Operationalizing the requirements selection process with study selection procedures from systematic literature reviews

Kai Petersen and Nauman Bin Ali

Blekinge Institute of Technology, Karlskrona, Sweden
{kai.petersen, nauman.ali}@bth.se

Abstract. Context: Software organizations working in a market-driven environment have to select requirements from a large pool to be prioritized and put into backlogs for the development organization.
Objective: This paper proposes an approach based on study selection in systematic literature reviews and translates the concept to requirements engineering. The rational for doing so is that the selection processes used there have been effective (selecting and finding relevant papers) and efficient (possible to use for a high number of studies, in some cases 10,000 research contributions had to be evaluated).
Method: This paper can be classified as a solution proposal, and utilizes hypothetical examples to explain and argue for the method design decisions.
Results: The process proposed consists of three main phases, namely establish selection criteria, evaluate selection criteria, and apply selection. On a more fine-grained level, nine activities are specified.
Conclusion: Given that the process has been effective and efficient in a similar context, our proposition to be evaluated in future research contributions is that the process leads to effective and efficient decision making in requirements selection.

Key words: Requirements selection, study selection, inclusion and exclusion criteria

1 Introduction

Requirements selection is a challenging task, as a very high number of requirements need to be processed. In particular, in a market-driven context this is a challenge where the requirements may come from market units, and also from research and development done within the organization. This phenomenon is referred to as requirements overload [1]. In such a situation there is very little time to spend on an individual requirement, hence instead of assigning a very specific value of priority relative to other requirements, the primary goal is to make a decision of relevance (e.g. on whether to include or exclude the requirement for the product). Thereafter, the requirements that are included as relevant are further investigated and prioritized.

Interestingly, a similar problem to the requirements selection is faced by researchers conducting systematic literature reviews. In systematic literature reviews [2] all relevant studies for answering a research question should be identified. When searching for the studies using keywords, there is usually a lot of noise in the results, and often over 90% of the papers have to be discarded due to a lack of relevance (e.g. [3, 4, 5]). Thereby, it is essential that no relevant papers are discarded during the so-called inclusion exclusion phase, which is usually conducted based on the title and the abstract of the studies. The selection process focuses on making sure that no relevant studies get lost, for example by providing means to objectify criteria, and also identifying the need of clarification (e.g. if an abstract is unclear). At the same time, literature studies have shown to deal with a very high number of decisions, in some cases over 10,000 studies had to be assessed for inclusion and exclusion, while assessing 1,000 studies is a common situation.

The problem translates very well to requirements selection as we have a high number of requirements, we need to define clear inclusion and exclusion criteria to not miss any essential requirements, and have to come to an objective decision in a relative short amount of time for the assessment to be feasible. Thus, in the context of systematic literature reviews, procedures for study selection have been identified, and also evaluated [6, 7]. In this paper, we propose utilizing the selection procedures from systematic literature reviews for requirements selection. We present the process of selection for literature studies, and explain how it could translate to requirements engineering. We also propose how to evaluate the approach to determine improvements to the procedure, and assess the benefit for practice.

The remainder of the paper is structured as follows: Section 2 presents the related work. The solution proposal is described in Section 3, followed by a discussion in Section 4. Section 5 concludes the paper.

2 Related work

The focus of this paper is on requirements selection, which is different from requirements prioritization. In requirements prioritization a comparative analysis of requirements should be made, which is a time-intensive activity [8]. How time intensive it is depends on the prioritization technique (e.g. ranking versus Analytical Hierarchy Process) [8].

In requirements selection (in particular screening), the goal is to immediately remove requirements from the initial set that are not relevant from a system and stakeholder perspective. That is, rather than deciding on the priority the decision is on whether to include or exclude the requirement (referred to as triage, a term inspired from the medical domain [9]).

The related work focuses on introducing the concept of requirements triage, and on existing requirements selection and screening approaches. Further details (e.g. specific criteria for selection) will be revisited in the context of the proposed selection process, where applicable.

