Protocol for Extending an existing Tertiary study of Systematic Literature
Reviews in Software Engineering

Barbara A. Kitchenham
O. Pearl Brereton
David Budgen

EPIC Technical Report
EBSE-2008-006

June 2008

Software Engineering Group
School of Computer Science and Mathematics
Keele University
Keele, Staffs
ST5 5BG, UK

and

Department of Computer Science
University of Durham
Science Laboratories
Durham,
DH1 3LE UK

Version 3

Abstract

This document is a protocol for extending a completed tertiary study that investigated the adoption of evidence-based software engineering and, in particular, the use of systematic literature reviews to aggregate software engineering primary studies. The original study was based on manual search of 13 journals and conference proceedings. This protocol defines a plan to extend the original tertiary study to include additional primary studies found by an electronic search of multiple digital libraries.
.

CONTENTS

1.   Background

Kitchenham et al. [2] recently completed a tertiary study aimed at investigating the adoption of evidence-based software engineering and in particular the use of systematic literature reviews to aggregate software engineering primary studies. The tertiary study was based on a manual search of 13 journals and conference proceedings (see Table 1) from 1st January 2004 to 30th June 2007

This study will extend the scope of the original study by means of an automated search of a variety of digital libraries. There are two main reasons for undertaking this study:
1. This study is being undertaken as a part of the EPSRC EPIC project which is investigating issues associated with the adoption and use of systematic literature reviews [1]. It will be organized as an EPIC case study investigating the value of manual and automated search processes ([1], RQ6) and the importance of broad search processes ([1], RQ3 and provisionally RQ2).
2. If the results of this study are available before September, 2008, they will be used to extend a journal paper describing the original tertiary study that has been accepted subject to amendments by Information and Software Technology.

Table 1 Sources used in original tertiary study

Source
Information and Software Technology (IST)
Journal of Systems and Software
IEEE Transactions on Software Engineering
IEEE Software
Communications of the ACM (CACM)
ACM Surveys
Transactions on Software Engineering Methods (TOSEM)
Software Practice and Experience
Empirical Software Engineering Journal (EMSE)
IEE Proceedings Software (now IET Software)
Proceedings International Conference on Software Engineering (ICSE 04, 05, 06,07)
Proceedings International Symposium of Software Metrics (Metrics04, Metrics05)
Proceedings International Symposium on Empirical Software Engineering (ISESE 04, 05, 05)

The research questions to be addressed by this study are exactly the same as those addressed by the original study:
- How much EBSE activity has there been since 2004?
- What research topics are being addressed?
- Who is leading EBSE research?
- What are the limitations of current research?

2 Roles and Responsibilities

This study will be the responsibility of an RA supported by a Supervisor (David Budgen). Other members of the EPIC team (Pearl Brereton, Barbara Kitchenham, Steve Linkman, Mark Turner, Mahmood Niazi) will act as additional members of the SLR team to provide support for the selection, quality evaluation and data extraction, and reporting process.

3 Search Process

The search process for this study will be based on an automated search of the following digital libraries covering the time period 1st January 2004-30th June 2007:

- ACM
- IEEE
- SCOPUS
- Google Scholar
- Web of Science
- SpringerLink

3.1 Search Strings

The RA will develop the search strings used for the automated search using the following terms:

- Systematic
- Literature
- Review
- Survey

The RA will develop specific search strings for each digital library and will revise the strings until they are able to detect all the papers found by the original manual search (see Appendix 1).

3.2 Inclusion criteria

Articles on the following topics, published between Jan 1st 2004 and June 30th 2007, will be included

- Systematic Literature Reviews (SLRs) i.e. Literature surveys with defined research questions, search process, data extraction and data presentation
- Meta-analyses (MA)
- SLRs relating to Software Engineering related topics (i.e. topics related to the development, maintenance, project & quality management of software intensive applications)

3.3 Exclusion Criteria

The following types of studies will be excluded

- Informal literature surveys (no defined research questions, no search process, no defined data extraction or data analysis process).
- Papers discussing process of EBSE, SLRs or Meta-analyses.

- SLRS related to Information Systems topics such as those related to evaluating installed systems/applications
- SLRs related to HCI topics such as usability testing & evaluation

When an SLR has been published in more than one journal/conference, both versions of the study will be reviewed for purposes of data extraction.

