Formalizing a Systematic Review Updating Process

Oscar Dieste
Facultad de Informática
Universidad Politécnica de Madrid
28660 Boadilla del Monte, Spain
odieste@fi.upm.es

Marta López
Facultad de Informática
Universidad Complutense de Madrid
C/Jose García Santesmases S/N
28040 Madrid, Spain
mlf@fdi.ucm.es

Felicidad Ramos
INDRA Systems
Carretera Loeches 9
28850. Torrejón de Ardoz, Spain
framos@indra.es

Abstract

The objective of a systematic review is to obtain empirical evidence about the topic under review and to allow moving forward the body of knowledge of a discipline. Therefore, systematic reviewing is a tool we can apply in Software Engineering to develop well founded guidelines with the final goal of improving the quality of the software systems. However, we still do not have as much experience in performing systematic reviews as in other disciplines like medicine, and therefore we need detailed guidance. This paper presents a proposal of a improved process to perform systematic reviews in software engineering. This process is the result of the tasks carried out in a first review and a subsequent update concerning the effectiveness of elicitation techniques.

1. Introduction

The requirements elicitation activity focuses on the analyst-customer interaction, with the aim of gathering high quality information about the future software [31]. To collect this information we can apply different requirements elicitation techniques [27]. However, despite the existence of a large set of techniques, there are not clear guidelines which define when to apply those techniques in practice [23,43]. So far, the applicability of elicitation techniques is based in the recommendation made by experts in Requirements Engineering (RE), with very few exceptions like [32].

Surprisingly, an examination of the scientific literature reveals the existence of a number of experiments about elicitation techniques, such as [1,2,5,6,8,9,33,36,37,39] to cite a few. All this empirically-based knowledge was underemployed and largely ignored. Therefore in 2005 we decided to summarize the existing empirical evidence about elicitation techniques following a procedure similar to Juristo et al. [28], but enriched with a, in that moment, recent proposal by Kitchenham [30].

Kitchenham's proposal aimed at performing Systematic Reviews (SR) in Software Engineering (SE). SR is a procedure extensively used in many disciplines (i.e. Medicine, Psychology, Education, etc.) to gather and synthesize scattered experimental results and generate pieces of knowledge. Those pieces of knowledge can serve as a foundation to establish empirically-based recommendations in any field.

We performed an initial SR [10,15,16] comprising 30 experiments and, some time later, we carried out an update of that review [17,18] including 13 new experiments. Apparently this work should be simple and easy because we have an SR process successfully applied and the pieces of knowledge obtained in the previous review. However, the update turned into a complex work because some issues were not anticipated by the procedures of SR and therefore some problems arose [14].

Even if new SR proposals have been made (i.e.: [3,34]), they have not addressed the problems experienced by the authors. This situation called for an improvement of existing SR procedures. A proposal for a improved process is presented in this paper, which has been structured as follows: section 2 describes the related work; section 3 briefly describes the initial SR as well as the subsequent update; Finally, section 4 describes the improved review

[†] Now the author is with Fraunhofer IESE, 67663 Kaiserslautern, Germany. Email: oscar.dieste@iese.fraunhofer.de

978-0-7695-3302-5/08 $25.00 © 2008 IEEE
DOI 10.1109/SERA.2008.35

process, improved taking into account the lessons learned so far.

2. Related work

SRs are usually performed in disciplines with a strong empirical component, like Medicine or Psychology. The reason is the great quantity of experiments available. In the SE field, SRs are not usually carried out due to the shortage of experiments (although their number is increasing year after year.) Another problem is derived from immaturity of the experimentation in SE. The poor quality of many experiments, the problems of reporting and the lack of proper statistical analysis make impossible the direct application of other disciplines' recommendations for SR (i.e. Medicine [25]) .

The first attempts to carry out SRs in SE used ad hoc procedures like [26,28,40]. However, a desirable attribute of any review is to be systematic and repeatable for future applications. In 2004, Kitchenham proposed a procedure to perform SRs [30] based on Khan [29]. The structure of this procedure is shown in Figure 1.