Requirements triage: As the pool of resources working with the requirements (e.g. in refining, designing, testing, and implementing them) is limited, the main goal of triage is to select the most relevant requirements to focus on [10]. Davis proposes three main activities: *prioritize requirements* (into categories such as must satisfy, need not satisfy, and could incorporate), *estimate resources*, and *select a subset of requirement* to maximize the probability of success of the product with the available resources in mind. As pointed out by Davis, the process of triage itself is difficult to implement. Based on experience from practice key recommendations were made, such as maintaining a list of requirements, record dependencies between them, link requirements to the estimated effort and importance, etc.

Simmons [9] explains how triage in medicine maps to triage in requirements engineering. An important part of triage is to have objective criteria for classifying patients in medicine considering severity of their condition and effort required to treat them. The criteria examples were objective and classified into physiology (e.g. blood pressure, Glasgow Coma scale), injury anatomy (e.g. flail chest, spinal cord injury), and injury mechanism (e.g. fell of 20 feet or more). Simmons maps examples of criteria for requirements selection to the categories. In particular, he points out that the criteria should be quantifiable through measures and indicators. Though, no concrete measures are presented as in medicine well defined units of measures are defined based on physics and biology research, which is not the case for software engineering, where measures are often ambiguous (e.g. defect density, SLOC).

Also efforts have been undertaken to automate the requirements prioritization process by clustering requirements [11], that are then analyzed and prioritized by a person. Hence, the clustering itself may be beneficial for triage, or at least may find groups of requirements belonging together, which may be treated together in a triage step.

Processes: Khurum et al. [12] propose a *Method* for *Early Requirements Triage* and *Selection* (MERTS), i.e. to conduct the screening of papers. It consists of the following three phases:

- Early requirements triage: Here, first the goal and strategy for the product are defined, stating where to go, how to get there, and defining concrete actions how to get there. Also, weights are assigned to the answers of these questions if more than one answer is given. Hence, the relative priority of the goals becomes obvious. In a sense, this defines the criteria and the importance of the criteria. Each requirement is then assessed against the criteria to determine how much it contributes on a scale 0-100.
- Requirements selection for release: In this step, the road-map is defined based on the assessment of the requirements in the first phase (e.g. with techniques for requirements prioritization), and the resources needed are estimated.
- Reasoning: In reasoning one should argue why the decisions made would lead to success, which is in a sense a reflective step to gain further confidence in the decisions made.

MERTS has been empirically evaluated in a student experiment [13], where students should conduct triage in relation to a predefined strategy using the MERTS templates and using natural language where the product strategy was formulated narratively. It was found that MERTS was achieving significantly more correct answers. Though, a limitation of MERTS is that the requirements have to be on the same level of abstraction, hence a prerequisite is to have abstraction models in place such as [14], otherwise the approach is infeasible in practice (cf. [12]). We contend that the need for comparability is more significant when we have to make a decision about the relative importance, whereas the early selection is primarily about discarding irrelevant requirements as soon as possible.

In selection processes, often multiple stakeholders are involved, which has not been incorporated in MERTS. An important contribution to mention here has been provided by Regnell et al. [15], presenting how to prioritize requirements when multiple stakeholders are involved. The process consists of the activities to establish a candidate list of requirements, individual prioritization by each stakeholder, combining the prioritizes, and all stakeholders giving feedback on results.

Our proposal complements triage processes by prescribing how to systematically involve multiple stakeholders in the selection to objectify the selection criteria and with that potentially increase the effectiveness on making the "right decision" when selecting requirements. Following a systematic process has shown to be beneficial in identifying more relevant studies [7]. Hence, having objectified decision making increases efficiency, and coming to the right decision overall reduces rework. This was motivating the solution proposal presented in this paper.

3 Solution proposal

We first describe the selection process conducted in the context of literature studies. Thereafter, we present how the process translates to requirements selection.

