena, et al., Quantitative determination of the relationship between internal validity and bias in software engineering experiments: consequences for systematic literature reviews, 2011 International Symposium on Empirical Software Engineering and Measurement, IEEE, 2011, pp. 285–294.
[99] O. Dieste, A. Grimán, N. Juristo, Developing search strategies for detecting relevant experiments, Emp. Softw. Eng. 14 (5) (2009) 513–539.
[100] O. Dieste, N. Juristo, Systematic review and aggregation of empirical studies on elicitation techniques, IEEE Trans. Software Eng. 37 (2) (2011) 283–304.
[101] O. Dieste, A.G. Padua, Developing search strategies for detecting relevant experiments for systematic reviews, First International Symposium on Empirical Software Engineering and Measurement (ESEM 2007), IEEE, 2007, pp. 215–224.
[102] T. Dingsøyr, Postmortem reviews: purpose and approaches in software engineering, Inf. Softw. Technol. 47 (5) (2005) 293–303.
[103] S. Djokic, G. Succi, W. Pedrycz, M. Mintchev, Meta analysisa method of combining empirical results and its application in object-oriented software systems, OOIS 2001: 7th International Conference on Object-Oriented Information Systems, Springer, 2001, pp. 103–112.
[104] B. Doolin, Alternative views of case research in information systems, Australas. J. Inf. Syst. 3 (2) (1996).
[105] P.S.M. Dos Santos, G.H. Travassos, On the representation and aggregation of evidence in software engineering: a theory and belief-based perspective, Electron Notes Theor. Comput. Sci. 292 (2013) 95–118.
[106] G. Du, J. McElroy, G. Ruhe, A family of empirical studies to compare informal and optimization-based planning of software releases, Proceedings of the 2006 ACM/IEEE international symposium on Empirical software engineering, ACM, 2006, pp. 212–221.
[107] T. Dybå, T. Dingsøyr, Strength of evidence in systematic reviews in software engineering, Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement, ACM, 2008, pp. 178–187.
[108] T. Dybå, T. Dingsøyr, G.K. Hanssen, Applying systematic reviews to diverse study types: An experience report. Empirical Software Engineering and Measurement, 2007. ESEM 2007. First International Symposium on, 7 (2007), pp. 225–234.
[109] T. Dybå, V.B. Kampenes, D.I. Sjøberg, A systematic review of statistical power in software engineering experiments, Inf. Softw. Technol. 48 (8) (2006) 745–755.
[110] T. Dybå, D.I. Sjøberg, D.S. Cruzes, What works for whom, where, when, and why?: on the role of context in empirical software engineering, Proceedings of the ACM-IEEE international symposium on Empirical software engineering and measurement, ACM, 2012, pp. 19–28.
[111] H.M. Edwards, S. McDonald, S.M. Young, The repertory grid technique: its place in empirical software engineering research, Inf. Softw. Technol. 51 (4) (2009) 785–798.
[112] C. Exton, G. Avram, J. Buckley, A. LeGear, An experiential report on the limitations of experimentation as a means of empirically investigating software practitioners. Proceedings of PPIG 2007: 19th Annual Meeting of the Psychology of Programming Interest Group (2007), (2007).
[113] R. Feldt, A. Magazinius, Validity threats in empirical software engineering research-an initial survey. Proceedings of the 22nd International Conference on Software Engineering & Knowledge Engineering (SEKE'2010), (2010), pp. 374–379.
[114] N. Fenton, S.L. Pfleeger, R.L. Glass, Science and substance: a challenge to software engineers, IEEE Softw. 11 (4) (1994) 86.
[115] D.M. Fernández, B. Penzenstadler, M. Kuhrmann, Pattern-based guideline to empirically analyse software development processes, Evaluation & Assessment in Software Engineering (EASE 2012), 16th International Conference on, IET, 2012, pp. 136–145.
[116] B. Freimut, T. Punter, S. Biffl, M. Ciolkowski, State-of-the-art in empirical studies, Report: ViSEK/007/E, Fraunhofer Inst. of Experimental Software Engineering (2002).
[117] R.D. Galliers, Choosing appropriate information systems research approaches: a revised taxonomy, In Proceedings of Information Systems Research: issues, methods and practical guidelines, Citeseer, 1991.
[118] H. Gallis, E. Arisholm, T. Dyba, An initial framework for research on pair programming, Empirical Software Engineering, 2003. ISESE 2003. Proceedings. 2003 International Symposium on, IEEE, 2003, pp. 132–142.
[119] M. Galster, D. Tofan, P. Avgeriou, On integrating student empirical software engineering studies with research and teaching goals, Evaluation & Assessment in Software Engineering (EASE 2012), 16th International Conference on, IET, 2012, pp. 146–155.
[120] R.E. Garcia, E.N. Höhn, E.F. Barbosa, J.C. Maldonado, An ontology for controlled experiments on software engineering. Proceedings of the 20nd International Conference on Software Engineering & Knowledge Engineering (SEKE'2008), (2008), pp. 685–690.
[121] R.E. Garcia, M.C.F. de Oliveira, J.C. Maldonado, M. Mendonça, Visual analysis of data from empirical studies, International Workshop on Visual Languages and Computing, (2004).
[122] D. Gessenharter, A.-M. Merten, A. Raschke, N.F. Porta, Experiences on Using Software Experiments in the Validation of Industrial Research Questions, Software Process and Product Measurement, Springer, 2008, pp. 86–94.
[123] M. Ghafari, M. Saleh, T. Ebrahimi, A federated search approach to facilitate systematic literature review in software engineering, Int. J. Softw. Eng. Appl. 3 (2) (2012) 13.
[124] R.L. Glass, Pilot studies: what, why and how, J. Syst. Softw. 36 (1) (1997) 85–97.
[125] O. Gómez, N. Juristo, S. Vegas, Replication, reproduction and re-analysis: Three ways for verifying experimental findings, Proceedings of the 1st international workshop on replication in empirical software engineering research (RESER 2010), Cape Town, South Africa, (2010).
[126] O.S. Gómez, N. Juristo, S. Vegas, Replications types in experimental disciplines, Proceedings of the 2010 ACM-IEEE international symposium on empirical software engineering and measurement, ACM, 2010, p. 3.
[127] J.M. González-Barahona, G. Robles, On the reproducibility of empirical software engineering studies based on data retrieved from development repositories, Empirical Software Engineering 17 (1–2) (2012) 75–89.
[128] T.R. Green, M.M. Burnett, A.J. Ko, K.J. Rothermel, C.R. Cook, J. Schonfeld, Using the cognitive walkthrough to improve the design of a visual programming experiment, Visual Languages, 2000. Proceedings. 2000 IEEE International Symposium on, IEEE, 2000, pp. 172–179.
[129] L. Guzmán, C. Lampasona, C. Seaman, D. Rombach, Survey on research synthesis in software engineering, Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, ACM, 2014, p. 2.
[130] T. Hall, V. Flynn, Ethical issues in software engineering research: a survey of current practice, Empirical Softw. Eng. 6 (4) (2001) 305–317.
[131] J. Hannay, M. Jørgensen, The role of deliberate artificial design elements in software engineering experiments, IEEE Trans. Software Eng. 34 (2) (2008) 242–259.
[132] J.E. Hannay, D.I. Sjoberg, T. Dyba, A systematic review of theory use in software engineering experiments, IEEE Trans. Software Eng. 33 (2) (2007) 87–107.
[133] B.H. Hansen, K. Kautz, Grounded theory applied-studying information systems development methodologies in practice, Proceedings of the 38th Annual Hawaii International Conference on System Sciences, IEEE, 2005. 264b–264b
[134] H.-J. Happel, S. Seedorf, Applications of ontologies in software engineering, Proc. of Workshop on Sematic Web Enabled Software Engineering"(SWESE) on the ISWC, Citeseer, 2006, pp. 5–9.
[135] R. Harrison, N. Badoo, E. Barry, S. Biffl, A. Parra, B. Winter, J. Wuest, Directions and methodologies for empirical software engineering research, Emp. Softw. Eng. 4 (4) (1999) 405–410.
[136] E. Hassler, J.C. Carver, N.A. Kraft, D. Hale, Outcomes of a community workshop to identify and rank barriers to the systematic literature review process, Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, ACM, 2014, p. 31.
[137] W. Hayes, Research synthesis in software engineering: a case for meta-analysis, Software Metrics Symposium, 1999. Proceedings. Sixth International, IEEE, 1999, pp. 143–151.

