Performing systematic literature review in software engineering

Zlatko Stapić
Faculty of Organization and Informatics
University of Zagreb
Pavlinska 2, 42000 Varaždin, Croatia
zlatko.stapic@foi.hr

Eva García López, Antonio García Cabot
Luis de Marcos Ortega
Computer Science Department
University of Alcalá
Ctra. Barcelona km 33.6, 28871 Alcalá de Henares (Madrid), Spain
{eva.garcial, a.garciac, luis.demarcos}@uah.es

Vjeran Strahonja
Faculty of Organization and Informatics
University of Zagreb
Pavlinska 2, 42000 Varaždin, Croatia
vjeran.strahonja@foi.hr

Abstract. *In the field of software engineering (SE) the first primary studies about Systematic Literature Review (SLR) have been conducted during the last decade. By adapting guidelines from other disciplines Kitchenham in 2004 created guidelines on performing the SLR in a field of SE. Since then the SLR is used in SE regularly, but many obstacles still remain.*

This paper aims to give an short but complete overview of all phases and stages that should be undertaken in SLR, and although based on mentioned guidelines, this paper takes into consideration feedback and experiences reported by other authors and discusses the possible approaches along with advantages and dissadvantages of the method.

Keywords. systematic literature review, SLR, overview, software engineering

1 Introduction

In order to perform comprehensive and thorough analysis of existing research about a given topic, a systematic approach should be undertaken and existing methodologies should be reviewed in such a manner which will result in a solid basis for the rest of the research. Such analysis could be undertaken by performing different methods and approaches, such as systematic literature review, systematic mapping studies, tertiary reviews discussed by [1], or narrative review, conceptual review, rapid review and several other types presented by [2].

In general, whenever a literature review is performed it could be done by a systematic (following stated procedures and steps) or unsystematic (just reading and taking notes) approach. There are different reasons to perform *systematic literature review* (SLR). The most common reason is to summarize the existing evidence concerning a treatment or a technology. SLR could also be used to identify any gaps in current research in order to suggest areas for further investigation or to provide a framework/background in order to appropriately position new research activities. In addition, there are other general reasons to use a systematic rather than unsystematic approach, such as the purpose of the research, the scientific approach, the quality expectations or the existence of previous researches on the selected topic.

According to [3] the key feature that distinguishes SLR from traditional narrative reviews lies in their explicit attempt to minimize the chances of making wrong conclusions which could be the result of biases either in primary studies or in the review process itself.

The summary of possible activities that should be undertaken in order to perform a systematic literature review in the field of *software engineering* (SE) is presented in this paper and is based on guidelines presented in [1] as well as on additional discussions from other authors which are cited in the text.

The paper is divided into following sections. Section 2 gives an overview of the phases and activities of SLR, the section 3 focuses on a SLR performed by a single researcher (e.g. PhD student), section 4 discusses advantages and disadvantages of SLR and section 5 concludes the paper.

2 Systematic literature review in SE

The method that has been more widely used for different analysis in the field of software engineering is Systematic Literature Review (SLR). "A systematic literature review is a means of evaluating and interpreting all available research relevant to a particular research question, topic area, or phenomenon of interest. Systematic reviews aim to present a fair evaluation of a research topic by using a

trustworthy, rigorous, and auditable methodology" [1].

2.1 The history of SLR

The guideline for systematic reviews that aimed to help software engineering researchers was proposed by Kitchenham in [4] and was created as adaptation of several existing guidelines from other disciplines, mainly medicine. Although the three proposed phases of systematic review (planning the review, conducting the review and reporting the review) in general were not criticized, some authors [5] [6] [7] found that Kitchenham described them at a relatively high level and partially inappropriate to conduct for researchers in the field of software engineering. In favor of this goes the fact that Kitchenham in 2007 published a new version of technology report [1] with the aim to propose new and more comprehensive guidelines of performing a systematic literature review for researchers and PhD students in the field of software engineering. The basis for this guideline remained the same: existing guidelines used by medical researchers, but was reinforced by several books and discussions with researches from other fields.