3.1 Study selection process in systematic reviews

The reliability of a systematic literature review relies heavily on the selection process in identifying relevant studies [16]. Therefore, to reduce the effects of personal bias and to bring traceability to the selection process several measures are undertaken. An overview of this process is shown in Figure 1.

Starting with research questions, study selection criteria are formulated in a way that they can be objectively assessed. As the research question is providing the scope of the study, the selection criteria are based on the question. The selection criteria are reviewed by other reviewers for clarity, objectivity of assessment and alignment with the research questions. The "think-aloud" approach of applying the criteria is used on a few papers to externalize a reviewer's understanding of the criteria. This helps to highlight misunderstandings or ambiguities in the criteria. Based on the review and "think-aloud" application of the selection criteria, it may need to be updated. An example of typical inclusion and exclusion criteria selected from [4] is presented below.

> **Example inclusion criteria:** Overall, eight different inclusion criteria were defined, for example:
>
> – The article is peer reviewed
> – The article discusses a model/framework of strategic release planning or post release planning analysis of strategic release planning.
> – The article will be included if it discusses a validation of existing model of strategic release planning or post release analysis.
>
> Three exclusion criteria were included, two examples are:
>
> – Articles related to only operational release planning will be excluded.
> – Articles related to re-planning of a release on operational level will be excluded.

In the example, it is noteworthy that some criteria are easier to evaluate (e.g. peer review), while others require more thought (e.g. assessing whether a model is presented focused on strategic release planning and not operational release planning).

Following this, the reviewers independently apply the selection criteria on a randomly selected subset of papers. This "pilot selection" is a confidence building step to assess the wider applicability of the criteria. For example, we may find out that in a majority of the papers, the information we seek to assess a paper's relevance is not available in the title and abstract. Hence, the assessment could not be made based on that information in title and abstract of a paper alone, which is usually the first step before reading the full-text.

Furthermore, using inter-rater agreement statistics like Kappa [17] we can assess the level of agreement between the reviewers and indirectly the unambiguity of formulation and the objectivity of the criteria. Different researchers [18, 19] defined ranges of Kappa values to define the quality of agreement. For example, if a Kappa value of 0.85 would be obtained Landis and Koch [18] would classify it as almost perfect agreement. All the disagreements should be discussed (especially with a low Kappa value) to understand the reasons that lead to it and the selection criteria should be further elaborated to avoid these disagreements in the future. Depending on the results of the pilot and changes to the criteria, we may need to repeat the pilot till we are confident in the objectivity of the criteria.

Now, we are ready to perform selection on the complete set of the papers. The recommended practice is to have at least two reviewers assessing the relevance of each paper [20]. This improves the reliability of the results as there is a *"safety net"* if one reviewer makes a mistake (which is highly likely given