[138] A. Höfer, W.F. Tichy, Status of empirical research in software engineering, Empirical Software Engineering Issues. Critical Assessment and Future Directions, Springer, 2007, pp. 10–19.
[139] M. Höst, P. Runeson, Checklists for software engineering case study research. Empirical Software Engineering and Measurement, 2007. ESEM 2007. First International Symposium on, (2007), pp. 479–481.
[140] M. Host, C. Wohlin, T. Thelin, Experimental context classification: incentives and experience of subjects, Proceedings. 27th International Conference on Software Engineering, 2005. ICSE 2005. IEEE, 2005, pp. 470–478.
[141] S.E. Hove, B. Anda, Experiences from conducting semi-structured interviews in empirical software engineering research, 11th IEEE International Software Metrics Symposium (METRICS'05), IEEE, 2005, pp. 10–pp.
[142] J. Hughes, S. Jones, Reflections on the use of grounded theory in interpretive information systems research, Proceedings of the 11th European Conference on Information Systems, ECIS 2003 (2003) 62.
[143] S. Imtiaz, M. Bano, N. Ikram, M. Niazi, A tertiary study: experiences of conducting systematic literature reviews in software engineering, Proceedings of the 17th International Conference on Evaluation and Assessment in Software Engineering, ACM, 2013, pp. 177–182.
[144] M. Ivarsson, T. Gorschek, A method for evaluating rigor and industrial relevance of technology evaluations, Emp. Softw. Eng. 16 (3) (2011) 365–395.
[145] R.H. Jackson, P.T. Boggs, S.G. Nash, S. Powell, Guidelines for reporting results of computational experiments. report of the ad hoc committee, Math. Program 49 (1) (1990) 413–425.
[146] S. Jalali, C. Wohlin, Systematic literature studies: database searches vs. backward snowballing, Proceedings of the ACM-IEEE international symposium on Empirical software engineering and measurement, ACM, 2012, pp. 29–38.
[147] S.L. Jarvenpaa, G.W. Dickson, G. DeSanctis, Methodological issues in experimental is research: experiences and recommendations, MIS Q. (1985) 141–156.
[148] A. Jedlitschka, L.C. Briand, The role of controlled experiments working group results, Empirical Software Engineering Issues. Critical Assessment and Future Directions, Springer, 2007, pp. 58–62.
[149] A. Jedlitschka, M. Ciolkowski, Towards evidence in software engineering, Empirical Software Engineering, 2004. ISESE'04. Proceedings. 2004 International Symposium on, IEEE, 2004, pp. 261–270.
[150] A. Jedlitschka, N. Juristo, D. Rombach, Reporting experiments to satisfy professionals information needs, Empirical Softw. Eng. 19 (6) (2014) 1921–1955.
[151] A. Jedlitschka, D. Pfahl, Experience-based model-driven improvement management with combined data sources from industry and academia, Empirical Software Engineering, 2003. ISESE 2003. Proceedings. 2003 International Symposium on, IEEE, 2003, pp. 154–161.
[152] A. Jedlitschka, D. Pfahl, Reporting guidelines for controlled experiments in software engineering, 2005 International Symposium on Empirical Software Engineering, 2005. IEEE, 2005, pp. 10–pp.
[153] J. Ji, J. Li, R. Conradi, C. Liu, J. Ma, W. Chen, Some lessons learned in conducting software engineering surveys in china, Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement, ACM, 2008, pp. 168–177.
[154] H. Jordan, S. Beecham, G. Botterweck, Modelling software engineering research with rsml, Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, ACM, 2014, p. 44.
[155] M. Jørgensen, T. Dybå, B.A. Kitchenham, Teaching evidence-based software engineering to university students, 11th IEEE International Symposium on Software Metrics (METRICS 2005), (2005), p. 24.
[156] M. Jørgensen, D. Sjøberg, Generalization and theory-building in software engineering research, Empirical Assess. Softw. Eng. Proc. (2004) 29–36.
[157] N. Juristo, Ana María Moreno, Sira Vegas, Towards building a solid empirical body of knowledge in testing techniques. ACM SIGSOFT Software Eng. Notes 29 (5) (2004) 1–4.
[158] N. Juristo, O.S. Gómez, Replication of software engineering experiments, Empirical software engineering and verification, Springer, 2012, pp. 60–88.
[159] N. Juristo, O.S. Gómez, Replication of software engineering experiments, Empirical software engineering and verification, Springer, 2012, pp. 60–88.
[160] N. Juristo, S. Vegas, Using differences among replications of software engineering experiments to gain knowledge, Proceedings of the 2009 3rd International Symposium on Empirical Software Engineering and Measurement, IEEE Computer Society, 2009, pp. 356–366.
[161] N. Juristo, S. Vegas, The role of non-exact replications in software engineering experiments, Empirical Softw. Eng. 16 (3) (2011) 295–324.
[162] N. Juristo, S. Vegas, M. Solari, S. Abrahão, I. Ramos, A process for managing interaction between experimenters to get useful similar replications, Inf. Softw. Technol. 55 (2) (2013) 215–225.
[163] V. Kampenes, Quality of design, analysis and reporting of software engineering experiments: A systematic review, Faculty of Mathematics and Natural Sciences, University of Oslo, 2007 Ph.D. thesis.
[164] V.B. Kampenes, B. Anda, T. Dybå, Flexibility in research designs in empirical software engineering, Eval.Assess. Softw. Eng. (EASE 2008). BCS (2008).
[165] V.B. Kampenes, T. Dybå, J.E. Hannay, D.I. Sjøberg, A systematic review of effect size in software engineering experiments, Inf. Softw. Technol. 49 (11) (2007) 1073–1086.
[166] V.B. Kampenes, T. Dybå, J.E. Hannay, D.I. Sjøberg, A systematic review of quasi-experiments in software engineering, Inf. Softw. Technol. 51 (1) (2009) 71–82.
[167] A. Karahasanoviæ, B. Anda, E. Arisholm, S.E. Hove, M. Jørgensen, D.I. Sjøberg, R. Welland, Collecting feedback during software engineering experiments, Empirical Softw. Eng. 10 (2) (2005) 113–147.
