A Systematic Mapping Study on Empirical Evaluation of Software Requirements Specifications Techniques

Nelly Condori-Fernandez[1], Maya Daneva[2], Klaas Sikkel[2], Roel Wieringa[2]
Oscar Dieste[3], Oscar Pastor[1],
Universidad Politecnica de Valencia[1], University of Twente[2],
Universidad Politecnica de Madrid[3]
{nelly,opastor}@pros.upv.es[1]; {m.daneva, k.sikkel}@utwente.nl[2], roelw@cs.utwente.nl[2];
odieste@fi.upm.es[3]

Abstract

This paper describes an empirical mapping study, which was designed to identify what aspects of Software Requirement Specifications (SRS) are empirically evaluated, in which context, and by using which research method. On the basis of 46 identified and categorized primary studies, we found that understandability is the most commonly evaluated aspect of SRS, experiments are the most commonly used research method, and the academic environment is where most empirical evaluation takes place.

1. Introduction

Increasing attention is being paid to empirical validation of SRS proposals, but there is still a need for software researchers to develop a better understanding of which software methods function best and why. Evidence-based software engineering plays an important role in this because it provides the means by which current best evidence from research can be integrated with practical experience and human values in the decision-making process for software development and maintenance [3] [10]. The core tool of this evidence-based paradigm is the Systematic Review[1], which has received much attention lately in software engineering [1][2][6]. However, systematic mapping studies, frequently used in other research fields, have largely been neglected by software engineering researchers [9][9][11]. Mapping study is a research method that provides an overview of a research area and allows us to identify the quantity and type of research and results available within it [7].

---

[1] Secondary research method that aims at systematically gathering and analyzing all evidence available on specific topic in an objective, unbiased and consistent manner [12]

This paper presents the results of a mapping study to identify and categorize a set of primary studies covering all quality aspects of the requirements specification process and product currently being considered by the researchers. The mapping study addresses the following research questions (RQs): 1) Which are the most investigated quality aspects of the SRS techniques? 2),3) In what study settings and in what problem domains are these aspects investigated? and 4) What research method was used in the evaluation of the aspects most studied?

In the following section we describe our review process, and Section 3 addresses the limitations found and discusses implications identified from this study.

2. The systematic mapping process

We carried out our systematic mapping study in three stages, which are presented below.

2.1. Stage 1: Defining Scope, search strategy and selection criteria

The scope of this study was as follows:
Population. Set of articles describing empirical studies in industry, academia and government reporting empirical evaluations. *Intervention.* Any empirical study involving SRS, specification languages, methods, techniques and tools. *Outcomes.* Quantity and type of evidence relating to the evaluation of requirements specification. *Study design.* Experiment, case study, experience reports, action research, observational study, survey.

The search strategy comprises the identification of *search terms* and the selection of *search resources*.
With respect to *search terms*, we used a search string consisting of two parts. The first part will be related to the type of studies that we wish to include in the study: (1) experiment, (2) action research (3) experience

report, (4) experimental study, (4) experimental comparison, (6) experimental analysis, (7) experimental evidence, and (8) empirical study.

The second part will be related to the specific technology to be reviewed: (9) requirements specification technique, (10) requirements specification method, (11) requirements specification approach, (12) requirements modeling (US) (13) requirements modelling (UK), (14) requirements model, (15) requirements specification, (16) specification language, (17) modeling language and (18) requirements specification process. Then, we used Boolean OR to join alternate terms and synonyms, and Boolean AND to join two major parts.

With respect to *search resource*, we considered using mainly Scopus. However, other complementary search resources (IEEE Digital Library, ACM Digital Library, and manual search) were also used since Scopus only includes partial articles of some conferences that are relevant to our research.

To select papers from the retrieved results, we used the following *inclusion criteria*: I1: The paper empirically evaluates one or more requirements specification approaches either in industrial or academic or government settings; I2: The paper empirically compares two or more requirements specification approaches. I3: In the case of dissimilar and similar replications, each of them was considered.

We also considered the following *exclusion criteria*: E1: The paper theoretically evaluates one or more features of an SRS technique; E2: The paper presents an approach to the theoretical evaluation of SRS technique; E3: Empirical studies on the evaluation of such approaches are also excluded; E4: Empirical studies that evaluate software artifacts produced in analysis, design and implementation phases; E5: If two papers publish the same empirical results, one of them is excluded; E6: Any paper that is not accessible is excluded; E7: We excluded posters, summaries of articles, tutorials, and panels.

2.2. Stage 2: Selecting primary studies

The selection process comprises four iterations: the first three are carried out by three reviewers, while the last iteration is by one evaluator. Prior to the process, the list of papers from SCOPUS was broken down in three parts of around 50 papers. In the first iteration, each part was independently reviewed by two reviewers. Each reviewer applied the inclusion/exclusion criteria to each paper, based on the paper title and abstract only. In the next iteration, those papers, which reviewer considered undetermined, were reviewed again but including the introduction and

conclusions. In the third iteration, the two reviewers compared their results and when they disagreed regarding the inclusion of a paper, they discussed their positions until they reached consensus. Those papers, which the two reviewers deemed undetermined, were reviewed by the third reviewer, based on whole article. This fourth iteration was meant to reduce the threat to internal validity of our results.