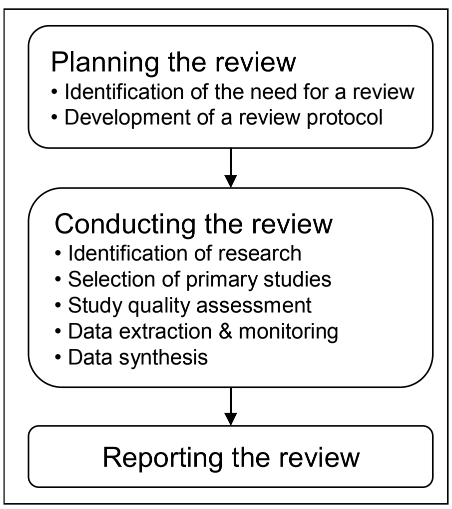

Figure 1. SR process, according to Kitchenham

During the planning phase, the need for a systematic review is detected, research objectives are listed and a review protocol is defined. Such protocol specifies the research question and the methods that will be used to perform the SR. The conducting phase involves the primary studies' identification, selection and evaluation taking into account the inclusion and exclusion criteria established in the SR protocol. Reviewers extract data from the studies in order to analyze and synthesize them during the data synthesis stage and finally communicating at the reporting phase. Usually SR can be reported in two different formats: technical report (or a section of a PhD thesis) and journal/conference paper. The phases shown in Figure 1 appear to be sequential, but in practice iteration between phases happen. For example, if any of the criteria established during planning turns out to be inadequate during conduction, then it can be changed provided that the appropriate corrective measures are applied to the review work already performed.

Since its introduction, most of the SRs carried out in SE have followed Kitchenham's proposal with few modifications. (i.e: [16,20,21]). Lately, Biolchini et *al*. developed a new SR process [3,35]. Biolchini's process follows closely Kitchenham's SR process, as seen in Figure 2.

Biolchini's proposal includes a template to perform SR. This template is highly influenced by Mendes [34]. The objectives of the template are: 1) to reduce the overload of SR planning and execution and 2) to serve as a guideline to SE researchers when conducting the systematic review.

3. Initial SR and subsequent update

We performed an initial SR [10] in 2005. This review identified 564 publications from the SCOPUS, IEEEXPLORE and ACM DL databases, as well as Google. We selected and extracted data from 26 of those publications. The selected publications contained 30 empirical studies. These studies were designed to test 43 elicitation techniques and 50 different response variables. We got a hundred separate results from those experiments. The aggregation generated 18 pieces of evidence about the interviewing, laddering, sorting and protocol analysis elicitation techniques [16].[15]

However, the initial SR disregarded 27 potentially interesting publications because it was impossible to obtain the full paper on time to review. As the conclusions of the SR are contingent upon the available evidence, it was clear that a more thorough search (e.g. in international library services) was desirable. It made possible to obtain 13 out of those 27 publications. The other 14 publications (e.g. [11,22]) were considered impossible to locate, as they are quite old, grey literature. Not all those 13 publications were useful. Four of them did not contain empirical studies at all or were papers published twice, so that they were discarded. The other nine[7] were useful and gave 13 empirical studies ([42] contained 2 different studies while [4] contained 3).

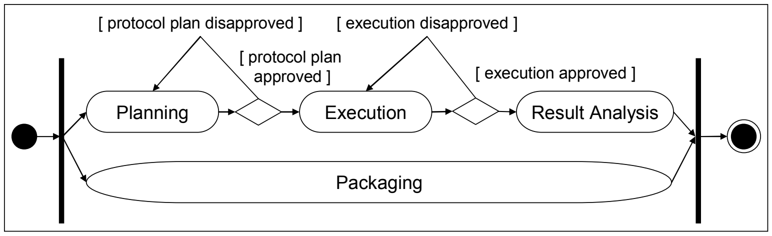

Figure 2. SR process, according to Biolchini

At first, we thought of applying the same procedure than in the initial SR. It was the most appropriate way of work, because a detailed protocol [12] and a detailed set of results [13] were already available. However, while executing it, many problems arose. For example, because there were no descriptions of the elicitation techniques in the initial SR, the work of classifying the elicitation techniques tested in the new experiments became very difficult. Likewise, other problems, related with data extraction, codification of results and aggregation also showed up [14]. Those problems forced us to deviate somehow from the initial plan. With the aim of improving the SR process, we tracked and collected all modifications to the original process. For reasons of space the details of the detected problems and their solutions are not included in this paper. Finally, we developed a new proposal for a SR process, described in the next section.