[168] A. Karahasanoviæ, B. Anda, E. Arisholm, S.E. Hove, M. Jørgensen, D.I. Sjøberg, R. Welland, Collecting feedback during software engineering experiments, Empirical Softw. Eng. 10 (2) (2005) 113–147.
[169] J. Karn, A. Cowling, Using ethnographic methods to carry out human factors research in software engineering, Behav. Res. Methods 38 (3) (2006) 495–503.
[170] M. Kasunic, Designing an effective survey, Technical Report, Carnegie Mellon Software Engineering Institute. CMU/SEI-2005-HB-004, 2005.
[171] D. Kirk, S.G. MacDonell, Investigating a conceptual construct for software context, Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, ACM, 2014, p. 27.
[172] C. Kirsopp, M. Shepperd, Making inferences with small numbers of training sets, IEE Proceedings-Software 149 (5) (2002) 123–130.
[173] B. Kitchenham, Desmet: a method for evaluating software engineering methods and tools technical report tr96-09, Department of Computer Science, University of Keele, Staffordshire (1996).
[174] B. Kitchenham, Procedures for performing systematic reviews, Keele, UK, Keele University 33 (2004) (2004) 1–26.
[175] B. Kitchenham, H. Al-Khilidar, M.A. Babar, M. Berry, K. Cox, J. Keung, F. Kurniawati, M. Staples, H. Zhang, L. Zhu, Evaluating guidelines for empirical software engineering studies, Proceedings of the 2006 ACM/IEEE international symposium on Empirical software engineering, ACM, 2006, pp. 38–47.
[176] B. Kitchenham, H. Al-Khilidar, M.A. Babar, M. Berry, K. Cox, J. Keung, F. Kurniawati, M. Staples, H. Zhang, L. Zhu, Evaluating guidelines for empirical software engineering studies, Proceedings of the 2006 ACM/IEEE international symposium on Empirical software engineering, ACM, 2006, pp. 38–47.
[177] B. Kitchenham, P. Brereton, D. Budgen, Mapping study completeness and reliability-a case study, Evaluation & Assessment in Software Engineering (EASE 2012), 16th International Conference on, IET, 2012, pp. 126–135.
[178] B. Kitchenham, P. Brereton, Z. Li, D. Budgen, A. Burn, Repeatability of systematic literature reviews, Evaluation & Assessment in Software Engineering (EASE 2011), 15th Annual Conference on, IET, 2011, pp. 46–55.
[179] B. Kitchenham, P. Brereton, M. Turner, M. Niazi, S. Linkman, R. Pretorius, D. Budgen, The impact of limited search procedures for systematic literature reviews a participant-observer case study, Proceedings of the 2009 3rd International Symposium on Empirical Software Engineering and Measurement, IEEE Computer Society, 2009, pp. 336–345.
[180] B. Kitchenham, D. Budgen, P. Brereton, M. Turner, S. Charters, S. Linkman, Large-scale software engineering questions-expert opinion or empirical evidence? IET Softw. 1 (5) (2007) 161–171.
[181] B. Kitchenham, J. Fry, S. Linkman, The case against cross-over designs in software engineering, Software Technology and Engineering Practice, 2003. Eleventh Annual International Workshop on, IEEE, 2003, pp. 65–67.
[182] B. Kitchenham, S. Linkman, D. Law, Desmet: a methodology for evaluating software engineering methods and tools, Comput. Control Eng. J. 8 (3) (1997) 120–126.
[183] B. Kitchenham, S.L. Pfleeger, Principles of survey research: part 5: populations and samples, ACM SIGSOFT Softw. Eng. Notes 27 (5) (2002) 17–20.
[184] B. Kitchenham, S. Lawrence Pfleeger., Principles of survey research part 6: data analysis. ACM SIGSOFT Softw. Eng. Notes 28 (2) (2003).
[185] B. Kitchenham, L. Pickard, S.L. Pfleeger, Case studies for method and tool evaluation, IEEE Softw. 12 (4) (1995) 52.
[186] B. Kitchenham, D.I. Sjøberg, O.P. Brereton, D. Budgen, T. Dybå, M. Höst, D. Pfahl, P. Runeson, Can we evaluate the quality of software engineering experiments? Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, ACM, 2010, p. 2.
[187] B.A. Kitchenham, Evaluating software engineering methods and tool part 1: the evaluation context and evaluation methods, ACM SIGSOFT Softw. Eng. Notes 21 (1) (1996) 11–14.
[188] B.A. Kitchenham, Evaluating software engineering methods and toolpart 3: selecting an appropriate evaluation methodpractical issues, ACM SIGSOFT Softw. Eng. Notes 21 (4) (1996) 9–12.
[189] B.A. Kitchenham, Evaluating software engineering methods and toolpart 3: selecting an appropriate evaluation methodpractical issues, ACM SIGSOFT Softw. Eng. Notes 21 (4) (1996) 9–12.
[190] B.A. Kitchenham, Evaluating software engineering methods and tools, part 7: planning feature analysis evaluation, ACM SIGSOFT Softw. Eng. Notes 22 (4) (1997) 21–24.
[191] B.A. Kitchenham, O.P. Brereton, D. Budgen, Z. Li, An evaluation of quality checklist proposals-a participant-observer case study, 13th International Conference on Evaluation and Assessment in Software Engineering, (2009).
[192] B.A. Kitchenham, P. Brereton, M. Turner, M.K. Niazi, S. Linkman, R. Pretorius, D. Budgen, Refining the systematic literature review processtwo participant-observer case studies, Empirical Softw. Eng. 15 (6) (2010) 618–653.
[193] B.A. Kitchenham, D. Budgen, O.P. Brereton, The value of mapping studies: a participant observer case study, Proceedings of the 14th international conference on Evaluation and Assessment in Software Engineering, British Computer Society, 2010, pp. 25–33.
[194] B.A. Kitchenham, D. Budgen, O.P. Brereton, Using mapping studies as the basis for further research–a participant-observer case study, Inf. Softw. Technol. 53 (6) (2011) 638–651.
[195] B.A. Kitchenham, S. Charters, Guidelines for performing systematic literature reviews in software engineering, Technical Report, Keele University and Durham University Joint Report. EBSE 2007-001., 2007.
[196] B.A. Kitchenham, T. Dyba, M. Jorgensen, Evidence-based software engineering, Proceedings of the 26th international conference on software engineering, IEEE Computer Society, 2004, pp. 273–281.
[197] B.A. Kitchenham, L. Jones, Evaluating software engineering methods and tool part

