Second-Generation Systematic Literature Studies using Snowballing

Claes Wohlin
Blekinge Institute of Technology
SE – 371 79 Karlskrona
Sweden
+46-(0)455-385820
claes.wohlin@bth.se

ABSTRACT
Systematic literature studies have become standard practice in software engineering to synthesize evidence in different areas of the discipline. As more such studies are published, there is also a need to extend previously published systematic literature studies to cover new research papers. These first extensions become second-generation systematic literature studies. It has been asserted that snowballing would be a suitable search strategy for these types of second-generation studies, since newer studies ought to refer to previous research on a topic, and in particular to systematic literature studies published in an area. This paper compares using a snowballing search strategy with a published second-generation study using a database search strategy in the area of cross-company vs. within-company effort estimation. It is concluded that the approaches are comparable when it comes to which papers they find, although the snowballing approach is more efficient in this particular case.

CCS Concepts
• Software and its engineering→Software creation and management→Software development process management.

Keywords
Empirical research methods; systematic literature reviews; snowballing.

1. INTRODUCTION
Systematic literature reviews and systematic mapping studies have been established as a way to synthesize evidence in software engineering. Here, "systematic literature studies" is used as a collective term for systematic literature reviews and systematic mapping studies. Systematic literature studies form an important basis for the concept of evidence-based software engineering. Inspired by medicine, the concept of evidence-based software engineering was introduced by Kitchenham et al. [3]. In a similar way Webster and Watson [10] describe the use of systematic literature studies in information systems research.

EASE '16, June 01 - 03, 2016, Limerick, Ireland

ACM 978-1-4503-3691-8/16/06…$15.00
DOI: http://dx.doi.org/10.1145/2915970.2916006

To support systematic literature studies, guidelines for conducting systematic literature reviews have been developed [11] and [13]. Furthermore, systematic mapping studies have been highlighted as a complement to systematic literature reviews [8] and Petersen et al. published an update of the guidelines for conducting systematic mapping studies [9]. Kitchenham et al. [6] discuss the use of systematic mapping studies as a starting point for further research, and hence position the different types of literature studies relative to each other. The guidelines by Kitchenham and Charters [4] use database searches as the main form of searching the literature, although other forms are highlighted as complementary. As a complement Wohlin [13] proposed guidelines for conducting the literature search using snowballing, including both backward and forward snowballing.

As systematic literature studies have become more established and numerous studies have been published, the need to extend systematic literature studies has also increased. Extended studies refer to newer generation systematic literature studies on the same topic. However, "extended" studies may be misinterpreted to imply extending a study with a complementary perspective on a previous systematic literature study. Furthermore, in the future there may be a need to have extensions of already extended systematic literature studies, which means that the notion of an extension becomes cumbersome. Based on this, it is here chosen to introduce the concept of newer generation systematic literature studies to denote an extension in time of a systematic literature study, and the actual generation refers to the number of extension. For example, Kitchenham et al. [5] published a systematic literature review on cross- vs. within-company cost estimation, and Mendes et al. [7] published a second-generation study. The original study covered publications up until 2005 and the second-generation study extended the time frame to include studies published between 2006-2013.

The second-generation study by Mendes et al. [7] used the same method for identifying relevant studies to include as the original study (first-generation). However, Wohlin [13] asserted in the guidelines for snowballing search that snowballing would be particular beneficial to use in a second-generation study. More specifically, Wohlin wrote:

"In particular, it should be noted that snowballing is particularly useful for extending a systematic literature study, since new studies almost certainly must cite at least one paper among the previously relevant studies or the systematic study already conducted in the area. Thus, snowballing is by deduction a better approach than a database search for extending systematic literature studies. The actual evidence for this assertion is left for further research."

Thus, the objective here is to evaluate this assertion by replicating the second-generation study by Mendes et at. [7]. The remainder of the paper is structured as follows. Related work is presented in Section 2, while Section 3 outlines the research method used in the snowballing approach. In Section 4, the results are reported and the paper is concluded in Section 5.