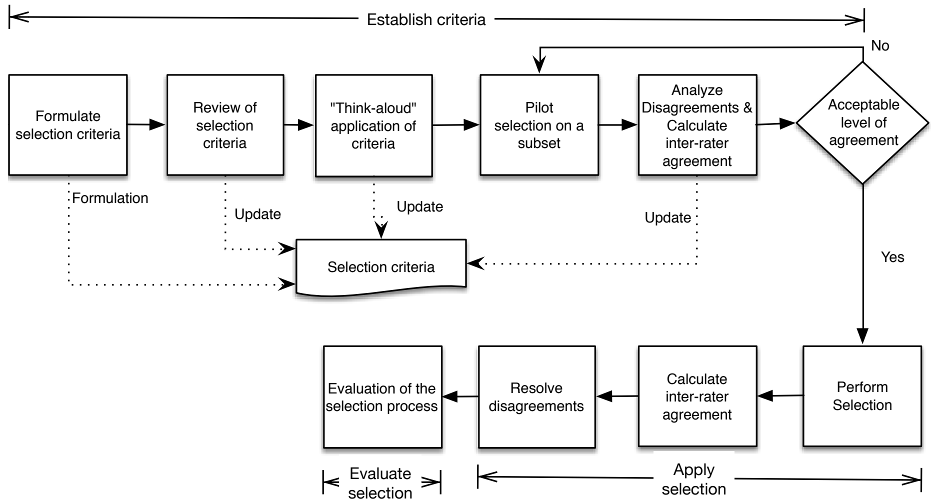

Fig. 1. Overview of the selection process

the typically high number of papers they have to sift through). In systematic literature reviews, the reviewers often use three categories to place each paper in terms of its relevance to the research questions: include (i.e. relevant), exclude (i.e. irrelevant) or uncertain (i.e. cannot make a decision based on the available information).

Even with all the measures that were taken to attain high objectivity in the criteria (which will likely be visible with the high-level of inter-rater agreement between the reviewers), there will still be mismatches between individual reviewer's judgments. The quality of reporting of the paper, quality of our criteria and also mistakes made by the reviewers will be the likely reasons for these mismatches.

In systematic literature reviews, various rules have been suggested and their implications on the results have been evaluated [6]. These rules take into account the level of disagreement (e.g. the extreme case is where some reviewer considers a paper as relevant and the others judge it as irrelevant) and the likely potential of the relevance of a paper (e.g. if more reviewers consider a paper as relevant and a few are uncertain, it tends to show relevance of the paper) to decide about the amount of effort that should be further spent on the papers or to exclude them altogether.

3.2 Requirements selection process

When approaching the study selection process from the perspective of requirements selection, the first obvious difference is in the input to the process. In study selection, it is the research questions that communicate the aim and goal of the study and guide the formulation and application of the selection criteria. In case of requirements selection such guidance would come from the product vision, goals, and strategies to reach the goals. Specifying product strategy and goals (cf. [21], [13], [22]) is beyond the scope of the selection process, but this provides critical input to it.

Apart from this difference of input to the process, we argue, that there is no modification required in activities to be conducted and their order [7] when using it for requirements selection. Though, on a more detailed level adaptation is needed for the requirements domain. These are discussed further in the remainder of the section.

Establish criteria Like the study selection process (as exemplified in Section 3.1), when specifying the selection criteria for requirements it is important to specify both inclusion and exclusion criteria. For example, questions detailing elements of the criteria could be like the following:

- Include a requirement when it fits customer goals.
- Exclude it if it is affecting a large part of the architecture, i.e. requires a relatively large effort.

It is important to highlight that like systematic literature reviews there are no fixed criteria applicable to all, and that individual companies need to investigate and set the criteria that are suitable for their context. Some useful input however can be taken in the form of attributes to be considered,criteria for prioritization, and guidance on how to formulate criteria from: Khurum et al. [23], and Riegel and Dörr [24].

The review of the criteria can be done in a meeting and employing techniques like "think-aloud" application (where one practitioner while applying the criteria also verbalizes his understanding of it). These steps help ensure alignment with goals, clarity of the formulation, and facilitate to align various perspectives (if practitioners have different roles in the product development e.g. product manager and lead architect).

Thereafter, a small scale dynamic evaluation of the criteria is performed by doing a pilot selection on a subset of requirements. Using inter-rater statistics level of mismatches between individual judgments can be evaluated and thus highlighting the need to further clarify the criteria or in turn the product goals from which these were derived. It is important to emphasize that this is an iterative process and may involve a few iterations before we have sufficient confidence in the criteria, and alignment between practitioners about the goals, criteria and their interpretation.

None of the existing selection procedures discussed in Section 2 provide such a systematic and structured approach that is geared to align various perspectives early on in the software requirements selection process and also reduces bias in selection.

Perform selection In these steps, the actual selection is performed. Like the study selection process it will be beneficial to have at least two practitioners judge each requirement. We argue that these practitioners should be the product owner and the lead architect of the product to somewhat represent both business and technical perspectives, this is very important as often their views are not aligned and can result in inclusion of unnecessary requirements [25].