5: the influence of human factors, ACM SIGSOFT Softw. Eng. Notes 22 (1) (1997) 13–15.
[198] B.A. Kitchenham, L. Jones, Evaluating sw eng. methods and tools, part 8: analysing a feature analysis evaluation, ACM SIGSOFT Softw. Eng. Notes 22 (5) (1997) 10–12.
[199] B.A. Kitchenham, S. Linkman, D. Law, Critical review of quantitative assessment, Softw. Eng. J. 9 (2) (1994) 43–53.
[200] B.A. Kitchenham, S.L. Pfleeger, Principles of survey research: part 3: constructing a survey instrument, ACM SIGSOFT Softw. Eng. Notes 27 (2) (2002) 20–24.
[201] B.A. Kitchenham, S.L. Pfleeger, Principles of survey research: part 4: questionnaire evaluation, ACM SIGSOFT Softw. Eng. Notes 27 (3) (2002) 20–23.
[202] B.A. Kitchenham, S.L. Pfleeger, L.M. Pickard, P.W. Jones, D.C. Hoaglin, K. El Emam, J. Rosenberg, Preliminary guidelines for empirical research in software engineering, IEEE Trans. Softw. Eng. 28 (8) (2002) 721–734.
[203] B.A. Kitchenham, L.M. Pickard, Evaluating software eng. methods and tools part 10: designing and running a quantitative case study, ACM SIGSOFT Softw. Eng. Notes 23 (3) (1998) 20–22.
[204] B.A. Kitchenham, L.M. Pickard, Evaluating software engineering methods and tools, part 11: analysing quantitative case studies, ACM SIGSOFT Softw. Eng. Notes 23 (4) (1998) 18–20.
[205] B.A. Kitchenham, L.M. Pickard, Evaluating software engineering methods and tools: part 9: quantitative case study methodology, ACM SIGSOFT Softw. Eng. Notes 23 (1) (1998) 24–26.
[206] H.K. Klein, M.D. Myers, A set of principles for conducting and evaluating interpretive field studies in information systems, MIS Q. (1999) 67–93.
[207] A.J. Ko, T.D. LaToza, M.M. Burnett, A practical guide to controlled experiments of software engineering tools with human participants, Empirical Softw. Eng. 20 (1) (2015) 110–141.
[208] J. Kontio, L. Lehtola, J. Bragge, Using the focus group method in software engineering: obtaining practitioner and user experiences, Empirical Softw. Eng., 2004. ISESE'04. Proceedings. 2004 International Symposium on, IEEE, 2004, pp. 271–280.
[209] J.L. Krein, C.D. Knutson, A case for replication: Synthesizing research methodologies in software engineering, RESER2010: proceedings of the 1st international workshop on replication in empirical software engineering research, (2010).
[210] A. Lam, B. Boehm, Experiences in developing and applying a software engineering technology testbed, Empirical Softw. Eng. 14 (5) (2009) 579–601.
[211] F. Lau, A review on the use of action research in information systems studies, Information systems and qualitative research, Springer, 1997, pp. 31–68.
[212] F. Lau, Toward a framework for action research in information systems studies, Inf. Technol. People 12 (2) (1999) 148–176.
[213] A.S. Lee, A scientific methodology for mis case studies, MIS Q. (1989) 33–50.
[214] T.C. Lethbridge, S.E. Sim, J. Singer, Softw. anthropology: performing field studies in software companies, Special Issue Empirical Studies IEEE Trans.Softw. Eng. (1998).
[215] T.C. Lethbridge, S.E. Sim, J. Singer, Studying software engineers: data collection techniques for software field studies, Empirical Softw. Eng. 10 (3) (2005) 311–341.
[216] S.A. Licorish, S.G. MacDonell, Adopting softer approaches in the study of repository data: a comparative analysis, Proceedings of the 17th International Conference on Evaluation and Assessment in Softw. Eng. ACM, 2013, pp. 240–245.
[217] C.M. Lott, H.D. Rombach, Repeatable software engineering experiments for comparing defect-detection techniques, Empirical Softw. Eng. 1 (3) (1996) 241–277.
[218] J. Lung, J. Aranda, S. Easterbrook, G. Wilson, On the difficulty of replicating human subjects studies in software engineering, 2008 ACM/IEEE 30th International Conference on Softw. Eng. IEEE, 2008, pp. 191–200.
[219] S. MacDonell, M. Shepperd, B. Kitchenham, E. Mendes, How reliable are systematic reviews in empirical software engineering? IEEE Trans. Softw. Eng. 36 (5) (2010) 676–687.
[220] H. Maimbo, G. Pervan, Designing a case study protocol for application in is research, Pacific Asia Conference on Information Systems, PACIS 2005 (2005) 106.
[221] C. Marshall, P. Brereton, Tools to support systematic literature reviews in software engineering: A mapping study, 2013 ACM/IEEE International Symposium on Empirical Softw. Eng. and Measurement, IEEE, 2013, pp. 296–299.
[222] C. Marshall, P. Brereton, B. Kitchenham, Tools to support systematic reviews in software engineering: A feature analysis, Proceedings of the 18th International Conference on Evaluation and Assessment in Softw. Eng. ACM, 2014, p. 13.
[223] L. Mathiassen, Collaborative practice research, Inf. Technol. People 15 (4) (2002) 321–345.
[224] L. McLeod, S.G. MacDonell, B. Doolin, Qualitative research on software development: a longitudinal case study methodology, Empirical Softw. Eng. 16 (4) (2011) 430–459.
[225] T. Mende, Replication of defect prediction studies: problems, pitfalls and recommendations, Proceedings of the 6th International Conference on Predictive Models in Softw. Eng. ACM, 2010, p. 5.
[226] M. Mendonça, D. Cruzes, J. Dias, M.C.F. de Oliveira, Using observational pilot studies to test and improve lab packages, Proceedings of the 2006 ACM/IEEE international symposium on Empirical software engineering, ACM, 2006, pp. 48–57.
[227] M.G. Mendonça, J.C. Maldonado, M.C. De Oliveira, J. Carver, S.C. Fabbri, F. Shull, G.H. Travassos, E.N. Höhn, V.R. Basili, A framework for software engineering experimental replications, Eng. of Complex Computer Systems, 2008. ICECCS 2008. 13th IEEE International Conference on, IEEE, 2008, pp. 203–212.