2. RELATED WORK

As mentioned above, the guidelines for systematic literature reviews [4] take database searches as a starting point. However, the guidelines do mention that complementary searches are needed too. The complementary searches include for example: reference lists, grey literature, specific research outlets (journals or conferences) and researchers in the field. Unfortunately, it is quite common that systematic literature studies are only based on one type of search, and hence the first method chosen plays an important role. The lack of complementary searches is understandable given the workload to conduct a thorough systematic literature study. Database searches are challenging for many reasons, including selection of databases, different interfaces for the databases, different ways of constructing search strings, different search limitations in the databases and identification of synonyms of terms used. This leads to two conclusions: 1) the first step in the search strategy often becomes the only step, i.e. database search (if following the guidelines [4]), and 2) we may miss important literature given the challenges with database searches.

Jalali and Wohlin [2] compared the use of database search vs. snowballing in the area of agile software development in a global context. The results from the two different approaches to searching the literature are comparable, i.e. both approaches resulted in a reasonable sample of the relevant literature. Although it is argued that systematic literature reviews should cover all relevant literature on a topic, it is not uncommon that only a sample will be found [12]. Furthermore, Badampudi et al. used snowballing as the first search method and then complemented it with a database search to evaluate the "goodness" of the snowballing approach. Experiences from using these two search methods are presented in [1].

Thus, snowballing may be a good alternative for the first method in a systematic literature study. However, more importantly here, it may be a very attractive alternative when conducting a second-generation systematic literature study (or higher generation systematic literature study). Guidelines for snowballing in information systems research are presented in [10], a summary guideline for software engineering is outlined in [11], and more detailed guidelines can be found in [13]. The relevant steps for the objective of conducting a snowballing search for a second-generation systematic literature study are summarized in here.

The first step is to formulate the research questions, which for a second-generation study means copying the research questions from the first-generation systematic literature study.

The second step in a snowballing search is to identify the start set from which the snowballing should be conducted. Searching, for example, with some suitable keywords in Google Scholar is one potential way of identifying a suitable start set. Google Scholar may be a good alternative to avoid bias in favour of any specific publisher, journal or conference series. The actual start set is only those papers in the tentative start set that at the end are included in the systematic literature study. Alternatively, for a second-generation study, it should be possible to use the first-generation systematic literature studies and the primary studies identified in the first-generation studies as a start set.

The third step is to start the snowballing procedure, which in normal cases include both backward and forward snowballing. Backward snowballing means using the reference list of the identified papers; forward snowballing refers to identifying new papers based on those papers citing the paper being examined.

In the fourth step, each paper citing the paper being examined is studied. A first screening is conducted using the information available in Google Scholar. If this information is insufficient for a decision, the citing paper is studied in more detail. First, the abstract is studied, and if this is insufficient, the place citing the paper already included is examined. If this is insufficient, the full text is studied to make a decision regarding the new paper. The outcome of the third step is a decision whether to include or exclude a paper. It should be noted that only included papers will be used in the snowballing procedure, and hence it is important to decide on inclusion or exclusion of a paper.

3. RESEARCH METHOD

The research method is in summary an instantiation of the four steps described in related work. The research questions of the systematic literature study are as for the first-generation systematic literature study [5] and its replication [7]. The research method used here to replicate the second-generation study by Mendes et al. [7] is based on specific research questions for the comparison between search strategies and conducting steps 2-4 of the snowballing procedure.

3.1 Research Questions

The following research questions are formulated for the comparison between the two second-generation systematic literature studies:

RQ1: Do the two searches identify the same papers?

This question may be divided into three aspects. First, to evaluate whether the snowballing approach actually leads to including the same papers, and second, even if this is not the case, the question is whether the papers were among those being evaluated for inclusion or exclusion. There is of course a risk that a paper included in [7] is excluded in the snowballing conducted here even if the paper is found. This is a matter of a different judgment or the researcher making a mistake, and hence this should be taking into account in the evaluation. Third, to evaluate whether the snowballing approach results in finding any additional papers, i.e. papers not found in [7].

RQ2: Is it possible to state if one search method is more efficient than the other?

Unfortunately, very few researchers keep track of the actual time conducting a systematic literature study, and hence it is unlikely that it is possible to compare the efficiency in terms of work hours conducting the two studies. Thus, as an approximation the number of papers actually evaluated in the two studies is compared.

3.2 Identification of Start Set

Given the objective to evaluate the assertion about the positive aspects of using snowballing for second-generation literature studies, the start set is created from the following two sources:

1. Published first-generation systematic literature studies on the topic.
2. Papers included in the first-generation systematic literature study or studies.

