Towards a Research Framework on Requirements Prioritization

Patrik Berander *, Kashif Ahmed Khan *
Blekinge Institute of Technology
Department of Systems and Software Engineering
PO Box 520
SE-372 25 Ronneby
+46 (0)457 385840
{patrik.berander@, kakb04@student.} bth.se

Laura Lehtola *
Helsinki University of Technology
Software Business and Engineering Institute
P.O.Box 9210
02015 TKK, Finland
+358(0)9 4516295
laura.lehtola@tkk.fi

ABSTRACT
There exist a large number of approaches for prioritization of software requirements. Despite of several empirical studies, there is still a lack of evidence of which approaches that are to prefer, since different studies have resulted in different conclusions. Reasons may be due to differences in contexts, variables measured, and data sets used. This paper presents a research framework for studies about requirements prioritization, which aims to enable building a more consistent knowledge base and stronger evidence. The framework facilitates comparison, replication, and high-level analysis of prioritization approaches by proposing suitable variables to measure. The basis of the framework comes from a systematic review conducted on requirements prioritization techniques, and is further refined through literature studies of similar frameworks in related areas, and in a research workshop. The framework supports researchers in conducting and reporting prioritization studies, and supports practitioners in getting information about different approaches.

Categories and Subject Descriptors
D.2 Software Engineering

General Terms
Management

Keywords
Requirements prioritization, requirements engineering, research framework

1. INTRODUCTION
Software engineering decision support plays a vital role in the value generation processes of software products [29], as it facilitates making the right decisions in order to develop the right products. Requirements prioritization has been recognized as one of the most important decision activities in the requirements engineering area [29] supporting such decisions.

SERPS'06, October 18–19, 2006, Umeå, Sweden.

Although there is rather much research made within the requirements prioritization area, little evidence exists regarding which approaches that are to prefer over others in different situations and environments. This is especially supported by the fact that different studies have shown different results when comparing the same or very similar approaches. Hence, it is not possible to get a real understanding of whether one approach is better than another. One reason for this may be that different studies are done in different contexts, measure and report different variables, and use different data sets. This makes it hard to compare different studies with each other to get an understanding of when different prioritization approaches are suitable to use. It also makes it hard to replicate studies, which is essential for maturing the body of knowledge in the area [39].

This paper addresses the existing problem by proposing a research framework that guides researchers on what to measure and report when conducting studies within about requirements prioritization. The purpose is to create a better and more consistent knowledge base. The recommendations presented in the framework can preferably be used as a checklist to make sure that nothing is missed when planning, conducting and reporting empirical studies. Further, the framework open up possibilities for researchers to analyze results from different studies at a high level in order to draw conclusions based on several studies. In addition, by having results presented consistently, it will be possible for researchers to easier and more confidently decide which studies to conduct and replicate to improve the evidence in the area. The objective with the framework is also to give support on how to design a study by specifying what variables to measure. Last, but not least, this may in the long run give opportunities for industrial professionals to give for example product type, type of requirements, number of requirements as input and get the most suitable prioritization approach as output.

To sum up; the framework presented in this paper is needed for guiding study arrangements and data collection, and to facilitate replication of studies and to get commensurable research results. The framework was developed through an analysis of the results from a systematic review (see e.g. [23]) and by looking at similar frameworks from other areas within software engineering. In addition, a two day workshop between researchers was conducted to complement the framework by consolidating the views of the three authors.

This paper is structured as follows. In Section 2, requirements prioritization is introduced together with a discussion about different levels of requirements prioritization. In addition, the

* All three authors have contributed equally to this paper.

status of the empirical evidence in the area is presented. In Section 3, the background for the framework and the work process used to develop the framework is presented. The proposed research framework for requirements prioritization is presented in Section 4. Section 5 presents an analysis of how well existing papers fulfill the framework while Section 6 discusses and concludes the paper.

2. REQUIREMENTS PRIORITIZATION

Requirements prioritization is defined as an activity during which the most important requirements for the system (or release) should be identified [38]. Commonly, only a limited set of requirements can be implemented in one release, but the product should, however, meet the needs of the customers and reach the markets in time [16]. When customer expectations are high, timelines short, and resources limited, the most essential functionality of the product should be delivered as early as possible [40] at the same time as the scope of each product release must be limited [37]. The challenge is therefore to select the 'right' requirements out of a given superset of candidate requirements so that all the different key interests, technical constraints and preferences of the critical stakeholders are fulfilled and the overall business value of the product is maximized [33]. Requirements facilitate this value generation process by delivering structures for finding those requirements that are most valuable to implement.