Evaluate selection Using inter-rater agreement and visualizing disagreements helps in the following ways:

- Making a conscious decision on how to resolves the mismatches in judgment
- Achieve alignment between stakeholders from business and technical perspective by highlighting disagreements upfront
- Also highlights the need to further clarify the product strategy/goal
- Avoiding overloading the development process by removing irrelevant requirements

The process in study selection is relatively easier to evaluate in terms of how many irrelevant papers got through the selection process and were later rejected in the study. Similar analysis could be done where the number of requirements that are later closed could be an indication of the quality of the selection process. Such requirements could also be an indication of unclear strategy or a change in it, but it is worth starting the analysis from this measure.

The number of relevant studies that are missed by the process can only be identified if the study is repeated or if another set of researchers independently perform the study. Similarly, a reflection on whether some relevant requirements were missed in the selection process will only be possible after the product has been on the market and through analysis of market success of features in competitors products. This could still however be a reflection on the product strategy and less on the selection process.

As highlighted earlier, some criteria are easier to judge than others. Hence, before checking all criteria, the objective and easy to apply ones should be checked first. This reduces the number of requirements to be checked for the criteria that are harder to evaluate. For example, if a requirement fits to the product strategy may be easier to check rather than determining the maximum effort a requirement should be allowed to consume before going into the backlog of the development organization.

The results of selection with two reviewers and three categories as suggested by Davis, can be visualized in a way similar to as shown in Table 1. Based on this table, we can now define decision rules how we deal with the requirements. We can, for example, say that requirements were both practitioners stated "Must satisfy" go into the backlog and are further refined and implemented. In systematic literature review, a category of "uncertain" has been introduced indicating that no decision could be made without further discussion and reflection. In general, if there are disagreements (see D in Table 1) discussions should take place, such as adding additional practitioners from marketing, product development, etc.

The strength of the proposed process is however in ensuring that the quality of product strategy/vision/goal is translated into only appropriate requirements getting selected through it. The success on the market may still be as good as the market analysis and business strategy, however, the operationalization

Table 1. Different scenarios for requirements selection with two practitioners (adapted from [7]).

Practitioner 1		Practitioner 2		
		Must satisfy	Could incorporate	Need not satisfy
	Must satisfy	A	B	D
	Could incorporate	B	C	E
	Need not satisfy	D	E	F

through this process of the product vision increases the chances of alignment of development with it.

4 Discussion and outlook

We first discuss the potential impact on the effectiveness of the requirements selection process. The goal of increasing the effectiveness is fulfilled if:

- the likelihood of excluding relevant requirements is reduced.
- the likelihood of including irrelevant requirements is reduced.

As we observed from utilizing study selection process in a variety of systematic reviews [5, 26], and by empirically evaluating them [7], the improvement of effectiveness may be supported in requirements engineering as well, which is achieved as follows:

Making criteria explicit: The first step is to make the criteria explicit through review, using "think-aloud" and pilot selection. When there are disagreements this leads to a discussion and reformulation of criteria. It may also lead to the addition of further items to the criteria, or may even indicate that the goals and strategies are not formulated well enough. As a consequence, the criteria are not concrete enough so that multiple practitioners could not come to the same conclusion while deciding on the same requirement.

The importance of checking alignment: A development organization has to be aligned in its goals and understanding of what the system should ideally deliver. Barney et al. investigated the alignment of different roles with regard to their priorities on functionalities, project metrics, and quality. They found misalignment in the organization. The decisions in Table 1 make this alignment explicit and also traceable, as the decisions of different people should be recorded for each requirement. Furthermore, the alignment can be measured using Kappa statistics. Only with the misalignment being explicit, organizations can take concrete actions to achieve a higher degree of alignment.