The second-generation systematic literature review [7] is not included in the start set, since the main objective is to compare the second-generation study using snowballing with the study using the same research method as the first-generation study [5].

3.3 Forward Snowballing

Given the objective of conducting a second-generation literature study, only forward snowballing is considered. However, potentially backward snowballing could have been used from the papers identified in the forward snowballing. This is not done, since the assertion above is based on the assumption that any prior-generation systematic literature study or papers included in prior-generation systematic literature studies ought to have been cited in newer relevant papers.

Thus, forward snowballing is only done from the start set and no iterations are conducted as outlined in the guidelines for snowballing [13]. The motivation being that it is hard to imagine that a paper published in 2006 or later does not refer to any of the papers in the start set. This would mean that a paper is authored, reviewed and published without any knowledge of the previous research on the topic, or at least without acknowledging the prior research. This is highly unlikely, and hence to make the search efficient, snowballing is only conducted based on the start set.

The citations to the paper being examined are studied using Google Scholar. Patents and quotes are removed in Google Scholar, and only citations from articles are considered. Forward snowballing is conducted by examining papers published in the time frame 2006-2013 that cite the papers in the start set.

3.4 Inclusion and Exclusion of Papers

Inclusion and exclusion are based on the original research questions in [5]. The papers citing the papers in the start set are first examined in Google Scholar, and if in doubt about inclusion and exclusion the full abstract is read. If this is insufficient to decide on inclusion or exclusion, the place of citation is studied and then finally the full paper is gone through if needed.

4. RESULTS

4.1 Start Set

Two systematic literature studies on the topic are found. The first paper is a conference paper from 2006, which was then extended into a journal article published in 2007. In addition, 10 primary studies from the journal paper are also included in the start set. In the journal paper [5], they are denoted S1-S10 and in the replication using a snowballing search, [13], they are referred to as P1-P10, although not listed in exactly the same order in the two papers. In total, the start set includes 12 papers for which the citations should be studied. Here, the numbering of the papers is taken from [13], and the two systematic literature studies are denoted P11 and P12 respectively. The 12 papers are:

P1. Mendes, E. and Kitchenham, B. A. 2004. Further comparison of cross-company and within-company effort estimation models for web applications. In *Proc. International Symposium on Software Metrics*. 348-357.

P2. Kitchenham, B.A., and Mendes, E. 2004. A comparison of cross-company and within-company effort estimation models for web applications. In *Proc. 8th International Conference on Empirical Assessment in Software Engineering*, 47-55.

P3. Mendes, E., Lokan, C., Harrison, R., and Triggs, C. 2005. A replicated comparison of cross-company and within-company effort estimation models using the ISBSG database. In *Proc. International Symposium on Software Metrics*.

P4. Briand, L.C., El-Emam, K., Maxwell, K., Surmann, D. and Wieczorek, I. 1999. An assessment and comparison of common cost estimation models. In *Proc. of the 21st International Conference on Software Engineering*, 313-322.

P5. Briand, L.C., Langley, T. and Wieczorek, I. 2000. A replicated assessment of common software cost estimation techniques. In. *Proc. of the 22nd International Conference on Software Engineering*, 377-386.

P6. Jeffery, R., Ruhe M. and Wieczorek, I. 2000. A comparative study of two software development cost modeling techniques using multi-organizational and company-specific data. *Information and Software Technology*, 42, 1009-1016.

P7. Jeffery, R., Ruhe, M. and Wieczorek, I. 2001. Using public domain metrics to estimate software development effort. In *Proc. 7th International Software Metrics Symposium*, 16-27.

P8. Wieczorek, I. and Ruhe, M. 2002. How valuable is company-specific data compared to multi-company data for software cost estimation? In *Proc. 8th International Software Metrics Symposium*, 237-246.

P9. Lefley, M. and Shepperd, M. J. 2003. Using genetic programming to improve software effort estimation based on general data sets. In *Proc. of GECCO*, 2477-2487.

P10. Maxwell, K., Wassenhove, L.V. and Dutta, S. 1999. Performance evaluation of general and company specific models in software development effort estimation, *Management Science*, 45, 6, 787-803.

P11. Kitchenham, B.A., Mendes, E., and Travassos, G.H. 2006. A systematic review of cross- vs. within-company cost estimation studies. In *Proc. 10th International Conference on Evaluation and Assessment in Software Engineering*.