[228] P. Mian, T. Conte, A. Natali, J. Biolchini, G. Travassos, A systematic review process for software engineering, ESELAW05: 2nd Experimental Softw. Eng. Latin American Workshop, (2005).
[229] J. Miller, Can results from software engineering experiments be safely combined? Softw. Metrics Symposium, 1999. Proceedings. Sixth International, IEEE, 1999, pp. 152–158.
[230] J. Miller, Applying meta-analytical procedures to software engineering experiments, J. Syst. Softw. 54 (1) (2000) 29–39.
[231] J. Miller, Triangulation as a basis for knowledge discovery in software engineering, Empirical Softw. Eng. 13 (2) (2008) 223–228.
[232] J. Miller, J. Daly, M. Wood, M. Roper, A. Brooks, Statistical power and its subcomponentsmissing and misunderstood concepts in empirical software engineering research, Inf. Softw. Technol. 39 (4) (1997) 285–295.
[233] J. Mingers, Combining is research methods: towards a pluralist methodology, Inf. Systems Res. 12 (3) (2001) 240–259.
[234] A. Mockus, B. Anda, D.I. Sjøberg, Experiences from replicating a case study to investigate reproducibility of software development, Proceedings of the 1st International Workshop on Replication in Empirical Softw. Eng. Research, (2010).
[235] P. Mohagheghi, R. Conradi, Vote-counting for combining quantitative evidence from empirical studies-an example, Proceedings of the International Symposium on Empirical Softw. Eng. (ISESE'04), (2004).
[236] T. Moher, G.M. Schneider, Methods for improving controlled experimentation in software engineering, Proceedings of the 5th international conference on Softw. engineering, IEEE Press, 1981, pp. 224–233.
[237] T. Moher, G.M. Schneider, Methodology and experimental research in software engineering, Int. J. Man. Mach. Stud. 16 (1) (1982) 65–87.
[238] E. Mumford, Advice for an action researcher, Inf. Technol. People 14 (1) (2001) 12–27.
[239] E. Murphy-Hill, G.C. Murphy, W.G. Griswold, Understanding context: creating a lasting impact in experimental software engineering research, Proceedings of the FSE/SDP workshop on Future of software engineering research, ACM, 2010, pp. 255–258.
[240] M. Myers, Investigating information systems with ethnographic research, Commun. AIS 2 (4es) (1999) 1.
[241] M.D. Myers, M. Newman, The qualitative interview in is research: examining the craft, Inf. Org. 17 (1) (2007) 2–26.
[242] A.A. Neto, T. Conte, A conceptual model to address threats to validity in controlled experiments, Proceedings of the 17th International Conference on Evaluation and Assessment in Softw. Eng. ACM, 2013, pp. 82–85.
[243] J. Nielsen, Advances in Human-computer Interaction (Vol. 3), Ablex Publishing Corp., Norwood, NJ, USA, 1992, pp. 69–82.
[244] J. Nielsen, Evaluating the thinking-aloud technique for use by computer scientists (1992b).
[245] B.J. Oates, G. Capper, Using systematic reviews and evidence-based software engineering with masters students, EASE, 9 (2009), pp. 20–21.
[246] M. Ohisson, P. Runeson, Experience from replicating empirical studies on prediction models, Softw. Metrics, 2002. Proceedings. Eighth IEEE Symposium on, IEEE, 2002, pp. 217–226.
[247] C. Okoli, S.D. Pawlowski, The delphi method as a research tool: an example, design considerations and applications, Inf. Manage. 42 (1) (2004) 15–29.
[248] T.I. Ören, Concepts and criteria to assess acceptability of simulation studies: a frame of reference, Commun. ACM 24 (4) (1981) 180–189.
[249] G. Paré, Enhancing the rigor of qualitative research: application of a case methodology to build theories of it implementation, The Qualitative Report 7 (4) (2002) 1–34.
[250] D.L. Parnas, The limits of empirical studies of software engineering, Empirical Softw. Eng., 2003. ISESE 2003. Proceedings. 2003 International Symposium on, IEEE, 2003, pp. 2–5.
[251] C. Passos, D.S. Cruzes, T. Dybå, M. Mendonça, Challenges of applying ethnography to study software practices, Proceedings of the ACM-IEEE international symposium on Empirical software engineering and measurement, ACM, 2012, pp. 9–18.
[252] S. Patil, A. Kobsa, A. John, D. Seligmann, Methodological reflections on a field study of a globally distributed software project, Inf. Softw. Technol. 53 (9) (2011) 969–980.
[253] W.M. Pereira, G.H. Travassos, Towards the conception of scientific workflows for in silico experiments in software engineering, Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Softw. Eng. and Measurement, ACM, 2010, p. 34.
[254] D.E. Perry, A.A. Porter, L.G. Votta, Empirical studies of software engineering: a roadmap, Proceedings of the conference on The future of Softw. engineering, ACM, 2000, pp. 345–355.
[255] D.E. Perry, S.E. Sim, S.M. Easterbrook, Case studies for software engineers, Softw. Eng., 2004. ICSE 2004. Proceedings. 26th International Conference on, IEEE, 2004, pp. 736–738.
[256] K. Petersen, N.B. Ali, Identifying strategies for study selection in systematic reviews and maps. Empirical Softw. Eng. and Measurement, 2011. ESEM 2011. International Symposium on, (2011), pp. 351–354.
[257] K. Petersen, R. Feldt, S. Mujtaba, M. Mattsson, Systematic mapping studies in software engineering, 12th international conference on evaluation and assessment in software engineering, 17 sn, 2008.
[258] K. Petersen, C. Gencel, Worldviews, research methods, and their relationship to validity in empirical software engineering research, Softw. Measurement and the 2013 Eighth International Conference on Softw. Process and Product Measurement (IWSM-MENSURA), 2013 Joint Conference of the 23rd International Workshop on, IEEE, 2013, pp. 81–89.
[259] K. Petersen, C. Wohlin, Context in industrial software engineering research, Proceedings of the 2009 3rd International Symposium on Empirical Softw. Eng. and Measurement, IEEE Computer Society, 2009, pp. 401–404.