4. Improved systematic review process

The original process in [10] contained 9 main activities, as shown in Table 1. The new process contains 10 activities and also decomposes further some previous atomic activities. The graphical view of the improved process is shown in Figure 3. Each step is described in the following sections.

4.1. Training on systematic reviewing

In this first step we will pursue: 1) to familiarize reviewers with the specific terminology of the field of knowledge in which the review will take place and 2) to make sure that the review process and activities are known to the reviewers. In order to guarantee and facilitate the later update of the review, it is desirable a thorough understanding of all topics related with the review.

4.2. Selection of primary studies

The goal of this step is the selection of the empirical papers available in the literature about the technology of interest. The empirical studies should be identified by means of using a keyword search in well-known accepted databases, such as IEEEXPLORE, ACM DL and/or SCOPUS databases, among others. Due to the fact that in SE there are not a lot of empirical studies, other empirical papers can be identified analyzing the bibliographical references of the studies identified by means of keyword search. An example of keyword search could be the following [16]:

> *(elicitation OR "requirements gathering" OR "requirements acquisition") AND (capture OR empirical OR experiment OR study OR review OR evaluation).*

Each identified empirical study will be evaluated by at least two work review members to decide on whether or not it should be selected. Any discrepancies were settled by consensus. The review team has to specify the conditions that the studies have to meet in order to be included in the SR. It is also important to realize that the results of several experiments can be used in different papers and, therefore, it is advisable to pay attention to duplicated empirical studies.

4.3. Reading of primary studies and data extraction

It is very difficult to perform a SR with the textual form in which empirical data are described in the primary studies. Due to this, and with the objective of formalizing the later aggregations, we extract the information into two forms: a data collection form and empirical study collection form. Of course, the content of these forms will depend on the field reviewed. Therefore, it is not possible to make generic forms applicable to all fields of knowledge.

The data collection form will contain data about the specific feature to be reviewed. In our case, this form had been used to state the significant differences

Table 1. Initial SR process

1. Selection of primary studies	A set of keywords were chosen to search for empirical studies in online databases. The references of the interesting papers were also looked through.
2. Reading of papers and data extraction	After reading each paper reviewers extract the data into a data extraction form. Only the data included in this form will be analyzed in the remaining processes of the SR.
3. Treatment codification	For each treatment, a unique and no-duplicate identifier is assigned. This id will be later used both to code the results and during the aggregation. Id's make possible a formal approach to SR, avoiding bias derived from free text and reviewer's interpretations.
4. Response variable codification	For each response variable, a unique and non-duplicate identifier is assigned. This response variable id will be later used both to code results and during the aggregation.
5. Experiment results codification	Rephrase the experiment results using the treatment and response variable ids. For instance, an experiment can show that interviews are better than protocol analysis as regards time (that is, interviews take less time than protocol analysis). If interviews would be coded as *X*, protocol analysis as *Y*, and time as ***t***, we can obtain the following formal expression: $X > Y$, regarding $\boldsymbol{t}$.
6. Discard low quality experiments	Before generalization and aggregation, the experimental design is identified and classified (i.e.: the study is a randomized experiment with parallel groups, or a pre-test/post-test design, etc.). Based on the design employed, a quality score is assigned. Finally, reviewers set up a threshold, and select only those experiments with scores equal or greater than the threshold.
7. Generalization of evidences	If reviewers study each technique and its variations in a one-by-one approach, they will not be able to obtain general findings. Instead of focusing on the small differences among techniques, we have to put attention in the general similarities. For example, if we have found *n* types of open interviews in the selected papers, we can generalize them into just 1 group of interview-based techniques, and work with that group instead of with the individual techniques.
8. Aggregation	The goal of the aggregation is to identify empirically-founded pieces of knowledge, which will be the final SR findings. By 'piece of knowledge' we mean the synthesis or union of compatible results from different empirical studies. The aggregation procedure employed is a variation of comparative analysis [38]. For example, if we have 2 studies *a* and *b*, which test techniques *X* and *Y* respect to a given response variable *t*, and both studies find that *X* > *Y*, then we can conclude that X > Y is likely true. We will say that *X* > *Y* is an aggregation of *a* and *b*. All aggregations will be coded. Below there is an example following [10]: AG03 = [D02, T(a)1/Ta(2)] T(a)2 > T(a)1 [D08, T(a)1/T(a)2] T(a)2 > T(a)1
9. Evidence decodification	The final activity of this process is the de-codification of aggregations in order to get a piece of knowledge understandable by any reader. For example, the de-codification of the previous aggregation means that technique T(a)2 (structured interviews) produces more customer needs (V(a)1) than T(a)1 (unstructured interviews).