It is widely accepted that requirements prioritization is a challenging and complex decision-making activity [6][26][27]. An important requirement in one release for a certain customer may not be as important in the next release for another customer [6]. Also, customers may be skeptical towards prioritizing because they are afraid that only the most important requirements get implemented. At the same time, developers may be skeptical since they do not want to admit that they are not able to implement all the requirements [40].

Despite of (or thanks to) the challenging nature of prioritization, several requirements prioritization approaches have been introduced. Such different approaches work on different measurement scales, focus on different aspects, and have different levels of sophistication [3]. The prioritization approaches introduced vary from high-level prioritization process descriptions to detailed prioritization algorithms. Approaches on different levels typically focus on solving some parts of the prioritization problem and put less emphasis on other challenges. Since different levels focus on different parts of the prioritization, it is not suitable to compare approaches from different levels.

One problem in the area of requirements prioritization seems to be that there does not exists a common vocabulary for how to denote the different levels, and what it is that characterizes each level. This has resulted in that different researchers use different names to denote approaches at the same level (e.g. method, technique, process, and algorithm). Since this vocabulary confusion may introduce problems, especially when it comes to empirical studies within the area, there is a need to structure the approaches into levels and to suggest what vocabulary that should be used for each level.

2.1 Hierarchical Division of Approaches

To clarify the field of prioritization and to form a common vocabulary, a hierarchical division of prioritization approaches is presented in this section. The four levels presented were defined by studying existing prioritization approaches (as part of the workshop presented in Section 3.2) and identifying commonalities and differences with regards to their characteristics. These different levels are yet rather coarse grained, and need further refinement. Since four different levels are proposed, and because different vocabulary is used at each level, the word approach is used in this paper when referring to all levels.

The recognition of hierarchy levels and a common vocabulary in the area is an important tool for enabling comparison of approaches with each other, and for understanding the viewpoint that they have for prioritization. Below, the levels are presented where higher-level approaches typically utilize lower-level approaches.

Level 1 (Activities): In all prioritization approaches, some activities need to be done by the people prioritizing the requirements to get the requirements prioritized. This level refers to these underlying activities where e.g. requirements are compared to each other pair-wise, Monopoly money are distributed between requirements, notes are used to put requirements in piles, etc.

Level 2 (Techniques): Prioritization techniques use the results from the activities level as input, possibly do some calculations with the data, and then present the priorities. The way the priorities are presented is unique for each technique. One example of a technique is numeral assignment [14], which presents priorities ordered in groups but do not prescribe how to end up with the result (activity). Other examples of techniques are: Analytical Hierarchy Process (AHP) [35][36], Binary Search Tree [18], and Hierarchical Cumulative Voting (HCV)[4].

Level 3 (Methods): Prioritization methods are usually specific to requirements engineering and are more sophisticated than techniques. Further, while a technique commonly focuses on getting a priority on one aspect (e.g. importance) that is not defined by the technique, a method commonly takes more variables (e.g. importance, dependencies) into account and also defines which ones to use. For example, the Cost-Value approach uses cost and value priorities from AHP (technique) to calculate a value-cost ratio and presents the result in a graph as input to release decision. Other examples of methods are EVOLVE [10], Quantitative Win-Win [33], and Wiegers' method [40].

Level 4 (Process): Prioritization process refers to the description of the steps needed in the organization to prioritize the requirements. This level includes issues like in which order things should be done, how different stakeholders should co-operate, and how prioritization of requirements fits to the selected software process etc. Examples of processes can be seen in [18] and [30].

2.2 Evidence on Requirements Prioritization

According to Ngo-The and Ruhe [29], there is rather much research done within the prioritization area (e.g. [18][20][25]). Nevertheless, when investigating the research made within the area, there seems to be little evidence regarding which approaches that are to prefer over others and in what situations and environments. This observation is especially supported by the fact that different studies have shown different results when comparing the same (or very similar) approaches.

In order to more structurally investigate the level of evidence present in the area, and if the initial observation is right or wrong, a systematic review of requirements prioritization was conducted within the scope of a master thesis (see [22]). A systematic review is "a means of evaluating and interpreting all available research relevant to a particular research question, topic area, or phenomenon of interest" [23] that stems from medical sciences. Within software engineering, a few attempts to conduct systematic reviews have recently been reported (e.g. [8][28]). The systematic review of requirements prioritization was conducted by one of the authors of this paper by carefully following the available guidelines about how to conduct systematic reviews (e.g. [5][21][23]). The reviewer can be regarded as independent, since no studies of his own were included. To limit the scope of the systematic review, and to find comparable studies, only approaches at the "technique" level (see Section 2.1) were included. This further means that studies comparing approaches at different levels, although one or more was at the “technique” level (e.g. [25]), was excluded from the study. However, the proposed framework in this paper covers all four levels, since variables needed to report the studies are not that level-dependent.