3.4   Initial Primary Study Selection Process

Once the search strings have been fully developed and tested, the RA will use the strings to extract the title and abstract all articles in the digital libraries found by the search strings.

Stage 1: The title and abstract of each article from each digital library will be reviewed against the inclusion and exclusion criteria and any papers that are clearly irrelevant will be excluded.

Stage 2. The lists from each digital library will be collated and duplicate references removed.

Stage 3 Candidate articles identified in the original tertiary study (see Appendix 1) and those later excluded from the study, (see Appendix 2) will be removed from the Stage 2 list of candidate studies

Stage 4 Full copies of all the candidate studies remaining after stage 3 will be reviewed against the inclusion and exclusion criteria.

Stage 5. The remaining studies will be checked for duplicate reports by checking papers that have the same authors and cover similar topics. At the end of this stage:
- Each article should have a unique reference number
- Each study should have a unique reference number, allowing for multiple studies in a single article and several articles reporting the same study.
- Each article will be identified as journal article, a conference paper, book chapter, technical report or other.

The RA will perform stages 1-3 and stage 5 supported by the supervisor. To complete Stage 4, the RA will assign each paper at random to two members of the research team and each person will decide whether or not to include the paper, excluded papers will be marked as either an irrelevant paper or an unstructured literature survey. Any disagreements about the inclusion of a paper will be mediated by another member of the team. The three team members must discuss the disagreement until the status of the article is finally resolved.

3.5   Final Primary Study Selection Process

The RA will review the references of each primary study remaining after Stage 5 and identify any additional candidate primary studies not already included in the selected primary studies. The candidate studies will be read by two other team members and marked for inclusion or exclusion. Any disagreements will be resolved by a third member of the team.

4   Quality Assessment

Each SLR identified by the selection process, will be evaluated using the DARE criteria [3]. The criteria are based on four questions:

- Are the review's inclusion and exclusion criteria described and appropriate?
- Is the literature search likely to have covered all relevant studies?
- Did the reviewers assess the quality/validity of the included studies?
- Were the basic data/studies adequately described?

The questions are scored as follows:

- Question 1: Y (yes), the inclusion criteria are explicitly defined in the study, P (Partly), the inclusion criteria are implicit; N (no), the inclusion criteria are not defined and cannot be readily inferred.
- Question 2: Y, the authors have either searched 4 or more digital libraries and included additional search strategies or identified and referenced all journals addressing the topic of interest; P, the authors have searched 3 or 4 digital libraries with no extra search strategies, or searched a defined but restricted set of journals and conference proceedings; N, the authors have search up to 2 digital libraries or an extremely restricted set of journals.
- Question 3: Y, the authors have explicitly defined quality criteria and extracted them from each primary study; P, the research question involves quality issues that are addressed by the study; N no explicit quality assessment of individual primary studies has been attempted.
- Question 4: Y Information is presented about each primary study; P only summary information is presented about papers; N the results of the individual studies are not specified.

The scoring procedure is Y=1, P=0.5 and N=0 or Unknown.

The RA will allocate each study to two members of the research team at random. The quality data will be extracted by both researchers independently and then cross-checked. Any disagreements will be mediated by a third team member. For all "Unknown" scores the authors of the primary study will be contacted to determine the appropriate score

5   Data Collection

The data extracted from each study will be the same as that used for the original tertiary study, i.e.:
- The source (i.e. the conference or journal that published the study).
- The year when the study was published. Note if the study was published in several difference sources all dates will be recorded and the first date will be used in any time-based analysis used to track the EBSE activity over time.
- Classification of study:
    - Type (Systematic Literature Review SLR, Meta-Analysis MA)

- o Scope (Research trends or specific research question).
- Main software engineering topic area.
- The author(s) and affiliation (organisation and country).
- Research question/issue.
- Whether the study made reference to an EBSE paper or the original SLR Guidelines (Kitchenham, 2004).
- How many primary studies were analysed.
- Whether the study proposed practitioner-oriented guidelines.
- Summary of study.
- Quality score for the study.

The data for each study will be extracted independently by the two researchers who performed the quality assessment for the study. Any disagreements will be mediated by a third member of the research team.