between the elicitation techniques, as well as other features of the empirical study like the number of experimental units, the factors and the response variables, the experimental problem, etc.

The empirical study collection form will contain data about the experiment itself. It will be used to assess the quality of the empirical study. In our case, quality was assessed by assigning each study to a level in the evidence hierarchy proposed by Kitchenham [30].

4.4. Setting up supporting tables

In this step two supporting tables are developed. These tables contain the details of the treatments and response variables used in each study. In our proposal, these tables contain the following data:

- Treatment table: for each study, the name of the technique/s, brief description of the technique and the assigned technique's ID.
- Response variable table: for each experiment, name of the response variable/s, brief description of the variable, and the assigned response variable's ID.

They are called 'supporting tables' because these tables will be a temporal storage of the results in order to facilitate the later aggregation. It is remarkable that these tables will be constantly changed during the execution of the SE whenever new information is extracted from primary studies.

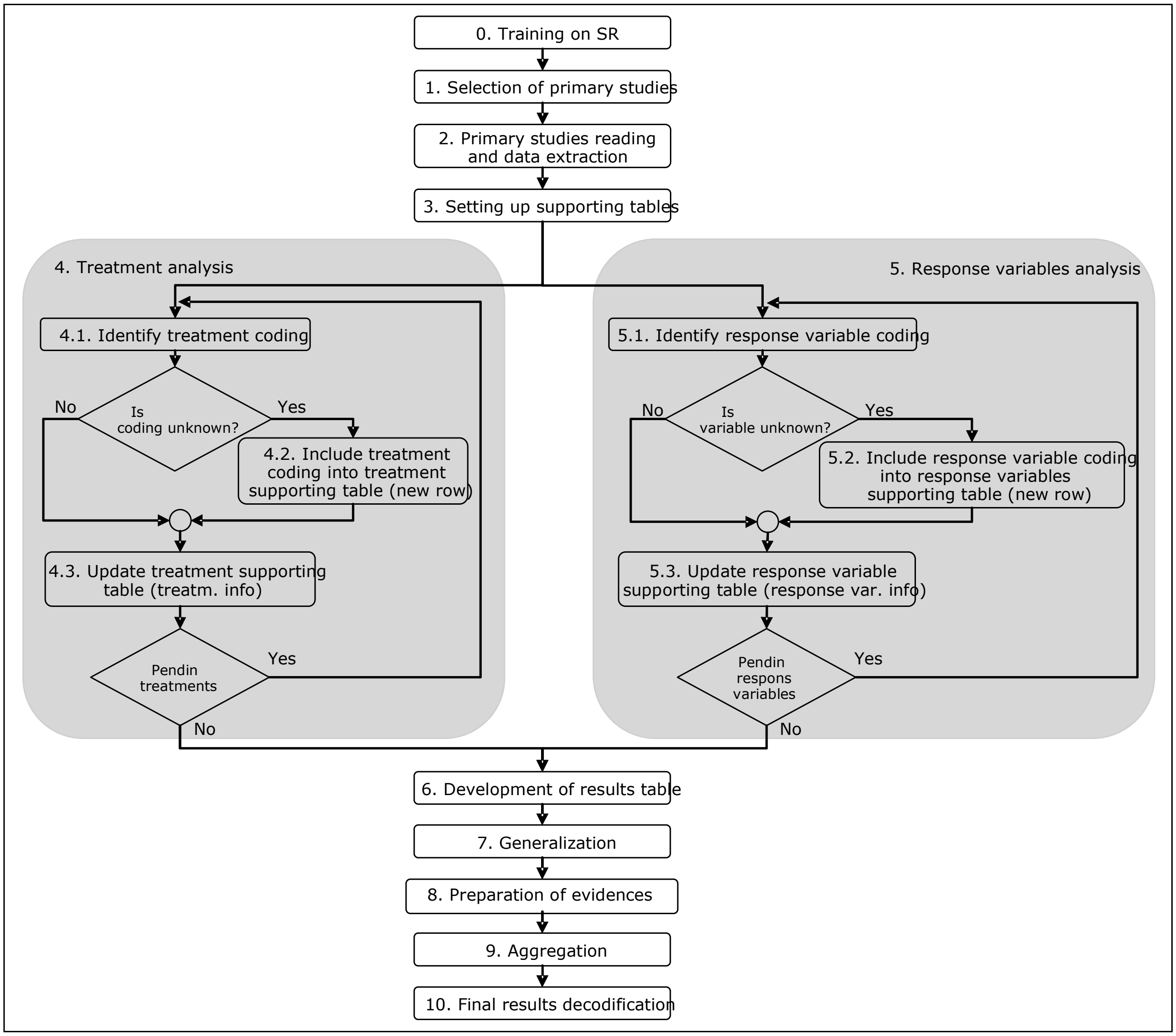

Figure 3. SR improved process.

With these tables, traceability of the experiments, treatments and response variables can be implemented and maintained all over the review.

During this step, it is also desirable to set-up glossary of treatments and response variables to ease the subsequent activities of the proposed process.

4.5. Treatment analysis and response variables analysis

As shown in Figure 3, each of these two steps will be performed iteratively until all treatments and response variables are codified. The general rules to perform these steps are the following:

- Do not use the name of treatments or response variables, but their descriptions (recorded in the glossary). The same treatment (or response variable) can be named differently in diverse studies because the names are a subjective naming, depending on each author. Do not forget to update the glossary of the review.
- Once a new treatment (or response variable) can be mapped to an existing one, use the same code for both of them. Avoid duplications. Sometimes this mapping could be a complex task because we can find several treatments (or response variables) with which the current treatment (or response variable) can be mapped. In those cases, we suggest to write down all the mapping combinations in the supporting tables, bounding the search space until finding out the most suitable option.

Anyway, if the same treatment (or response variable) is repeated in the supporting tables with different codes, the later generalization process can put them into the same group and consider them as equivalent (if generalization is correctly performed). However, our experience tell us that it is preferable to reserve generalizations to create groups of treatments (or response variables), and not to fix errors made during the codification of experiments.

4.6. Development of the results table

Only after identifying and codifying all treatments and response variables, the results table will be created. This table will show the results obtained in each analyzed experiment, using the treatment and response variable codifications. The general rules to perform this step are the following:

- Assign a unique id to each result described in the experiment, and relate it to the appropriate treatments and response variable/s.
- If for any reason the ids are modified in subsequent steps, make sure that all changes are propagated adequately.
- Sometimes 'no significant differences' (n.s.d.) are detected between the tested treatments. Use this acronym (n.s.d) to identify them.

4.7. Generalization

The generalization step is aimed at creating groups of similar treatments or variable responses to make aggregation of results more productive.