[260] S.L. Pfleeger, Design and analysis in software engineering: the language of case studies and formal experiments, ACM SIGSOFT Softw. Eng. Notes 19 (4) (1994) 16–20.
[261] S.L. Pfleeger, Experimental design and analysis in software engineering, Ann. Softw. Eng. 1 (1) (1995) 219–253.
[262] S.L. Pfleeger, Understanding and improving technology transfer in software engineering, J. Syst. Softw. 47 (2) (1999) 111–124.
[263] S.L. Pfleeger, B.A. Kitchenham, Principles of survey research: part 1: turning lemons into lemonade, ACM SIGSOFT Softw. Eng. Notes 26 (6) (2001) 16–18.
[264] S.L. Pfleeger, B.A. Kitchenham, Principles of survey research part 2: designing a survey, Softw. Eng. Notes 27 (1) (2002) 18–20.
[265] L.M. Pickard, B.A. Kitchenham, P.W. Jones, Combining empirical results in software engineering, Inf. Softw. Technol. 40 (14) (1998) 811–821.
[266] A. Pinsonneault, K. Kraemer, Survey research methodology in management information systems: an assessment, J. Manage. Inf. Syst. 10 (2) (1993) 75–105.
[267] C. Potts, Softw.-engineering research revisited, IEEE Softw. 10 (5) (1993) 19–28.
[268] R. Prikladnicki, A. Boden, G. Avram, C.R. de Souza, V. Wulf, Data collection in global software engineering research: learning from past experience, Empirical Softw. Eng. 19 (4) (2014) 822–856.
[269] R. Prikladnicki, R. Evaristo, D. Damian, J.L.N. Audy, Conducting qualitative research in an international and distributed research team: challenges and lessons learned, Hawaii International Conference on System Sciences, Proceedings of the 41st Annual, IEEE, 2008. 442–442
[270] R. Prikladnicki, R. Evaristo, K. Gallagher, L.T. Lopes, J.L.N. Audy, The role of culture in interpreting qualitative data: methodological issues in an exploratory study of cross-cultural distributed software development, the 13th Special Interest Group on Cross-Cultural Research in Information Systems (SIGCCRIS) at ICIS, Las Vegas (2005).
[271] T. Punter, M. Ciolkowski, B. Freimut, I. John, Conducting on-line surveys in software engineering, Empirical Softw. Eng., 2003. ISESE 2003. Proceedings. 2003 International Symposium on, IEEE, 2003, pp. 80–88.
[272] A. Rainer, S. Beecham, A follow-up empirical evaluation of evidence based software engineering by undergraduate students, 12th International Conference on Evaluation and Assessment in Softw. Eng., University of Bari, Italy, Citeseer, 2008.
[273] A. Rainer, S. Beecham, Supplementary guidelines and assessment scheme for the use of evidence based software engineering, Hatfield, UK, University Of Hertfordshire Technical Report CS-TR-469 (2008).
[274] A. Rainer, T. Hall, N. Baddoo, A preliminary empirical investigation of the use of evidence based software engineering by under-graduate students. Proceedings of the 10th International Conference on Evaluation and Assessment in Softw. Eng. EASE 2006, British Computer Society, 2006.
[275] A. Rainer, D. Jagielska, T. Hall, Softw. engineering practice versus evidence-based software engineering research, ACM SIGSOFT Softw. Eng. Notes, 30 ACM, 2005, pp. 1–5.
[276] M. Riaz, M. Sulayman, N. Salleh, E. Mendes, Experiences conducting systematic reviews from novices perspective, Proceedings of the 14th international conference on Evaluation and Assessment in Softw. Eng. British Computer Society, 2010, pp. 44–53.
[277] D.M. Ribeiro, M. Cardoso, F.Q. da Silva, C. França, Using qualitative metasummary to synthesize empirical findings in literature reviews, Proceedings of the 8th ACM/IEEE International Symposium on Empirical Softw. Eng. and Measurement, ACM, 2014, p. 47.
[278] H. Robinson, J. Segal, H. Sharp, Ethnographically-informed empirical studies of software practice, Inf. Softw. Technol. 49 (6) (2007) 540–551.
[279] H. Robinson, J. Segal, H. Sharp, Ethnographically-informed empirical studies of software practice, Inf. Softw. Technol. 49 (6) (2007) 540–551.
[280] G. Robles, D.M. Germán, Beyond replication: An example of the potential benefits of replicability in the mining of software repositories community, Proceedings of the 1st international workshop on replication in empirical software engineering sesearch (RESER 2010), (2010).
[281] F.P. Romero, J.A. Olivas, M. Genero, M. Piattini, Automatic extraction of the main terminology used in empirical software engineering through text mining techniques, Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement, ACM, 2008, pp. 357–358.
[282] K. Rönkkö, Ethnography and distributed software development, ICSE Workshop Beg, Borrow Or Steal: Using Multidisciplinary Approaches In Empirical Softw. Eng. Research, Limerick, Ireland, (2000).
[283] K. Rönkkö, Softw. Practice from the Inside: Ethnography Applied to Softw. Eng. Blekinge Institute of Technology, Department of Softw. Eng. and Computer Science, 2002 Ph.D. thesis.
[284] K. Rönkkö, Interpretation, interaction and reality construction in software engineering: an explanatory model, Inf. Softw. Technol. 49 (6) (2007) 682–693.
[285] K. Ronkko, O. Lindeberg, Y. Dittrich, 'bad practice'or'bad methods' are software engineering and ethnographic discourses incompatible? Empirical Softw. Eng., 2002. Proceedings. 2002 International Symposium n, IEEE, 2002, pp. 204–210.
[286] J. Rosenberg, A methodology for evaluating predictive metrics, Adv. Comput. 53 (2000) 285–318.
[287] P. Runeson, Using students as experiment subjects–an analysis on graduate and freshmen student data, Proceedings of the 7th International Conference on Empirical Assessment in Softw. Eng..–Keele University, UK, (2003), pp. 95–102.
[288] P. Runeson, M. Höst, Guidelines for conducting and reporting case study research in software engineering, Empirical Softw. Eng. 14 (2) (2009) 131–164.
[289] P. Runeson, M. Höst, Guidelines for conducting and reporting case study research in software engineering, Empirical Softw. Eng. 14 (2) (2009) 131–164.