Identify the need to improve requirements: Unclear and ambiguous requirements may go into the "uncertain" category and then specific actions have to be defined for that category (e.g. discussion, refine the requirement). If a very high number of requirements always ends up in the uncertain category, then there would be a clear need to investigate this. The reason may be the lack of clarity of the criteria, but also the formulation of the requirement that may be too abstract and does not provide concrete enough information. This situation can also be found in literature studies, where the abstract does not specify important information to make a decision.

Efficiency may also be improved, which is very important when making decisions on a very high number of items. As mentioned earlier, in literature studies it is common to make decisions on 1000 and more studies. For this, we provide the following suggestions.

Use multiple level of criteria: Not all criteria are equally difficult to evaluate. Hence, if a criterion is very easy and objective to apply it should be done first, thereby the total number of requirements to be looked at may be reduced. For example, basic and advanced criteria could be defined.

Use multiple levels of inclusiveness: Based on Table 1 different strategies could be followed. That is, we could define decision rules which requirements to take to the next step (e.g. detailed analysis of the requirements). The most inclusive strategy would be to take everything to the next step (i.e. A+B+C+D+E+F). Though, as we found in the context of literature studies the most inclusive strategy created overhead, while the overhead did not justify the value added to our results. Hence, finding good decision rules giving a good return (effectiveness) on investment (time spent to investigate requirements) may lead to efficiency gains and hence the ability to handle a high number of requirements.

Learning: When using the criteria identified for selection early on, time has to be invested to objectify criteria and spread them in the organization. As time goes on and decisions were made and compared, people in the organization will most likely familiarize with the criteria, which may decrease the time required for decision making.

In order to substantiate the propositions stated above, controlled experiments and industry evaluations of the process are the next steps to take in this research.

Controlled experiments: Khurum et al. [13] conducted an experiment to evaluate MERTS. They compared MERTS with the selection of requirements only using the narratively formulated goals. Similarly, we propose to compare the situation with narratively formulated text with having a set of defined inclusion and exclusion criteria. The objectivity of the criteria has to be determined prior to running the experiment. In the experimental package defined by Khurum et al. the relevant requirements were defined. Hence, the package would allow to evaluate the effectiveness as defined above.

Industry evaluation: Before conducting a live evaluation of the process, it should first be presented to the industry to gather feedback on the process. Thereafter, the process should be utilized by the practitioners, and feedback should be gathered (e.g. through interviews and questionnaires). It may also be important to participate in meetings taking place in relation to the process. This is useful to see how practitioners achieve agreements and decide to reformulate the inclusion and exclusion criteria, or even their product goals and strategies. Within the scope of the process, it is important to evaluate whether effective selection has been achieved with respect to the goals and strategies. Outside the process, the satisfaction of the customers has to be looked at, even though this leads to other confounding factors playing a role as well.

5 Conclusion

The proposed process provides a systematic way for requirements selection. The process is inspired by the selection of empirical studies in systematic literature reviews. The challenges in both processes are very similar, i.e. making an effective selection with respect to specified goals/research questions for a very high number of decision items (requirements/studies) This process can potentially contribute in the following ways:

– provides a transparent process where a conscious traceable decision is made about relevance of requirements
– brings multiple perspectives in the selection process, promotes communication and highlights the need for alignment
– reduces the likelihood of excluding relevant requirements
– reduces the likelihood of including irrelevant requirements

We have suggested concrete studies on how to proceed with the evaluation of the proposed process using controlled experiments and industrial case studies.