There are several approaches to perform this step, depending mainly on the quantity of information extracted from the primary studies. If we had many experiments, generalization would not be necessary because it is likely that we already have compatible results to aggregate. However, since in SE there are not a lot of published experiments and data is scarce, we follow the strategy in [10]. This strategy consists of grouping together similar treatments or response variables. This makes easier to later generate a greater number of pieces of knowledge.

4.8. Preparation of evidences

In this step the experimental results, coded in step 6 (see Figure 3), are updated by including the generalizations obtained up to now. The final set of evidences is renumbered (that is, the ids are reorganized), establishing a definitive numbering scheme. If we are performing the update of a previous review, it is necessary to keep in mind that the ids used whilst carrying out the update may be independent of the numbering applied in the preceding review. Due to this, now we have to check carefully the whole set of results (the findings of the past reviews plus those obtained during the update) and unify the numbering schemes.

It is prudent to pay attention and check carefully contradictory results. Maybe these disagreements could be the result of incorrect generalizations. A thorough analysis is advisable when performing this step.

Once the analysis has been done, the experimental results can be used as evidences for the aggregation step.

4.9. Aggregation

Once the new evidences are extracted, in this step we will aggregate them. As mentioned before, in this work we used a variation of the comparative analysis [38]. However, reviewers can consider other approaches. In the literature there are several procedures available, either interpretative procedures, such as content analysis or meta-ethnography [19], or non interpretative such as meta-analysis [41], case surveys [44] or vote-counting techniques [24].

The final output of the aggregation will be a set of codified pieces of knowledge AG_i, as shown in Table 1. Based on those aggregations, we will obtain pieces of knowledge which will be the final results of the SR. More concretely, the pieces of knowledge will be the conclusive aggregations, that is, those aggregations in which several studies find the same relationship among techniques.

If we are running an updating review, we have to apply the same procedure used in previous reviews. The new aggregation will obtain pieces of knowledge that can corroborate or contradict previous ones.

4.10. Final results de-codification

The codification of treatments, response variables and results is useful to perform the aggregation. However, the aggregations in codified form are hard to read and understand. Therefore, the last step of the proposed SR process is the de-codification of aggregations in order to obtain a textual description understandable by any reader.

5. Conclusions

In this paper, we propose a process to perform SR in SE: This SR process has been developed taking into account two main sources: 1) the initial in-depth study of the review process in other disciplines and 2) the lessons learned after performing an initial SR and a later update.

We do know that our experience in SR is still insufficient. For example, we only applied SR to elicitation techniques in RE, but not to other fields of SE. Likewise, we were unable to apply sophisticated aggregation techniques like meta-analysis [41] but we had to use variants of comparative analysis [38] instead. These and other limitations have an impact in the proposed process, limiting its applicability.

As future research, we plan to address those limitations. We will apply SRs to other fields like software inspections or reading techniques and we will use the acquired experience to improve the applicability of the proposed SR process.

6. References

[1] Agarwal, R. and Tanniru, M. R., "Knowledge Acquisition Using Structured Interviewing: An Empirical Investigation," *Journal of Management Information Systems*, vol. 7, pp. 123-141, Summer, 1990.

[2] Bech-Larsen, T. and Nielsen, N. A., "A comparison of five elicitation techniques for elicitation of attributes of low involvement products," *Journal of Economic Psychology*, vol. 20, pp. 315-341, 1999.

[3] Biolchini, J., Mian, P., Natali, A., and Travassos, G. H., COPPE/UFRJ, Rio de Janeiro, Brazil, ES 679/05, 2005.

[4] Bradburn, B. , "A comparison of knowledge elicitation methods," *International Conference on Engineering Design (ICED'91),* pp. 298-305, 1991.

[5] Breivik, E. and Supphellen, M., "Elicitation of product attributes in an evaluation context: A comparison of three elicitation techniques," *Journal of Economic Psychology*, vol. 24, pp. 77-98, 2003.

[6] Browne, G. J. and Rogich, M. B., "An Empirical Investigation of User Requirements Elicitation: Comparing the Effectiveness of Prompting Techniques," *Journal of Management Information Systems* , vol. 17, pp. 223-249, Spring, 2001.