6   Data Analysis

The RA will tabulate the data in exactly the same format as the original tertiary study [2]. These tables will be kept separately for analysis by the EPIC case study team. Then the data from the original study and this study will be integrated into combined tables indicating clearly whether the study was found in the initial study or the current study.

The RA will review the results of the initial study and amend the results as necessary given the additional data.

7   Dissemination

The results of the study will be used by the EPIC case study team to investigate the relevant case study research questions. In addition, if the study is completed before September 2008, Barbara Kitchenham will use the results to update the tertiary study paper provisionally accepted by IST.

8   Schedule

Task	Role Responsible	Required Completion Date
Complete protocol construction	SLR team (BAK)	May 30 2008
Test and finalize search strings	RA plus supervisor (DB)	RA start plus 1 week
Complete Primary Study selection	RA plus SLR Team	RA start + 3 weeks
Complete Data Extraction	RA plus SLR team	RA start + 6 weeks
Complete Analysis	RA plus supervisor	RA Start + 8 weeks
Update IST paper with results	SLR team (BAK)	20 September 2008

9   References

1. Brereton, O. P., Kitchenham, B.A. The Scope of EPIC Case Studies. EPIC technical Report EPIC-2007-01.
2. Kitchenham, Barbara, O. Pearl Brereton, David Budgen, Mark Turner, John Bailey, and Stephen Linkman. A Systematic Literature Review of Evidence-based Software Engineering, EBSE Technical Report EBSE-2007-03, November 2007.
3. Centre for Reviews and Dissemination. (2007) What are the criteria for the inclusion of reviews on DARE? Available at http://www.york.ac.uk/inst/crd/faq4.htm (accessed 24 July 2007).

Appendix 1 Candidate articles that were included in the study

1. Barcelos, R.F., and Travassos, G.H. (2006) Evaluation approaches for Software Architectural Documents: A Systematic Review, Ibero-American Workshop on Requirements Engineering and Software Environments (IDEAS). La Plata, Argentina.
2. Dyba, T., Kampenes, V.B. and Sjøberg, D.I.K. (2006) A systematic review of statistical power in software engineering experiments, Information and Software Technology, 48(8), pp. 745-755.
3. Galin, D. and Avrahami, M. (2005) Do SQA programs work - CMM works. A meta analysis. IEEE International Conference on Software - Science, Technology and Engineering.
4. Galin, D. and Avrahami, M. (2006) Are CMM Program Investments Beneficial? Analyzing Past Studies. IEEE Software 23(6), pp. 81-87.
5. Glass, R.L., Ramesh, V., and Vessey, I. (2004) An Analysis of Research in Computing Disciplines, CACM, Vol. 47, No. 6, pp. 89-94.
6. Grimstad, S., Jorgensen, M. and Molokken-Ostvold, K. (2006) Software effort estimation terminology: The tower of Babel, Information and Software Technology, 48 (4), pp. 302-310.
7. Hannay, J E., Sjøberg, D.I.K and Dybå. T. (2007)A Systematic Review of Theory Use in Software Engineering Experiments. IEEE Trans on SE, 33 (2), pp. 87-107.
8. Jørgensen, M. (2004) A review of studies on expert estimation of software development effort, Journal of Systems and Software, 70 (1-2), pp. 37-60.
9. Jørgensen, M (2007) Estimation of Software Development Work Effort: Evidence on Expert Judgement and Formal Models, International Journal of Forecasting, 3(3), pp. 449-462.
10. Jørgensen, M., and Shepperd, M. (2007) A Systematic Review of Software Development Cost Estimation Studies, IEEE Transactions on SE, 33(1), pp. 33-53.
11. Juristo, N., Moreno, A.M. and Vegas,S. (2004) Reviewing 25 years of Testing Technique Experiments, Empirical Software Engineering Journal, Issue 1-2, March, pp. 7-44.
12. Juristo, N., Moreno, A.M. Vegas,S. and Solari, M. (2006) In Search of What We Experimentally Know about Unit Testing, IEEE Software, 23 (6), pp.72-80.
13. Kitchenham, B., Mendes, E., Travassos, G.H. (2006) A systematic review of Cross-company vs. Within-Company Cost estimation Studies. Proceedings of EASE06, BSC, pp 89-98.