P12. Kitchenham, B.A., Mendes, E., and Travassos, G.H. 2007. Cross- vs. within-company cost estimation studies: A systematic review, *IEEE Transactions on Software Engineering*, 33, 5, 316-329.

4.2 Forward Snowballing

For each of the 12 papers, the following steps were taken:

- Search for the paper in Google Scholar (using the title of the paper), or find the Google Scholar page of one of the authors and identify the paper of interest in the publication list,
- Use the citation link to get all papers citing the paper,
- Remove ticks for patents and quotes, and then
- Change the time interval for citations to 2006-2013.

It turned out that for all 12 papers at least one of the authors has a Google Scholar page.

The total number of citations in the time interval 2006-2013, when the searches were conducted, was as follows:

P1 – 77; P2 – 45; P3 – 40; P4 – 179; P5 – 149; P6 – 92; P7 – 101; P8 – 32; P9 – 41; P10 – 38; P11 – 53 and P12 – 171.

The total number of citations becomes 794 for P1-P10, i.e. the primary studies, and 224 for the two systematic literature studies. Thus, in total 1018 citations were found in the forward snowballing using the Google Scholar pages of the authors (one per paper). The citation analysis was conducted in November 2015. Given that the analysis was done close to two years after the end of the time frame (2013), it was assumed that the numbers would be stable. However, some numbers have changed slightly since the analysis was done. This was observed when looking up

some of the citations in February 2015. However, these changes should not affect the study, since the changes should not affect the papers that can be found in from the database search; the regular databases (e.g. IEEE, ACM, Scopus and Springer) used in literature studies should be stable close to two years after the end of the investigated time interval. Thus, the results are based on the searches conducted in November 2015.

The forward snowballing from the two systematic literature studies were conducted first, since it was assumed that most relevant papers in the time frame 2006-2013 ought to refer to at least one of them. It is a little complicated by the following facts:

- The conference version covers papers up until 2004, and the journal version also includes papers from 2005. However, the difference is only one paper.
- The conference version is published in April 2006, and the journal version in May 2007. This implies that some papers published in 2006 and 2007 may be unable to refer to the journal version. Furthermore, some researchers may know about the conference version and since it fulfils their needs, they are not looking for another review paper. Thus, the conference version may be cited even if the journal version is available.

It is worth noting that there is a large overlap in citations. Several papers cite the same papers in the start set, and hence the actual workload was smaller than having to look at more than 1000 unique papers in Google Scholar. Furthermore, non-English papers and grey literature are quickly identified in the listing in Google Scholar. The papers were first evaluated looking at the information available in the listing in Google Scholar. If this was judged insufficient, the link from Google Scholar was used to obtain more information. Using the links in the web browser helped avoiding looking at the same paper several times, since the colour of links changed when the link has been used once. This made the search much more efficient than having to look at more than 1000 papers in Google Scholar.

At the end, 16 candidate papers were identified from the two systematic literature studies. No new candidate papers were found when looking at the 794 citations to the primary studies, i.e. P1-P10.

4.3 Inclusion and Exclusion

The 16 candidate papers were downloaded to more fully inspect the abstract, actual citation and if needed the full paper. The outcome of the inclusion and exclusion evaluation resulted in including 12 papers in total, and hence four papers were judged as not meeting the inclusion criteria. When dividing the 12 papers between the two systematic literature studies, the outcome was as follows[1]:

- Six papers are identified from P11 (conference paper).
- Seven papers are identified from P12 (journal article).

It should be noted that one of papers identified for P12 was also identified from P11. Thus, one paper refers to both P11 and P12. This means that 12 unique papers are identified from the forward snowballing from the two systematic literature studies. The 12 papers are:

[1] It should be noted that in [7], it is mentioned that all papers published after 2006 refer to P11. This is probably a mistake and it should be that they refer to P11 or P12.

P13. Lokan, C. and Mendes, E. 2006. Cross-company and single-company effort models using the ISBSG database: A further replicated Study, In *Proc. International Symposium on Empirical Software Engineering*, 75-84.

P14. Mendes, E., Di Martino, S., Ferrucci, F. and Gravino, C. 2007. Effort estimation: How valuable is it for a web company to use a cross-company data set, compared to using its own single-company data set? In *Proc. 16th International Conference on World Wide Web*, 963-972.