[7] Burton, A. M., Shadbolt, N. R., Rugg, G., and Hedgecock, A. P., "Knowledge elicitation techniques in classification domains," *Proceedings of the 8th Conference in Artificial Intelligence ECAI-88,* 1988.

[8] Burton, A. M., Shadbolt, N. R., Rugg, G., and Hedgecock, A. P., "The efficacy of knowledge acquisition techniques: A comparison across domains and levels of expertise," *Knowledge Acquisition*, vol. 2, pp. 167-178, 1990.

[9] Corbridge, B., Rugg, G., Major, N. P., Shadbolt, N. R. , and Burton, A. M., "Laddering - technique and tool use in knowledge acquisition," *Knowledge Acquisition*, vol. 6, pp. 315-341, 1994.

[10] Davis, A., Dieste, O., Hickey, A., Juristo, N., and Moreno, A. M., "Effectiveness of Requirements Elicitation Techniques: Empirical Results derived from a Systematic Review," *Proceedings of the IEEE International Conference on Requirements Engineering,* Minneapolis, Minessota, 2006.

[11] de Bont, C. J. P. M., Consumer Evaluation of Early Product-Concepts 1992. Delft University.

[12] Dieste, O."TR2. Protocolo de revisión de estudios empíricos," http://grise.ls.fi.upm.es/research_documents.php, 2006.

[13] Dieste, O."TR3. Resultados extraídos de los estudios empíricos seleccionados (http://grise.ls.fi.upm.es/research_documents.php)," http://grise.ls.fi.upm.es/research_documents.php, 2006.

[14] Dieste, O., Fernández, E., Grimán, A., López, M. , and Ramos, F., "Systematic Reviews: Some Practical Experiences," *Submitted to the 2nd Workshop on Empirical Software Engineering and Metrics.,* Kaiserslautern, Germany.

[15] Dieste, O., Juristo, J., and Shull, F., "Understanding the Customer: What Do We Know about Requirements Elicitation Techniques?," *IEEE Software*, vol. 25, pp. 11-13, 2008.

[16] Dieste, O. and Juristo, N., "Systematic Review and Aggregation of Empirical Studies on Elicitation Techniques," *IEEE Transactions on Software Engineering*, vol. 2008.

[17] Dieste, O., López, M., and Ramos, F., "Obtaining Well-Founded Practices about Elicitation Techniques by Means of an Update of a Previous Systematic Review," *Proceedings of SEKE'08,* Redwood City, CA USA.

[18] Dieste, O., López, M., and Ramos, F., "Formalizing a systematic review updating process," Universidad Complutense de Madrid, Technical Report, to appear.

[19] Dixon-Woods, M., Agarwal, S., Jones, D., Young, B., and Sutton, A., "Synthesising qualitative and quantitative evidence: a review of possible methods," *Journal of Health Services Research and Policy*, vol. 10, pp. 45-53B(9), 2005.

[20] Dyba, T., Arisholm, E., Sjoberg, D. I. K., Hannay, J. E., and Shull, F., "Are two heads better

than one? On the effectiveness of pair programming," *IEEE Software*, vol. 24, pp. 10-13, 2007.

[21] Dyba, T., Kampenes, V., and Sjoberg, D., "A systematic review of statistical power in software engineering experiments," *Information and Software Technology*, vol. 48, pp. 745-755, 2006.

[22] Geiwitz, J. , Kornell, J., and McCloskey, B. P., "An Expert System for the Selection of Knowledge Acquisition Techniques," Anacapa Sciences, Santa Barbara, CA, USA, Technical Report 785-2, 1990.

[23] Goguen, J. A. and Linde, C., " Techniques for requirements elicitation," *Proceedings of IEEE International Symposium on Requirements Engineering,* pp. 152-164, 1993.

[24] L.V. Hedges and I. Olkin. *Statistical Methods for Meta-Analysis*, Academic Press, 1985.