We selected 34 out of the 206 papers from SCOPUS; 4 of 29 papers from ACM digital library, 2 papers from WER, and 6 papers from REFSQ. Table 1 shows the top five publication channels. The MODELS, REFSQ and ICSE conferences seem to be the dominant forums. However, 21% of 46 selected papers were only published in journals.

Table 1. Top five publication channels (1987-2008)

Acronym	Type of publication	Percent
MODELS	Conference	13.0%
ICSE	Conference	13.0%
REFSQ	Conference	13.0%
ISESE-ESEM	Conference	8.7%
RE	Conference	6.5%

2.3. Stage 3: Classifying selected studies

Our classification criteria include SRS aspect studied, type of empirical study, study setting, and problem domain:

- *Aspect studied*: this refers to the quality properties investigated in an empirical study. As there are few quality models for evaluating different objects of study that are produced in the RE discipline, we had to carry out a process of similar terminology unification due to variety of terms used.
- *Type of empirical study*: We consider six types of empirical study, partly based on [8]: experiment, case study, experience reports, observational study, survey, and action research. The classification according to this criteria was supervised by an expert in research methodologies.
- *Study setting*: this refers to the context in which studies are realized. It can be in industry, government or academic settings. We considered also the combination of these as a mixed setting.
- *Domain*: we cannot evaluate an SRS technique without discussing where its use could be appropriate. We consider the following taxonomy of domains proposed by Kotonya [4]: command and control, embedded software, electronic commerce, real-time, management information systems (MIS), simulation, and virtual reality.

Considering these criteria, our four RQs were addressed, which are analyzed below.

1) *Which are the most investigated aspects of SRS techniques?*

31 aspects studied for our research were found. Table 2 reports the top-six most studied aspects: understandability, efficiency, correctness, defect rate, completeness, and consistency. 41.3% of the studies focus on understandability of SRS. We found also 11 aspects with only one occurrence. This might be indicative that more research is needed to understand these aspects of SRS approaches (e.g, appropriateness, intention to use, ease of analysis, perceived ease of use, etc.). Next, we analyze the relation between the top five aspect and other criteria considered in this study.

Table 2. SRS aspects most investigated

Aspect studied	Frequency	Percent
Understandability	19	41.3%
Efficiency	9	19.6%
Correctness	6	13.0%
Defect rate	5	10.9%
Completeness	5	10.9%
Consistency	4	8.7%

2) *In what study settings are these aspects investigated?*

Table 3 indicates that almost 58.7% of the 46 studies took place in an academic environment. Empirical studies in government settings are rarely undertaken.

Table 3. Distribution of study setting

Study setting	Frequency	Percent
Academic	27	58.7 %
Mixed	10	21.7 %
Industrial	8	17.4 %
Government	1	2.2 %

We mapped the 31 aspects being studied against the categories of settings. Table 4 presents the mapping result for the top five aspects. 84.2% of the studies on understandability are carried out in an academic context. Only 10.5% are done in an industry setting. We also note that 40% of the studies on the defect rate aspect are investigated in a mixed context. In addition, the completeness aspect is exclusively investigated in academic settings. However, none of the studies on the top five aspects is investigated in a government setting. This might be a preliminary indication that our knowledge of these aspects has been accumulated one-sidedly and was shaped, by and large, by what university researchers believe it is important to evaluate. This might or might not be what practitioners perceive as important.

Table 4. Aspects studied-types of settings

Aspects	Academic	Industrial	Mixed	Government
Understandability	84.2%	10.5%	5.3%	0%
Efficiency	44.4%	22.2%	33.3%	0%
Correctness	83.3%	0.0%	16.7%	0%
Defect rate	40.0%	20.0%	40.0%	0%
Completeness	100.%	0.0%	0.0%	0%

3) *In what problem domains are these aspects investigated?*

Table 5 shows that the dominant applications are MIS. However, this result is not significant since a large proportion of primary studies do not indicate the type of application used.

Table 5. Aspects studied-types of application

Aspects	Command & Control	E-Commerce	Embedded software	MIS	Real-time	Virtual reality	Not indicated
Understandability	5%	5%	5%	32%	5%	.0%	47%
Efficiency	.0%	.0%	.0%	67%	11%	11%	11%
Correctness	.0%	17%	17%	33%	.0%	.0%	33%
Defect rate	.0%	.0%	.0%	40%	.0%	.0%	60.%
Completeness	.0%	20.%	.0%	40%	.0%	.0%	40.%

4) *What research method was used in the evaluation of the aspect studied?*

Table 6 provides the answer to this question. It suggests that experiments are by far the most used research approach. 67.3% studies relied on experiments. 13% papers only used case study, and 0% were action research studies.