2.2 Performing SLR in software engineering

General steps of the methodology of SLR are defined by [1] and are as follows:

Phase 1: Planning the review
- Identification of the need for a review
- Commissioning a review (optional)
- Specifying the research question(s)
- Developing a review protocol
- Evaluating the review protocol (recommended)

Phase 2: Conducting the review
- Identification of research
- Selection of primary studies
- Study quality assessment
- Data extraction and monitoring
- Data synthesis

Phase 3: Reporting the review
- Specifying dissemination mechanisms
- Formatting the main report
- Evaluating the report (recommended)

Above mentioned stages and phases are not literally sequential as some of them could be repeated more than once and might involve iteration or reimplementation. Additionally, a negative evaluation result means a repetition of the part or of the whole process.

It is important to notice that even experienced scientists often have to change or adapt review protocol. For some authors this is the reason to criticize the methodology of existing reviews of not being completely objective or even conducting a fake rational design process. On the other side, there are authors, for example [7], who discuss the need of the protocol even if it is a subject of constant changes through the whole systematic review process. General and strong conclusion of the authors who performed a SLR is that the protocol is needed and that it increases the quality of the process.

2.3 Planning the review

The results of this phase should be a clearly defined review protocol containing the purpose and the procedures of the review. The summary of each step of this phase is presented below.

Identification of the need for a review arises from the preliminary research in the topic area. It is important to identify and review existing systematic reviews on the same topic. The review of existing SLRs is usually undertaken against appropriate and previously created evaluation criteria. The most common practice is to create a checklist or set of questions that should be examined for every existing SLR. There are several checklists proposed by different authors or organizations, such as [8], and depending on the level of complexity they usually use concepts of quality of defined inclusion and exclusion criteria or the level of literature and relevant studies coverage along with assessment of quality of included studies.

Commissioning a review is an optional task and whether or not will be performed is determined by the type and the stakeholders of the review process. When performing this step, the organization must provide a commissioning document (that will contain all important information about the required work such as project name, review questions, timetable and budget) or dissemination strategy.

Specifying the research questions is probably most important part of systematic review process as it is the base for all other activities. The research question will define which primary studies include or exclude from review, it will define data that should be extracted from the reviewed literature. The defined research question should be answered in the final systematic literature review report. There are several types of research questions that could be stated in the domain of software engineering. These questions may concern, for example, effect of software engineering technology, cost and risk factors, the impact of technology on different concepts et cetera, and the type of a question sometimes can determine the guidelines and procedures to be used. According to Kitchenham, it is important to create a right question, i.e. the question that is meaningful and important to practitioners and researchers, that will lead either to changes in current SE practice or to increased confidence in the value of current practice, or that will identify discrepancies between commonly held beliefs and reality. Finally, the right questions could be the questions that are primarily of interest to researchers

in order to identify and scope the future research activities.

Developing a review protocol is considered as the most important activity of the whole planning phase as it determines the rest of the SLR process. The output of this activity should be a detailed review protocol that specifies the methods that will be used to perform a planned systematic review.

Creating a protocol prior to systematic review is necessary to reduce the possibility of researcher bias. Staples and Niazi [7] claim that review protocol, as a concrete and formal plan of the systematic review, usually insinuates and suggests the structure of the final report. Protocol should also describe the background context for the research, the specific research questions, the planned search strategy, criteria for publication selection, the treatment of publication quality assessment, the data extraction plan, the data synthesis plan and a project plan. In general, these parts are definition of whole systematic review process, and usually it is impossible to predict all elements and obstacles in it. That is why some authors, for example [7], discuss that a protocol is a subject of constant changes through the whole systematic review process, but any changes made on protocol during the execution phases must be documented and properly reported.

In the guidelines, Kitchenham suggests that aspects of the protocol should be piloted during its development. In particular, the search terms, selection criteria, and data extraction procedures should be tried out before finalizing the protocol. In order to make the protocol definition activities less difficult, Biolchini [5] created a review protocol template which could be used in this stage.

Evaluating the review protocol is not necessary, but is a recommended step in the SLR process in order to improve its quality as the protocol is a critical element of any systematic review. The researchers must take into consideration several aspects (purpose of the research, desired quality, time, financial construction etc.) in order to agree with a procedure for evaluating the protocol.

In regards to these, there are several methods of evaluating a review protocol which could be used, such as author's review, peer review, review by supervisor, review by external experts or test of protocol execution.