[290] P. Runeson, A. Stefik, A. Andrews, Variation factors in the design and analysis of replicated controlled experiments, Empirical Softw. Eng. 19 (6) (2014) 1781–1808.
[291] P. Runeson, T. Thelin, Prospects and limitations for cross-study analyses–a study on an experiment series, 2nd Workshop in Workshop Series on Empirical Softw. Eng. (2003), pp. 133–142.
[292] C. Sadler, B.A. Kitchenham, Evaluating software engineering methods and toolpart 4: the influence of human factors, ACM SIGSOFT Softw. Eng. Notes 21 (5) (1996) 11–13.
[293] P.S.M.d. Santos, G.H. Travassos, Action research use in software engineering: An initial survey, Proceedings of the 2009 3rd International Symposium on Empirical Softw. Eng. and Measurement, IEEE Computer Society, 2009, pp. 414–417.
[294] L.P. Scatalon, R.E. Garcia, R.C.M. Correia, Packaging controlled experiments using an evolutionary approach based on ontology (s). Proceedings of the 23nd International Conference on Softw. Eng. & Knowledge Eng. (SEKE'2011), (2011), pp. 408–413.
[295] C.B. Seaman, Qualitative methods in empirical studies of software engineering, IEEE Trans. Softw. Eng. 25 (4) (1999) 557–572.
[296] P.B. Seddon, R. Scheepers, Towards the improved treatment of generalization of knowledge claims in is research: drawing general conclusions from samples, Eur. J. Inf. Syst. 21 (1) (2012) 6–21.
[297] J. Segal, The nature of evidence in empirical software engineering, Softw. Technology and Eng. Practice, 2003. Eleventh Annual International Workshop on, IEEE, 2003, pp. 40–47.
[298] J. Segal, Some parallels between empirical software engineering and research in human-computer interaction, Proceedings of the 7th International Conference on Empirical Assessment in Softw. Eng..–Keele University, UK, (2003), pp. 63–72.
[299] J. Segal, A. Grinyer, H. Sharp, The type of evidence produced by empirical software engineers, ACM SIGSOFT Softw. Eng. Notes, 30 ACM, 2005, pp. 1–4.
[300] M. Shaw, What makes good research in software engineering? Int. J. Softw. Tools Technol. Trans. 4 (1) (2002) 1–7.
[301] F. Shull, V. Basili, J. Carver, J.C. Maldonado, G.H. Travassos, M. Mendonça, S. Fabbri, Replicating software engineering experiments: addressing the tacit knowledge problem, Empirical Softw. Eng., 2002. Proceedings. 2002 International Symposium n, IEEE, 2002, pp. 7–16.
[302] F. Shull, J. Carver, L. Hochstein, V. Basili, Empirical study design in the area of high-performance computing (hpc), 2005 International Symposium on Empirical Softw. Eng., 2005. IEEE, 2005, pp. 10–pp.
[303] F. Shull, J. Carver, G.H. Travassos, An empirical methodology for introducing software processes, ACM SIGSOFT Softw. Eng. Notes, 26 ACM, 2001, pp. 288–296.
[304] F. Shull, D. Cruzes, V. Basili, M. Mendonça, Simulating families of studies to build confidence in defect hypotheses, Inf. Softw. Technol. 47 (15) (2005) 1019–1032.
[305] F. Shull, M.G. Mendoncça, V. Basili, J. Carver, J.C. Maldonado, S. Fabbri, G.H. Travassos, M.C. Ferreira, Knowledge-sharing issues in experimental software engineering, Empirical Softw. Eng. 9 (1–2) (2004) 111–137.
[306] F. Shull, R. Turner, An empirical approach to best practice identification and selection: the us department of defense acquisition best practices clearinghouse, 2005 International Symposium on Empirical Softw. Eng., 2005. IEEE, 2005, pp. 8–pp.
[307] F.J. Shull, J.C. Carver, S. Vegas, N. Juristo, The role of replications in empirical software engineering, Empirical Softw. Eng. 13 (2) (2008) 211–218.
[308] J.E. Sieber, Protecting research subjects, employees and researchers: implications for software engineering, Empirical Softw. Eng. 6 (4) (2001) 329–341.
[309] A. Sigfridsson, A. Sheehan, On qualitative methodologies and dispersed communities: reflections on the process of investigating an open source community, Inf. Softw. Technol. 53 (9) (2011) 981–993.
[310] S.E. Sim, S. Easterbrook, R.C. Holt, Using benchmarking to advance research: A challenge to software engineering, Proceedings of the 25th International Conference on Softw. Eng. IEEE Computer Society, 2003, pp. 74–83.
[311] J. Singer, Using the american psychological association (apa) style guidelines to report experimental results, Proceedings of Workshop on Empirical Studies in Softw. Maintenance, (1999), pp. 71–75.
[312] J. Singer, N. Vinson, Empirical software engineering research ethics, Proceedings of the 6th International Symposium on Softw. Metrics, IEEE Computer Society, 1999, p. 326.
[313] J. Singer, N.G. Vinson, Ethical issues in empirical studies of software engineering, IEEE Trans. Softw. Eng. 28 (12) (2002) 1171–1180.
[314] D.I. Sjøberg, Knowledge acquisition in software engineering requires sharing of data and artifacts, Empirical Softw. Eng. Issues. Critical Assessment and Future Directions, Springer, 2007, pp. 77–82.
[315] D.I. Sjoberg, B. Anda, E. Arisholm, T. Dyba, M. Jorgensen, A. Karahasanovic, E.F. Koren, M. Vokác, Conducting realistic experiments in software engineering, Empirical Softw. Eng., 2002. Proceedings. 2002 International Symposium n, IEEE, 2002, pp. 17–26.
[316] D.I. Sjøberg, B. Anda, E. Arisholm, T. Dybå, M. Jørgensen, A. Karahasanović, M. Vokáč, Challenges and Recommendations When Increasing the Realism of Controlled Softw. Eng. Experiments, Empirical methods and studies in software engineering, Springer, 2003, pp. 24–38.
[317] D.I. Sjøberg, T. Dybå, B.C. Anda, J.E. Hannay, Building theories in software engineering, Guide to advanced empirical software engineering, Springer, 2008, pp. 312–336.
[318] D.I. Sjoberg, T. Dyba, M. Jorgensen, The future of empirical methods in software engineering research, 2007 Future of Softw. Eng. IEEE Computer Society, 2007, pp. 358–378.
[319] D.I. Sjøberg, J.E. Hannay, O. Hansen, V.B. Kampenes, A. Karahasanovic, N.-K. Liborg, A.C. Rekdal, A survey of controlled experiments in software engineering, IEEE Trans. Softw. Eng. 31 (9) (2005) 733–753.
[320] M. Skoglund, P. Runeson, Reference-based search strategies in systematic reviews,