P15. Premraj, R. and Zimmermann, T. 2007. Building software cost estimation models using homogenous data. In *Proc. First International Symposium on Empirical Software Engineering and Measurement*, 393-400.

P16. Lokan, C. and Mendes, E. 2008. Investigating the use of chronological splitting to compare software cross-company and single-company effort predictions. In *Proc. 12th International Conference on Evaluation and Assessment in Software Engineering*, 136-145.

P17. Mendes, E., Di Martino, S., Ferrucci, F. and Gravino, C. 2008. Cross-company vs. single-company web effort models using the Tukutuku database: An extended study. *Journal of Systems and Software*, 81, 673-690.

P18. Mendes, E. and Lokan, C. 2008. Replicating studies on cross- vs. single-company effort models using the ISBSG database. *Empirical Software Engineering*, 13, 3-37.

P19. Ferrucci, F., Gravino, C., Di Martino, S. and Buglione, L. 2009. Estimating web application development effort employing COSMIC: A comparison between the use of a cross-company and a single-company dataset. In *Proc. 6th Software Measurement European Forum*, 77-89.

P20. Lokan, C. and Mendes, E. 2009. Using chronological splitting to compare cross- and single-company effort models: Further investigation. In *Proc. 32nd Australasian Computer Science Conference*, 47-54.

P21. Mendes, E. and Lokan, C. 2009. Investigating the use of chronological splitting to compare software cross-company and single-company effort predictions: A replicated study. In *Proc. 13th International Conference on Evaluation and Assessment in Software Engineering*, 11-20.

P22. Kocaguneli, E., Gay, G., Menzies, T., Yang, Y. and Keung, J. W. 2010. When to use data from other projects for effort estimation. In *Proc. IEEE/ACM International Conference on Automated Software Engineering*, 321-324.

P23. Ferrucci, F., Mendes, E. and Sarro, F. 2012. Web effort estimation: the value of cross-company data set compared to single-company data set. In *Proc. Predictive Models in Software Engineering*, 29-38.

P24. Minku, L. L. and Yao, X. 2012. Can cross-company data improve performance in software effort estimation? In *Proc. Predictive Models in Software Engineering*, 69-78.

It is noteworthy that most papers refer to either the conference paper or the journal article, and only one paper to both of them. This illustrates the need to look at citations to both versions of the systematic literature study. Furthermore, no additional paper is included based on citations to the primary studies published in 2005 or earlier. It illustrates that researchers do indeed build on systematic literature studies, at least based on the findings from this second-generation systematic literature study using snowballing.

4.4 Research Questions

In relation to RQ1, the following can be noted:

1. The second-generation study by Mendes et al. [7] identified 11 papers using the same database search as in the first-generation study [5].
2. The second-generation study using snowballing found 12 papers.
3. The two second-generation studies have nine papers in common.
4. The following two papers are included in [7], but not in the study reported here:
   - Kocaguneli, E. and Menzies, T. 2011. How to find relevant data for effort estimation? In *Proc. International Symposium on Empirical Software Engineering and Measurement*, 255-264. *Comment*: This paper does not refer to any of the primary studies published before 2006, i.e. P1-P10. However, the paper does refer to the journal version of the first-generation systematic literature study, i.e. P12.
   - Top, O. O., Ozkan, B., Nabi, M. and Demirors, O. 2011. Internal and external software benchmark repository utilization for effort estimation. In *Proc. Joint Conference 21st International Workshop on Software Measurement and 6th International Conference on Software Process and Product Measurement*, 302-307. *Comment*: This paper refers to five of the primary studies published before 2006, i.e. P1-P10, but not any of the first-generation systematic literature reviews, i.e. P11 or P12.
5. Three papers are not included in [7] that were judged as meeting the inclusion criteria here. These papers are: P15, P19 and P22. *Comment*: It is unknown whether these three papers were not found in the database search forming the basis for the study reported in [7], or if the authors decided to exclude the studies or if the studies were missed.

A more detailed study of papers P15, P19 and P22 provides the following:

For P15: The abstract of P15 gives the impression that the paper is primarily about cross-company cost models. However, in the paper the authors write: "To develop cross-company cost models to compare their prediction accuracy against company-specific cost models." Thus, the paper is judged to meet the inclusion criteria. This paper is available in the IEEE and ACM databases.