1. Question Formularization
   - 1.1. Question Focus
   - 1.2. Question Quality and Amplitude
     - Problem
     - Question
     - Keywords and Synonyms
     - Intervention
     - Control
     - Effect
     - Outcome Measure
     - Population
     - Application
     - Experimental Design
2. Sources Selection
   - 2.1. Sources Selection Criteria Definition
   - 2.2. Studies Languages
   - 2.3. Sources Identification
     - Sources Search Methods
     - Search String
     - Sources List
   - 2.4. Sources Selection after Evaluation
   - 2.5. References Checking
3. Studies Selection
   - 3.1. Studies Definition
     - Studies Inclusion and Exclusion Criteria Definition
     - Studies Types Definition
     - Procedures for Studies Selection
   - 3.2. Selection Execution
     - Initial Studies Selection
     - Studies Quality Evaluation
     - Selection Review
4. Information Extraction
   - 4.1. Information Inclusion and Exclusion Criteria Definition
   - 4.2. Data Extraction Forms
   - 4.3. Extraction Execution
     - Objective Results Extraction
       - **i)** Study Identification
       - **ii)** Study Methodology
       - **iii)** Study Results
       - **iv)** Study Problems
     - Subjective Results Extraction
       - **i)** Information through authors
       - **ii)** General Impressions and Abstractions
   - 4.4. Resolution of divergences among reviewers
5. Results Summarization
   - 5.1. Results Statistical Calculus
   - 5.2. Results Presentation in Tables
   - 5.3. Sensitivity Analysis
   - 5.4. Plotting
   - 5.5. Final Comments
     - Number of Studies
     - Search, Selection and Extraction Bias
     - Publication Bias
     - Inter-Reviewers Variation
     - Results Application
     - Recommendations

Figure 1 - Protocol template proposed by Biolchini et al. in [5]

2.4. Conducting the review

According to Kitchenham's guidelines, conducting the review phase consists of five obligatory stages:

Identification of research is a first step in conducting a review that will result in a list of entire population of publications relevant to the research questions and obtained by performing a search strategy. The search strategy should be the same as

stated in the review protocol, and it should be stated in such a manner that it allows the study to be replicable and open to external review. It is also good to break down a research question and to identify initial search strings according to population, intervention, comparison, outcomes, context and study design. Additionally, it is important to create a list of synonyms, abbreviations and alternative spellings. Process of performing a SLR must be transparent and replicable. This means that the whole process should be properly documented: the review must be documented, the search should be documented and unfiltered search results should be saved and retained for possible reanalysis, and any changes in review protocol made during this or any other subsequent stage should also be documented. An important prerequisite for repeatability of research is that the data that are the subject of analysis are time-stamped. Many of these documents will not be presented in the final report but could also be published and a reference to them could be available in the final report.

In an attempt to perform an exhaustive search Brereton et al. [9] identified seven electronic sources as most relevant sources to Software Engineers (IEEExplore, ACM Digital Library, Google Scholar, Citeseer library, INSPEC, ScienceDirect and EI Compendex), and they also discuss about considering the use of additional sources (SpringerLink, Web of Science and SCOPUS) from publishers or bibliographical databases. Several authors also tried to identify a list of relevant journals and conferences in the field of software engineering, for example, [10] and [1] obtained a list of relevant journals and conferences.

Selection of primary studies is performed on all identified (potentially relevant) studies by applying an inclusion and exclusion criteria in order to assess their actual relevance. The selection criteria are also decided during the protocol definition but if necessary they can be redefined during this process. The identification of research will usually end up with a great number of articles that do not answer to the research question (because the keywords may have different meanings or may be used in the studies that are not in the focus of SLR research topic). The *inclusion criteria* will define which of these studies to include in the set of relevant ones, and the *exclusion criteria* could be applied on the already selected studies in order to identify those that do not meet additional conditions; or on the initial list of studies in order to remove irrelevant ones. Inclusion and exclusion criteria should be based on the research question, but could be defined based on study types. For example, only quantitative studies will be taken into consideration.