[25] Higgins, J. P. T. and Green, S. *Cochrane handbook for systematic review of interventions*, Chichester, UK: John Wiley and Sons, Ltd., 2006.

[26] Hu, Q., "Evaluating Alternative Software Production Function," *IEEE Transactions on Software Engineering*, vol. 23, pp. 379-387, 1997.

[27] Jones, S. R. and Miles, J. C., "The Use of a Prototype System for Evaluating Knowledge Elicitation Techniques," *Expert Systems*, vol. 15, pp. 83-97, 1998.

[28] Juristo, N. , Moreno, A. M., and Vegas, S., "Reviewing 25 Years of Testing Technique Experiments," *Journal of Empirical Software Engineering*, vol. 9, pp. 7-44, 2004.

[29] Khan, K., Khalid, S., Riet, G. T., Glanville, J., Sowden, A. J., and J.K., NHS Centre for Reviews and Dissemination. University of York, CRD Report Number 4 (2nd edition), 2001.

[30] B.A. Kitchenham. *Procedures for performing systematic reviews*, Keele University TR/SE-0401, 2004.

[31] P. Loucopoulos and V. Karakostas. *Systems Requirements Engineering*, McGraw-Hill, 1995.

[32] Maiden, N. A. M. and Rugg, G., "ACRE: selecting methods for requirements acquisition," *Software Engineering Journal*, vol. 11, pp. 183-192, 1996.

[33] Marakas, G. M. and Elam, J. J., "Semantic structuring in analyst and representation of facts in requirements analysis," *Information Systems Research*, vol. 9, pp. 37-63, 1998.

[34] Mendes, E., "A Systematic Review of Web Engineering Research," *ACM/IEEE International Symposium on Empirical Software Engineering (ISESE'05),* Noosa heads, Australia.

[35] Mian, P., Conte, T., Natali, A., Biolchini, J., and Travassos, G., "A Systematic Review Process to Software Engineering," *Proc. of the 2nd Experimental Software Engineering Latin American Workshop (ESELAW'05),* 2005.

[36] Moody, J. W., Will, R. P., and Blanton, J. E., "Enhancing knowledge elicitation using the cognitive interview," *Expert Systems with Applications*, vol. 10, pp. 127-133, 1996.

[37] Pitts, M. G. and Browne, G. J., "Stopping Behavior of Systems Analysts During Information Requirements Elicitation," *Journal of Management Information Systems*, vol. 21, pp. 203-226, 2004.

[38] C.C. Ragin. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*, Berkeley, CA: University of California Press, 1987.

[39] Schweickert, R., Burton, A. M., Taylor, N. K., Corlett, E. N., Shadbolt, N. R., and Hedgecock, A. P., "Comparing knowledge elicitation techniques: A case study," *Artificial Intelligence Review*, vol. 1, pp. 245-253, 1987.

[40] Shull, F., Carver, J., Travassos, G. H., Maldonado, J. C. , Conradi, R., and Basili, V. R. Replicated Studies: Building a Body of Knowledge about Software Reading Techniques. In: *Lecture Notes on Empirical Software Engineering*, ed. N. Juristo and A.M. Moreno. Singapore: World Scientific, 2003.pp. 39-84.

[41] S.E. Straus , W.S. Richardson, P. Glasziou, and R.B. Haynes. *Evidence-Based Medicine*, Churchill Livingstone, 2005.

[42] Wood, L. E. , Davis, T. C., Clay, S. L., Ford, J. M., and Lammersen, S., "Evaluation of interviewing methods and mediating representations for knowledge acquisition," *International Journal of Expert Systems*, vol. 8, pp. 1-23, 1995.

[43] Wooten, T. C. and Rowley, T. H., "Using anthropological interview strategies to enhance knowledge acquisition," *Expert Systems with Applications*, vol. 9, pp. 469-482, 1995.

[44] Yin, R. K. and Heald, K. A., "Using the Case Survey Method to Analyze Policy Studies ," *Administrative Science Quarterly*, vol. 20, pp. 371-381, 1975.