For P19: From the abstract it is quite clear that the paper considers both cross-company and company-specific data and compares the findings. In the abstract, it is stated: "This paper will present a further case study in such direction, presenting and discussing results from an empirical study carried out using data from an Italian single company dataset as well as from the public benchmarking repository ISBSG r10,". Thus, the paper is included. It is worth noting that a general search on Google does not provide any links to the regular databases used in database searches, and hence it is highly likely that it is possible to find this paper using a database search strategy.

For P22: This is a short paper. The abstract is quite clear about that the paper addresses cross-company vs. company-specific cost models. In the abstract, the authors state: "Collecting the data required for quality prediction within a development team is time-consuming and expensive. An alternative to make predictions using data that crosses from other projects or even other companies." Thus, this paper is included. The paper can be found in the ACM database.

To summarize with respect to RQ1, there is a large overlap in the studies identified using the two different search strategies. It is positive to note that all papers found in the database search in [7] were also found in the snowballing search, although the author of this paper excluded two of the papers in the screening process. The latter may have been avoided if having several researchers conducting the screening of papers. This is not seen as critical for the results, since the main objective is to compare the search strategies as such and not actually conducting a second-generation systematic literature study.

A possible explanation for the exclusion of these two papers is that only the title and first part of the abstracts are visible in Google Scholar, and the text visible was not sufficiently clear in motivating a closer look at the papers. Thus, the two papers did not make in the first screening in Google Scholar. If they would have made it, the inclusion is quite clear when actually looking at the full abstracts.

The fact that three more papers (P15, P19 and P22) were included when using snowballing may be due to several reasons as explained above. Alternatively, it may be because the author of this paper made a different judgment than the authors of [7]. Furthermore, P19 does not seem to be available in the standard databases and hence it is impossible to find using only a database search strategy.

The results related to RQ1 indicate that the papers found are mostly the same, and hence the results are comparable. The differences are probably more related to the subjective evaluation of the researchers than the actual search strategy. However, it is hard to know definitively if the papers found in snowballing were also found in the database search, although P19 was most likely not found in the database search.

Turning to RQ2, the question is about the efficiency in finding the relevant papers for a second-generation systematic literature study. The database searches in [7] resulted in 1641 candidate papers, which were reduced to 100 papers by two of the authors in a preliminary screening. The authors of the paper looked at these 100 papers and 11 papers were selected for inclusion. This should be compared with 1018 candidate papers in the snowballing search, and 16 papers being evaluated after the first screening. Thus, in this case, fewer papers are evaluated using snowballing than a database search strategy, but the outcomes are comparable. It is also worth noting that having papers from seven different databases, as is the case in [7], require quite some work to consolidate into one list and to identify duplicates among the papers found. This is much less of an issue in snowballing, since the links in Google Scholar are used, and the change in colour based on web pages already visited helps a lot to identify duplicates in snowballing. It has of course to be ensured that the pages have not been visited before for any other reason.

4.5 Discussion

As formulated in RQ1, the first evaluation criterion is the degree of overlap in papers identified. This is essential to be able to consider using a less well-established search strategy than a database search, or more precisely less well-established in software engineering. The guidelines for systematic literature reviews [4] have made database searches the standard starting point for systematic literature studies in software engineering. The

guidelines by Webster and Watson [10] in information systems research take snowballing as the starting point. The guidelines in [13] were intended to make snowballing more common in software engineering too. In particular, snowballing was asserted to be very suitable for second-generation systematic literature studies. The findings here support this assertion.

However, it is impossible to claim generalizability based on one comparison of two second-generation systematic literature studies. It should be noted that the replicated systematic literature review is a little special, since one of the authors is also an author of several of the primary study. However, it is not believed that this fact invalidates the findings. Anyway, more comparisons are needed, although it is logical that snowballing is a good alternative when conducting second-generation (or higher) systematic literature studies. It is highly unlikely that authors of newer papers do not refer to older papers and in particular do not refer to systematic literature reviews on the topic. Furthermore, even if authors of newer papers do not know about the previous work or ignore it, knowledgeable reviewers would comment on the weakness in related work. Finally, editors and program chairs would most likely not accept a paper that did not refer to any of the previous work on a specific topic. With this in mind, and the comparable results with respect to the papers found and the fact that the snowballing was more efficient in this case, snowballing is a competitive and viable option as the first search strategy used when conducting a second-generation systematic literature study.