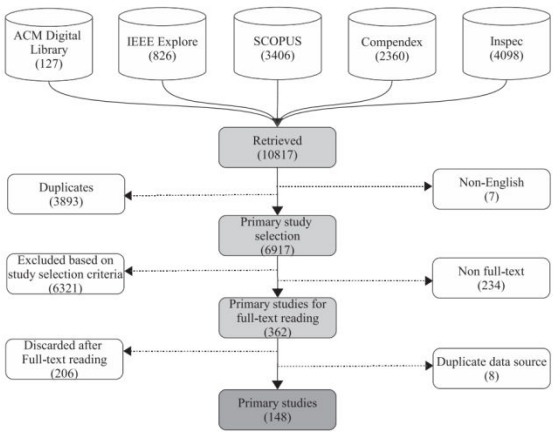

Figure 2 - Example of study selection process [3]

Study quality assessment is the second most important part of this phase. The idea of this process is to analyze and assess the quality of each primarily selected study in order to be finally included in data extraction and reporting process. In general, the aim of assessing the quality is to make sure that the study findings are relevant and unbiased. However, this is not a simple process as, according to Kitchenham, there is no agreed definition of study "quality". Some authors, for example [8], discuss that the study quality assessment procedures mainly depend on the type of the study. The following elements should be assessed regardless of the study type:

- Appropriateness of study design to the research objective.
- Risk of bias.
- Choice of outcome measure.
- Statistical issues.
- Quality of reporting and intervention.
- Generalizability.

The most usual tool (quality instrument) used to assess the quality of studies is *checklist*. Usage of checklists ensures that all assessed studies are evaluated critically and in a standardized way. According to [8] there are many different checklists and scales already available, and they can be used or adapted to meet the requirements of the review, but also a usual way is to create a new detailed checklist which will cover the bias and validity in the focus of specific research. In the literature several types of biases are recognized that should be addressed in a *checklist*. According to Kitchenham, checklist should also include consideration of biases and validity problems that can occur at the different stages of the study (design, conduct, analysis and conclusions). An accumulated list of questions was created in [1] and organized with respect to study stage and study type. The total of 59 questions is divided into four mentioned stages and cover quantitative empirical studies, correlation (observational) studies, surveys and experiments. This example checklist should not be used literally, but rather as a pool of questions in order to use, for specific study, only appropriate ones. The same process was conducted on qualitative studies, and resulted in 18 questions that could be used. Simple examples of quality assessment of SE studies are presented in [3], [11], [12], [13] and [14].

Data extraction and monitoring as a next stage aims to accurately and without bias record the appropriate information from selected papers. Researchers usually, during the protocol definition phase, define extraction forms which are used in this activity. The design of data extraction forms is not a trivial task while forms should be designed to collect all information needed to address the review questions and the study quality criteria. Basically, data extraction forms should contain questions needed to answer the review questions and quality evaluation criteria. There is no firm guidance on how to define these questions as they are different for every specific SLR process. On the other side, there are several elements that are considered to be common to all forms in order to provide standard information. According to Kitchenham these elements could be: name of the reviewer; date of data extraction; title, authors, journal, publications details; and space for additional notes. Combining the examples presented in [1] and [15] we can conclude that in general, data extraction form could include the following parts (sections): extraction information (data extractor, data checker and date of extraction), general study information (study identifier, title and publication details), questions to answer review questions, questions to assess study quality, and data summary. It is important not to include multiple studies with same data in a systematic review in order to avoid results with bias. This could be a serious threat if different sets of publications are analyzed by different researchers. On the other side, it is also important to contact the authors if it is identified that some data are missing or they were poorly reported. Examples on data extraction could be found in [14], [15] and [16].

Data synthesis is the final step in the review conduction phase. During this activity extracted data are collected and summarized. In general, there are two types of data synthesis: descriptive (narrative) synthesis and quantitative synthesis [8]. In order to draw reliable conclusions, synthesis should consider the strength of evidence, explore consistency and discuss inconsistencies. The synthesis approach should be defined by the protocol and is determined by the type of research questions, but also by the type of available studies and by the quality of data. For example, it is not wise to perform a statistical analysis on the numerical data if used publications are not randomized or do not cover whole population, or if there are studies with poor quality and with the biased results. In addition, according to guidance of Centre for Reviews and Dissemination (CRD) [8], narrative and quantitative approaches are not mutually exclusive, and according to [9] "software engineering systematic reviews are likely to be qualitative in nature". Regardless of the synthesis type, the synthesis should begin with a creation of a summary of included studies. The studies included in the review are usually presented in a table which covers all their important details (type, interventions, number and characteristics of participants, outcomes etc.). In the same (or in another) table could be also presented elements of study quality and risk of bias. Additionally, this descriptive process should be explicit rigorous and should help to conclude if the studies are similar and reliable to synthesize [8]. Kitchenham [1] also adds that the extracted data should be tabulated in a consistent manner with the review questions and structured to highlight similarities and differences between study outcomes. When systematic literature review includes quantitative and qualitative studies, Kitchenham suggests that researchers should "synthesize the quantitative and qualitative studies separately, and then attempt to integrate the results by investigating whether the qualitative results can help explain the quantitative results". When there is a considerable difference in the quality of studies, Kitchenham suggests the sensitivity analysis to be performed in order to determine if the low quality publications have significant impact on synthesis results. Sensitivity analysis could also be performed on different subsets of primary studies to determine the robustness of results. Examples of different methods and approaches of presentation of systematized data could be found in Chapter 1.3.5. of [8].