Table 6. Distribution of empirical research

Empirical research	Frequency	Percent
Experiment	31	67.3 %
Case Study	6	13.0 %
Observational Study	4	8.7 %
Experience Report	4	8.7 %
Survey	1	2.2 %
Action Research	0	0.0 %

3. Discussion

This mapping study reported on the SRS aspects being investigated in the RE literature, the settings in which evaluation of SRS takes place, and the research methods being used for such evaluation.

Two main limitations of this study are identified: (i) bias in the selection of publications to be included due to our access to 'relevant' sources depending on the appropriateness of search strings used. The diversity of terms used in empirical software engineering means that we may have missed some relevant studies. Thus, there is a need to develop ontologies for describing the findings of these empirical studies [5]. In addition, exclusion of papers written in a language other than English leads to biased estimates of the effectiveness of the selection process. This could not be avoided since English was the only feasible common language for the revision team. ii) Robust categorizations for analysis; in our research, although four classification criteria were used, we had difficulties with correct identification of 1) the type of empirical study, e.g. most of the papers when they refer to 'case study' in fact mean 'proof of concept'. 2) Aspect studied; as everyone has their own interpretation of what quality term to use. 3) Problem domain; it was not possible to obtain an exhaustive list of all possible domains where business users may decide to use software systems.

Our study revealed that very few real-world case studies have been published. The majority of academic work so far has focused specifically on experiments, meaning that the general applicability of results may be compromised as a result. More technical action research will be necessary in order to understand the problems of using SRS techniques in specific contexts, where stakeholders have different roles and needs that would impact on any empirical evaluation. More research to evaluate SRS techniques in real-life settings is therefore required.

We found 31 aspects of SRS which were studied. A key question is which of these aspects need further study. Clearly, aspects such as understandability, efficiency are important, but which aspects are actually problematic in the real world? Our position is that problematic aspects need to be studied first. It might be the case that understandability is studied so often because it is easier to study in an experimental context, and not because it is the most important problem in the real world. It would therefore be worthwhile as a complement to this systematic mapping study to carry out a systematic review focusing on empirical evaluation of the understandability of SRS.

4. Acknowledgements

5. References

[1] A. Herrmann, M. Daneva: Requirements Prioritization Based on Benefit and Cost Prediction: An Agenda for Future Research, 16th International Requirements Engineering Conference, IEEE Computer Society, 8-12 September 2008, Barcelona, Spain, pp. 125-134

[2] A. M. Davis, Ó. Dieste, A. M. Hickey, N Juristo, A. Moreno, "Effectiveness of Requirements Elicitation Techniques: Empirical Results Derived from a Systematic Review". 14th International Conference on Requirements Engineering (RE 2006), IEEE Computer Society, 11-15 September 2006, Minneapolis, USA, pp. 176-185.

[3] B. Kitchenham O. Pearl, D. Budgen M. Turner J. Bailey, S. Linkman, "Systematic Literature reviews in software engineering – A systematic literature review", Information and Software Technology, 51(1): 7-15, January 2009.

[4] G. Kotonya, I. Sommerville, S. Hall, "Towards A Classification Model for Component-Based Software Engineering Research". 29th. EUROMICRO Conference, Belek-Antalya, Turkey, IEEE Computer Society, 3-5 September 2003, pp. 43-52

[5] J. Calmon, P. Gomes, A. Cruz, T. Uchôa, G. Travassos (2007): "Scientific research ontology to support systematic review in software engineering". Advanced Engineering Informatics 21(2): 133-151

[6] K. Ahmed, "A Systematic Review of Software Requirements Prioritization", Master Thesis in Software Engineering, School of Engineering Blekinge Institute of Technology, Sweden, October 2006.

[7] K. Petersen, R. Feldt, M. Shahid, M. Mattsson, "Systematic Mapping Studies in Software Engineering", 12th International Conference on Evaluation and Assessment in Software Engineering (EASE), Department of Informatics, University of Bari, Italy, June 2008.

[8] P. Tonella, M. Torchiano, D. Du Bois, T. Systa, "Empirical studies in reverse engineering: state of the art and future trends", Empirical Software Engineering Journal, Springer, 2007

[9] R. Pretorius, D. Budgen: "A mapping study on empirical evidence related to the models and forms used in the uml", Proceedings of the Second International Symposium on Empirical Software Engineering and Measurement, ESEM, October 9-10, 2008, Kaiserslautern, Germany, pp. 342-344

[10] T. Dybå, T. Dingsøyr. "Strength of evidence in systematic reviews in software engineering", Proceedings of the Second International Symposium on Empirical Software Engineering and Measurement, ESEM, October 9-10, 2008, Kaiserslautern, Germany, pp. 178-187.

[11] J. Bailey, D. Budgen, M. Turner, B. Kitchenham, P. Brereton, S. Linkman, "Evidence relating to Object-Oriented software design: A survey," First International Symposium on Empirical Software Engineering and Measurement (ESEM), 2007, pp.482-484.

[12] D. I. K. Sjøberg, T. Dybå, M. Jørgensen: "The Future of Empirical Methods in Software Engineering Research" International Workshop on the Future of Software Engineering, FOSE 2007, May 23-25, 2007, Minneapolis, USA, pp. 358-378.