References

1. Potts, C.: Invented requirements and imagined customers: requirements engineering for off-the-shelf software. In: Requirements Engineering, 1995., Proceedings of the Second IEEE International Symposium on, IEEE (1995) 128–130
2. Kitchenham, B.: Procedures for performing systematic reviews. (2004)
3. Loniewski, G., Insfran, E., Abrahão, S.: A systematic review of the use of requirements engineering techniques in model-driven development. In: Model driven engineering languages and systems. Springer (2010) 213–227
4. Svahnberg, M., Gorschek, T., Feldt, R., Torkar, R., Saleem, S.B., Shafique, M.U.: A systematic review on strategic release planning models. Information and Software Technology **52**(3) (2010) 237–248
5. Ali, N.B., Petersen, K., Wohlin, C.: A systematic literature review on the industrial use of software process simulation. Journal of Systems and Software **97**(0164-1212) (2014) 65–85
6. Petersen, K., Ali, N.B.: Identifying strategies for study selection in systematic reviews and maps. Empirical Software Engineering (2011)
7. Ali, N.B., Petersen, K.: Evaluating strategies for study selection in systematic literature studies. In: International Symposium on Empirical Software Engineering and Measurement, ACM (2014)
8. Berander, P., Andrews, A.: Requirements prioritization. In: Engineering and managing software requirements. Springer (2005) 69–94
9. Simmons, E.: Requirements triage: what can we learn from a" medical" approach? Software, IEEE **21**(4) (2004) 86–88

10. Davis, A.M.: The art of requirements triage. Computer **36**(3) (2003) 42–49
11. Duan, C., Laurent, P., Cleland-Huang, J., Kwiatkowski, C.: Towards automated requirements prioritization and triage. Requirements engineering **14**(2) (2009) 73–89
12. Khurum, M., Aslam, K., Gorschek, T.: A method for early requirements triage and selection utilizing product strategies. In: Software Engineering Conference, 2007. APSEC 2007. 14th Asia-Pacific, IEEE (2007) 97–104
13. Khurum, M., Gorschek, T., Angelis, L., Feldt, R.: A controlled experiment of a method for early requirements triage utilizing product strategies. In: Requirements Engineering: Foundation for Software Quality. Springer (2009) 22–36
14. Gorschek, T., Wohlin, C.: Requirements abstraction model. Requirements Engineering **11**(1) (2006) 79–101
15. Regnell, B., Höst, M., och Dag, J.N., Beremark, P., Hjelm, T.: An industrial case study on distributed prioritisation in market-driven requirements engineering for packaged software. Requirements Engineering **6**(1) (2001) 51–62
16. Wohlin, C., Runeson, P., da Mota Silveira Neto, P.A., Engström, E., do Carmo Machado, I., de Almeida, E.S.: On the reliability of mapping studies in software engineering. Journal of Systems and Software **86**(10) (2013) 2594–2610
17. Cohen, J.: Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological bulletin **70**(4) (1968) 213
18. Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. biometrics (1977) 159–174
19. Fleiss, J.L., Levin, B., Paik, M.C.: Statistical methods for rates and proportions. John Wiley & Sons (2013)
20. Kitchenham, B., Brereton, P.: A systematic review of systematic review process research in software engineering. Information & Software Technology **55**(12) (2013) 2049–2075
21. van Lamsweerde, A.: Goal-oriented requirements engineering: a guided tour. In: Requirements Engineering, 2001. Proceedings. Fifth IEEE International Symposium on. (2001) 249–262
22. Hauser, J.R., Clausing, D.: The house of quality. (1988)
23. Khurum, M., Gorschek, T., Wilson, M.: The software value map an exhaustive collection of value aspects for the development of software intensive products. Journal of Software: Evolution and Process **25**(7) (2013) 711–741
24. Riegel, N., Dörr, J.: A systematic literature review of requirements prioritization criteria. In: 21st Int. Working conference on Requirements Engineering: Foundations for Software Quality, Springer (2015)
25. Barney, S., Wohlin, C., Chatzipetrou, P., Angelis, L.: Offshore insourcing: A case study on software quality alignment. In: Global Software Engineering (ICGSE), 2011 6th IEEE International Conference on, IEEE (2011) 146–155
26. Edison, H., Bin Ali, N., Torkar, R.: Towards innovation measurement in the software industry. Journal of Systems and Software **86**(5) (2013) 1390–1407