2.5. Reporting the review

The aim of the final phase of the systematic literature review process is to write the results of the review in a form suitable to dissemination channel and target audience or parties. The results are usually written in a form of a systematic review report.

Specifying dissemination strategy and mechanisms is usually performed during the project commissioning activities, or if there was no commissioning phase, then dissemination strategy and mechanisms should be defined in the review protocol. Kitchenham discusses that apart from disseminating the results in academic journals and conferences, scientists should consider performing other dissemination activities that might include direct communication with affected bodies, publishing the results on web pages, posters or practitioner-oriented magazines etc. If the results are to be published in a conference or journal, or any other publication with restricted number of pages, then the reference to a document (technical report, PhD thesis or similar) that contains all information should be provided.

Formatting the main report is a stage in which Kitchenham adopted the suggested structure of systematic review report given in CRD's guidelines from 2001. Although the mentioned original guidelines are updated in [8], the version presented by Kitchenham is sufficient in the field of software engineering.

Evaluating the report is the final step in the systematic literature review process. This activity depends mainly on the type of the publication. Papers

submitted to the scientific conference or scientific journal will be reviewed by independent peer reviewers. Doctoral dissertations will be reviewed by supervisors and by the committee during the examination process. Finally, if the publication is a technical review, it is also advisable to subject the materials to an independent evaluation. In this case, this final review could be done by the same expert panel that was created to review the research protocol. The results of the review, if negative, could require repetition of one or more phases in the systematic literature review process.

3 Light SLR

Described *Systematic Literature Review* process is defined in such way to be undertaken by large group of researchers, and this raises the question if these guidelines could be used while single researcher (like PhD student) performs SLR. The authors of the guidelines conclude that not all of the mentioned stages (activities) are obliged to be performed if a PhD student is performing a research, so they proposed a light version of the SLR process containing the most important steps:

- Developing a protocol
- Defining the research question(s).
- Specifying what will be done to address the problem of a single researcher applying inclusion/exclusion criteria and undertaking all the data extraction.
- Defining the search strategy.
- Defining the data to be extracted from each primary study including quality data.
- Maintaining lists of included and excluded studies.
- Using the data synthesis guidelines.
- Using the reporting guidelines.

Although the version of the process is *light*, the mentioned tasks should be performed according to the guidelines and should take into consideration the notes guidelines' authors gave for PhD students and single researchers.

4 Discussion

As every other method and approach, SLR also has several advantages and disadvantages. Kitchenham identified three main groups of advantages of using systematic literature review. (1) The methodology is well-defined; (2) it enables researchers to provide the information available in the wide range of sources; (3) and in the case of quantitative data, it is possible to perform some meta-analysis and to extract information that single study cannot provide [1]. Additionally, if compared to unstructured methods, like simple literature review, the SLR has many advantages (described in the SLR process) that make the results of such analysis more reliable and more likely to be unbiased.

On the other hand, major disadvantage of this approach is that it requires much more effort and time in comparison to simple literature review and this is exacerbated by large number of review points: search term pilot reviews, protocol reviews, initial selection reviews, final selection reviews, data extraction reviews, and data analysis reviews [7]. Kitchenham also adds that the usage of meta-analysis could be a disadvantage as it can detect small and unimportant biases.

Biolchini [5] states that a conduction of SLR in software engineering is much harder than in other disciplines, for example medicine. Same authors point out that the overall process is difficult to conduct, especially the activities of protocol development, searching and evaluating studies.

Additionally, execution of this method depends on solid literature coverage of the focused phenomenon, and subsequently it cannot be used to explore new, revolutionary, phenomena which are not well covered in literature.

Finally, even experienced authors are likely to change the review protocol during the implementation phase, and that brings the problem of documenting the whole process.