Proceedings of the 13th international conference on Evaluation and Assessment in Softw. Eng. British Computer Society, 2009, pp. 31–40.
[321] D. Smite, C. Wohlin, R. Feldt, T. Gorschek, Reporting empirical research in global software engineering: a classification scheme, 2008 IEEE International Conference on Global Softw. Eng. IEEE, 2008, pp. 173–181.
[322] M. Solari, Identifying experimental incidents in software engineering replications, 2013 ACM/IEEE International Symposium on Empirical Softw. Eng. and Measurement, IEEE, 2013, pp. 213–222.
[323] M. Solari, S. Vegas, Classifying and analysing replication packages for software engineering experimentation, 7th International Conference on Product Focused Softw. Process Improvement (PROFES 2006)-Workshop Series in Empirical Softw. Eng. (WSESE). Amsterdam, Paises Bajos, (2006).
[324] M. Staples, M. Niazi, Experiences using systematic review guidelines, J. Syst. Softw. 80 (9) (2007) 1425–1437.
[325] K.-J. Stol, B. Fitzgerald, Uncovering theories in software engineering, Softw. Eng. (GTSE), 2013 2nd SEMAT Workshop on a General Theory of, IEEE, 2013, pp. 5–14.
[326] Y. Sun, Y. Yang, H. Zhang, W. Zhang, Q. Wang, Towards evidence-based ontology for supporting systematic literature review, Evaluation & Assessment in Softw. Eng. (EASE 2012), 16th International Conference on, IET, 2012, pp. 171–175.
[327] C. Thomson, M. Holcombe, Factors explaining external quality in 54 case studies of software development projects, Proceedings of the 13th international conference on Evaluation and Assessment in Softw. Eng. British Computer Society, 2009, pp. 75–78.
[328] F. Tomassetti, G. Rizzo, A. Vetro, L. Ardito, M. Torchiano, M. Morisio, Linked data approach for selection process automation in systematic reviews, Evaluation & Assessment in Softw. Eng. (EASE 2011), 15th Annual Conference on, IET, 2011, pp. 31–35.
[329] P. Tonella, M. Torchiano, B. Du Bois, T. Systä, Empirical studies in reverse engineering: state of the art and future trends, Empirical Softw. Eng. 12 (5) (2007) 551–571.
[330] M. Turner, B. Kitchenham, D. Budgen, O. Brereton, Lessons learnt undertaking a large-scale systematic literature review, Proceedings of Evaluation & Assessment in Softw. Eng. EASE 2008, 8 (2008).
[331] S. Vegas, N. Juristo, A. Moreno, M. Solari, P. Letelier, Analysis of the influence of communication between researchers on experiment replication, Proceedings of the 2006 ACM/IEEE international symposium on Empirical software engineering, ACM, 2006, pp. 28–37.
[332] J.M. Verner, J. Sampson, V. Tosic, N.A. Bakar, B.A. Kitchenham, Guidelines for industrially-based multiple case studies in software engineering, 2009 Third International Conference on Research Challenges in Information Science, IEEE, 2009, pp. 313–324.
[333] S. Viller, I. Sommerville, Coherence: an approach to representing ethnographic analyses in systems design, Human–Comput. Interact. 14 (1–2) (1999) 9–41.
[334] R.H. Von Alan, S.T. March, J. Park, S. Ram, Design science in information systems research, MIS Q. 28 (1) (2004) 75–105.
[335] J. Webster, R.T. Watson, Analyzing the past to prepare for the future: writing a literature review, MIS Q. (2002) xiii–xxiii.
[336] E.J. Weyuker, Empirical software engineering research-the good, the bad, the ugly, 2011 International Symposium on Empirical Softw. Eng. and Measurement, IEEE, 2011, pp. 1–9.
[337] R. Wieringa, Design science as nested problem solving, Proceedings of the 4th international conference on design science research in information systems and technology, ACM, 2009, p. 8.
[338] R. Wieringa, A unified checklist for observational and experimental research in software engineering (version 1), Technical Report, Centre for Telematics and Information Technology, University of Twente. TR-CTIT-12-07, 2012.
[339] R. Wieringa, N. Condori-Fernandez, M. Daneva, B. Mutschler, O. Pastor, Lessons learned from evaluating a checklist for reporting experimental and observational research, Proceedings of the ACM-IEEE international symposium on Empirical software engineering and measurement, ACM, 2012, pp. 157–160.
[340] R. Wieringa, M. Daneva, N. Condori-Fernandez, The structure of design theories, and an analysis of their use in software engineering experiments, 2011 International Symposium on Empirical Softw. Eng. and Measurement, IEEE, 2011, pp. 295–304.
[341] R. Wieringa, H. Heerkens, Designing requirements engineering research, Comparative Evaluation in Requirements Eng., 2007. CERE'07. Fifth International Workshop on, IEEE, 2007, pp. 36–48.
[342] R. Wieringa, H. Heerkens, B. Regnell, How to write and read a scientific evaluation paper, 2009 17th IEEE International Requirements Eng. Conference, IEEE, 2009, pp. 361–364.
[343] R. Wieringa, N. Maiden, N. Mead, C. Rolland, Requirements engineering paper classification and evaluation criteria: a proposal and a discussion, Requirements Eng. 11 (1) (2006) 102–107.
[344] C. Wohlin, Guidelines for snowballing in systematic literature studies and a replication in software engineering, Proceedings of the 18th International Conference on Evaluation and Assessment in Softw. Eng. ACM, 2014, p. 38.
[345] C. Wohlin, Writing for synthesis of evidence in empirical software engineering, Proceedings of the 8th ACM/IEEE International Symposium on Empirical Softw. Eng. and Measurement, ACM, 2014, p. 46.
[346] C. Wohlin, A. Aurum, Towards a decision-making structure for selecting a research design in empirical software engineering, Empirical Softw. Eng. 20 (6) (2015) 1427–1455.
[347] P. Woodall, P. Brereton, Conducting a systematic literature review from the perspective of a ph. d. student, Proceedings of the 10th international conference on Evaluation and Assessment in Softw. Eng. British Computer Society, 2006. 130–130
[348] H.K. Wright, M. Kim, D.E. Perry, Validity concerns in software engineering research, Proceedings of the FSE/SDP workshop on Future of software engineering research, ACM, 2010, pp. 411–414.
[349] J.L. Wynekoop, N.L. Russo, Studying system development methodologies: an examination of research methods, Inf. Syst. J. 7 (1) (1997) 47–65.
[350] F. Xia, On the vital role and difficulty of definition and measurement for the validation of s.e. theory, Proceedings of the 6th International Conference on Empirical Assessment in Softw. Eng. (2002), pp. 1–9.
[351] S. Xu, V. Rajlich, Dialog-based protocol: an empirical research method for cognitive activities in software engineering, 2005 International Symposium on Empirical Softw. Eng., 2005. IEEE, 2005, pp. 10–pp.
[352] A. Yamashita, L. Moonen, Assembling multiple-case studies: potential, principles and practical considerations, Proceedings of the 18th International Conference on Evaluation and Assessment in Softw. Eng. ACM, 2014, p. 36.
[353] M.V. Zelkowitz, D. Wallace, Experimental validation in software engineering, Inf. Softw. Technol. 39 (11) (1997) 735–743.
[354] M.V. Zelkowitz, D.R. Wallace, Experimental models for validating technology, IEEE Comput. 31 (5) (1998) 23–31.
[355] H. Zhang, M.A. Babar, On searching relevant studies in software engineering, IEEE, 2010 http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=675630&isnumber=14870.
[356] H. Zhang, M.A. Babar, An empirical investigation of systematic reviews in software engineering, 2011 International Symposium on Empirical Softw. Eng. and Measurement, IEEE, 2011, pp. 87–96.
[357] H. Zhang, M.A. Babar, X. Bai, J. Li, L. Huang, An empirical assessment of a systematic search process for systematic reviews, Evaluation & Assessment in Softw. Eng. (EASE 2011), 15th Annual Conference on, IET, 2011, pp. 56–65.
[358] H. Zhang, M.A. Babar, P. Tell, Identifying relevant studies in software engineering, Inf. Softw. Technol. 53 (6) (2011) 625–637.

Update

Information and Software Technology
Volume 110, Issue , June 2019, Page 190

DOI: https://doi.org/10.1016/j.infsof.2019.03.011

Information and Software Technology 110 (2019) 190

Contents lists available at ScienceDirect

Information and Software Technology

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

Corrigendum to "CERSE-Catalog for empirical research in software engineering: A Systematic mapping study" [Information and Software Technology 105 (2019) 117–149]

Jefferson Seide Molléri, Kai Petersen, Emilia Mendes

BTH - Blekinge Tekniska Högskola Sweden

The author regrets a duplicated entry in the reference list. The correct reference for entry [289] is:

289. Yvonne Dittrich, Kari Rönkkö, Jeanette Eriksson, Christina Hansson, and Olle Lindeberg. Cooperative method development. *Empirical Software Engineering*, 13(3):231–260, Jun 2008.

The mistake affects the information provided regarding action research in Section 4.3 and Appendix A2.2. The methodological support in [289] is the only instrument we identify aiming the support for action research according to particular software engineering needs.

The author would like to apologize for any inconvenience caused.

DOI of original article: 10.1016/j.infsof.2018.08.008

https://doi.org/10.1016/j.infsof.2019.03.011