5. CONCLUSIONS

As published systematic literature studies grow older, there is a need to update the evidence in an area. Thus, more second-generation systematic literature studies will be conducted and published, and later higher-generation studies. It has been asserted that snowballing ought to be a good search strategy for second-generation studies, due to the fact that newer research ought to refer either to first-generation systematic literature studies or at least some of the primary papers included in a first-generation study. Based on this assertion a snowballing search strategy was used to replicate a database-driven second-generation study using a database search strategy.

It is concluded that in this study, snowballing finds all papers included as primary studies based on a database search [7]. Furthermore, it finds some additional papers. In particular, one of these papers could not be found in the database search, since the proceedings are not in one of the major databases normally used for conducting systematic literature studies. Snowballing has potentially found some additional papers, although this is judged as very hard to evaluate in retrospect. Furthermore, the snowballing search strategy is judged to be more efficient, at least, in this case. Fewer papers had to be screened and duplicates were very easily identified from links on the web indicating that a specific paper has already been evaluated.

In summary, it is concluded that a snowballing search strategy is suitable to use when conducting second- and higher-generation systematic literature studies. In particular, this will be the case until primary studies actually are written for synthesis. Advice for writing for synthesis can be found in [14].

6. ACKNOWLEDGMENTS

I would like to express my sincere thanks to Deepika Badambudi, Samireh Jalali and Prof. Rafael Prikladnicki for our discussions and collaboration on snowballing as a search method in systematic literature studies in software engineering.

7. REFERENCES

[1] Badampudi, D., Wohlin, C. and Petersen, K. 2015. "Experiences from using snowballing and database searches in systematic literature studies. In *Proceedings 19th International Conference on Evaluation and Assessment in Software Engineering*, Article No. 17.

[2] Jalali, S. and Wohlin, C. 2012. Systematic literature studies: Database searches vs. backward snowballing. In *Proceedings 6th International Symposium on Empirical Software Engineering and Measurement*, 29-38.

[3] Kitchenham, B. A., Dybå, T. and Jørgensen, M. 2004. Evidence-based software engineering. *In Proceedings of 27th IEEE International Software Engineering Conference*, 273-281, IEEE Computer Society, 2004.

[4] Kitchenham B. A. and Charters S. 2007. *Guidelines for performing systematic literature reviews in software engineering*. Version 2.3, EBSE Technical Report, EBSE-2007-01, Keele University.

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

[6] Kitchenham B. A., Budgen, D. and Brereton, O. P. 2011. Using mapping studies as the basis for further research – a participant-observer case study. *Information and Software Technology* 53, 6, 638-651.

[7] Mendes, E., Kalinowski, M., Martins, D., Ferrucci, F. and Sarro, F. 2014. Cross- vs. within-company cost estimation studies revisited: an extended systematic review. In *Proceedings 18th International Conference on Evaluation and Assessment in Software Engineering*, 129-138.

[8] Petersen K., Feldt R., Mujtaba S. and Mattsson M. 2008. Systematic mapping studies in software engineering. In *Proceedings 12th International Conference on Evaluation and Assessment in Software Engineering*.

[9] Petersen K., Vakkalanka S. and Kuzniarz, L. 2015. Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology* 64, 8, 1-18.

[10] Webster, J. and Watson, R. T. 2002. Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly* 26, 2, xiii-xxiii.

[11] Wohlin C. and Prikladnicki R. 2013. Systematic literature reviews in software engineering. *Information and Software Technology* 55, 6, 919-920.

[12] Wohlin, C., Runeson, P., da Mota Silveira Neto, P. A., Engström, E., do Carmo Machado, I. and de Almeida, E. S. 2013. On the reliability of mapping studies in software engineering. *Journal of Systems and Software* 86, 10, 2594-2610.

[13] Wohlin, C. 2014. Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering. In *Proceedings 18th International Conference on Evaluation and Assessment in Software Engineering*, 321-330.

[14] Wohlin, C. 2014. Writing for Synthesis of Evidence in Empirical Software Engineering", *In Proceedings 8th International Symposium on Empirical Software Engineering and Measurement*, 46:1-4.