5 Conclusion

The method overview presented in this paper is based on the Kitchenham's guidelines presented in [1] and is expanded by reported feedback of the researchers, mainly from the field of software engineering.

The process of systematic literature review is not easy to perform, but the general opinion of the authors who have performed SLR is that this method is useful and could be used to decrease the biases and to increase the review quality. They also note that the usage of this method has significant obstacles in the field of software engineering in comparison to other fields, for example, the field of health sciences.

In order to overcome the obstacles specific for SE field, it is important to keep in mind that the scope of the review should be limited by choosing clear and narrow research questions and that the whole process should be in advance well defined by putting an considerable efforts in creation of feasible review protocol.

SLR can be a monolithic batch process, that periodically provides reports and digests but in terms of scientific research, much closer is a view on SLR as an interactive and iterative process.

As SLR method still emerges in the field of software engineering, the idea of publishing the replications of existing systematic reviews is welcomed, along with the idea of creation of centralized index of existing literature reviews.

Acknowledgments

The research presented in this paper was conducted as a part of a PhD research and is partially founded by the scholarship granted to the first author from Croatian Science Foundation.

References

[1] B. Kitchenham and S. Charters, "Guidelines for performing Systematic Literature reviews in Software Engineering Version 2.3," Keele University and University of Durham, Technical report EBSE-2007-01, 2007.

[2] M. Petticrew and H. Roberts, *Systematic reviews in the social sciences: a practical guide*. Malden, MA: Blackwell Pub., 2005.

[3] T. Dybå and T. Dingsøyr, "Strength of evidence in systematic reviews in software engineering," 2008, pp. 178–187.

[4] B. Kitchenham, "Procedures for Performing Systematic Reviews," Software Engineering Group; National ICT Australia Ltd., Keele; Eversleigh, Technical report Keele University Technical Report TR/SE-0401; NICTA Technical Report 0400011T.1, 2004.

[5] J. Biolchini, P. Gomes Mian, A. Candida Cruz Natali, and G. Horta Travassos, "Systematic Review in Software Engineering," PESC, Rio de Janeiro, Technical report RT - ES 679 / 05, 2005.

[6] P. Mian, T. Conte, A. Natali, J. Biolchini, and G. Travassos, "A Systematic Review Process for Software Engineering," in *ESELAW '05: 2nd Experimental Software Engineering Latin American Workshop*, 2005.

[7] M. Staples and M. Niazi, "Experiences using systematic review guidelines," *Journal of Systems and Software*, vol. 80, no. 9, pp. 1425–1437, 2007.

[8] Centre for Reviews and Dissemination, University of York, *Systematic reviews: CRD's guidance for undertaking reviews in health care.* York: Centre for Reviews and Dissemination, 2009.

[9] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *Journal of Systems and Software*, vol. 80, no. 4, pp. 571–583, 2007.

[10] J. Hannay, D. Sjoberg, and T. Dyba, "A Systematic Review of Theory Use in Software Engineering Experiments," *IEEE Transactions on Software Engineering*, vol. 33, no. 2, pp. 87–107, 2007.

[11] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering – A systematic literature review," *Information and Software Technology*, vol. 51, no. 1, pp. 7–15, 2009.

[12] B. Kitchenham, R. Pretorius, D. Budgen, O. Pearl Brereton, M. Turner, M. Niazi, and S. Linkman, "Systematic literature reviews in software engineering - A tertiary study," *Information and Software Technology*, vol. 52, no. 8, pp. 792–805, 2010.

[13] B. Kitchenham, P. Brereton, M. Turner, M. Niazi, S. Linkman, R. Pretorius, and D. Budgen, "The impact of limited search procedures for systematic literature reviews - A participant-observer case study," 2009, pp. 336–345.

[14] M. Unterkalmsteiner, T. Gorschek, A. K. M. M. Islam, C. K. Cheng, R. B. Permadi, and R. Feldt, "Evaluation and Measurement of Software Process Improvement - A Systematic Literature Review," *IEEE Transactions on Software Engineering*, 2012.

[15] M. Jørgensen, "Estimation of Software Development Work Effort:Evidence on Expert Judgment and Formal Models," *International Journal of Forecasting*, vol. 23, no. 3, pp. 449–462, 2007.

[16] T. Dybå and T. Dingsøyr, "Empirical studies of agile software development: A systematic review," *Information and Software Technology*, vol. 50, no. 9–10, pp. 833–859, 2008.