14. Kitchenham, B., Mendes, E., Travassos, G.H. (2007) A Systematic Review of Cross- vs. Within-Company Cost Estimation Studies, IEEE Trans on SE, 33 (5), pp 316-329.
15. Mair, C. and Shepperd, M. (2005) The consistency of empirical comparisons of regression and analogy-based software project cost prediction, International Symposium on Empirical Software Engineering, pp 509-518.
16. Mendes, E. (2005) A systematic review of Web engineering research. International Symposium on Empirical Software Engineering, pp. 498-507.
17. Moløkken-Østvold, K.J., Jørgensen, M. Tanilkan, S.S., Gallis,H., Lien, A.C. and Hove, S.E. (2004) A Survey on Software Estimation in the Norwegian Industry, Proceedings Software Metrics Symposium, pp. 208-219.
18. Petersson, H., Thelin, T, Runeson, P, and Wohlin, C. (2004) Capture-recapture in software inspections after 10 years research – theory, evaluation and application, Journal of Systems and Software, 72, pp. 249-264.
19. Ramesh, V., Glass, R. L.; Vessey, I. (2004) Research in computer science: an empirical study, Journal of Systems and Software, 70(1-2), pp.165-176.
20. Runeson, P., Andersson, C., Thelin, T., Andrews, A. and Berling, T. (2006) What do we know about Defect Detection Methods? IEEE Software, 23(3), pp. 82-86.
21. Sjøberg, D.I.K., Hannay, J.E., Hansen, O., Kampenes, V.B., Karahasanovic, A., Liborg, N.K. and Rekdal, A.C. (2005) A survey of controlled experiments in software engineering. IEEE Transactions on SE, 31 (9), pp.733-753.
22. Torchiano, M. and Morisio, M. (2004) Overlooked Aspects of COTS-Based Development. IEEE Software, 21 (2), pp. 88-93.
23. Zannier, C, Melnick, G. and Maurer, F. (2006) On the Success of Empirical Studies in the International Conference on Software Engineering.ICSE06, pp. 341-350.

Appendix 2 Candidate articles that were not included in the study

1. T Mens and T, Tourwé, A survey of software refactoring, TSE 30(2), 2004 pp 126-139
2. S. Balsamo, A. Di Marco, P. Inverardi, Model-based Performance Prediction in Software Development, TSE 30(5), 2004 pp295-309
3. S. Mahmood, R. Lai and Y.S. Kim, Survey of component-based software development, IET Software, 2007 1(2), pp 57-66
4. D.C. Gumm, Distribution Dimensions in Software Development, IEEE Software 23(5) 2006 pp 45-51
5. M. Shaw and P Clements, The Golden Age of Software Architecture, IEEE Software 2006 23(2) pp31-39
6. M. Aberdour, Achieving Quality in Open Source Software, IEEE Software 2007 24(1), pp 58-64
7. D. Damian, Stakeholders in Global Requirements Engineering: Lessons learnt from practice., IEEE Software 2007 24(2), pp 21-27
8. E. Folmer and J. Bosch, Architecting for usability: a survey, JSS 70, 2004 pp 61-78
9. Hochstein & Lindvall, Combating architectural degeneration: a survey, IST 47, 2005 pp 643-656

10. S. Mahmood, R. Lai, Y.S. Kim, J.H. Kim, S.C. Park, H.S. h, A survey of component-based system quality assurance and assessment, IST 47, 2005 pp 693-707
11. J. Estublier, D. Leblang, A. van der Hoek, R. Conradi, G. Clemm, W. Tichy, D. Wiborg-Weber, Impact of software engineering research on the practice of software configuration management, TOSEM 2005 pp 383-430
12. Barbara G. Ryder, Mary Lou Soffa, Margaret Burnett, The impact of software engineering research on modern programming languages, TOSEM , 2005 pp 431-477
13. J. Ma and J. V. Nickerson, Hands-on, Simulated and Remote Laboratories: A Comparative Literature Review, ACM Survey , 38(3), 2006 pp 1-24
14. S. Wagner, A literature survey of the Quality Economics of Defect-Detection Techniques, ISESE, 2006