Within the systematic review, a comprehensive search strategy was utilized where eight papers finally qualified to be included, based on the study selection criteria. This can be seen as a small number of papers, but these were the only ones presenting empirical studies with the approaches investigated exclusively at the “technique” level. The papers were carefully examined according to the predefined quality assessment questions and the data from the studies were extracted into a data extraction form. As a part of the systematic review, the results where synthesized in order to identify commonalities and differences in the results, along with reasons for potential differences. The process of doing the systematic review is not in the focus of this paper and will therefore not be described in detail. Detailed description of the study can be found in [22].

When synthesizing the results, it was apparent that different studies came up with different results. For example, in Karlsson [14] it was concluded that the Analytical Hierarchy Process (AHP) was superior to numeral assignment regarding time consumption, while the studies by Karlsson *et al.* and Ahl showed that Planning Game (which basically is an extended way of doing numeral assignment where also ranking is introduced) was superior to AHP for the same variable (time) [1][19].

The problem discussed above may not be huge if it is possible to analyze why, when, and how the results differ. However, as the result showed, different studies reported differently about the study setting (e.g. some did not report on experience of subjects, others not on the number of subjects) as well as measured different variables (e.g. some did not measure time, others not accuracy), and in different ways (see Section 5 for more information). This means that it was not possible to get any answers as to why the results differed, and when one technique is to prefer over another. Hence, clear evidence of whether one technique is better than another could not be found, even though some preliminary evidence of course is better than no results at all. It is apparent that more studies are required to really get knowledge about when a prioritization approach is useful. Further, it is important that such studies report on the same variables, in the same way, to make findings comparable and hence be able to determine the suitability of different approaches.

3. CREATION OF THE FRAMEWORK

As can be understood by the previous section, some structure within the research field studying requirements prioritization approaches is needed. The sections below present the background of the framework presented in this paper (Section 3.1) as well as the process of creating it (Section 3.2).

3.1 Background

In currently available studies, different variables are not seldom left out when reporting the results. The reasons for this may differ, but one reason could be that the variables are not seen as essential in that isolated study. Instead, the focus is put on finding out if a technique or method is better than another one for some isolated variable. However, to get reliable evidence, different prioritization approaches needs to be compared with each other, and tested in various environments with different aspects (e.g. importance, cost). However, this situation is not unique for requirements prioritization, but has been recognized in other areas of software engineering as well. Hence, it is interesting to investigate how such situations have been addressed in other areas.

In Gallis *et al.*, a framework for evaluating pair programming is suggested [9]. In this framework, independent (e.g. technique), dependent (e.g. quality), and context variables (e.g. type of task) are proposed to be measured and monitored in different research studies related to pair programming. Similarly, Runeson *et al.* presents a framework for comparing different models to identify fault-prone components by giving researchers a template for how to characterize and evaluate the capabilities of different models [34]. The reasons that are highlighted as the motivation for the frameworks in these two papers are basically the same reasons as the ones highlighted in this paper. Hence, it may be a good idea to use these as input when developing a similar framework in the area of requirements prioritization.

3.2 Process of Creating the Framework

The first step when eliciting and deciding upon the variables to include in the framework was to consult the result from the systematic review (see Section 2.2) and extract the variables that were reported in the included studies. With the systematic review as input, a two-day workshop was held where two of the authors of this paper (the ones not involved in the systematic review) used the variables found in the studies included in the systematic review and complemented them with variables from studies about prioritization approaches that had been excluded from the systematic review. Studies excluded from the systematic review include (but are not limited to) studies focusing on prioritization approaches on other levels than “technique” level (see 2.1). Furthermore, papers representing similar frameworks in other related areas (e.g. [9][34]) were studied, which made it possible to find relevant variables that have been reported in other areas but not yet within the requirements prioritization area. In addition to these sources, previous experience from research within the area of requirements prioritization was used as input for the work. This means that the systematic review was the main input for the framework, but the framework was complemented during the workshop and from additional literature sources. The main reason for complementing the systematic review was to take into account variables missed in the included studies as well as including

variables that reside on other levels than technique level (see Section 2.1). Based on these inputs, variables were defined and classified into three main classes: dependent, independent, and context variables.

When all variables were defined, classified, and structured, the third researcher was invited to critically evaluate the framework. The exclusion of the third author before this point was done to allow him to be able act as an outsider without preconceived thoughts about what it should look like (even though it was not entirely possible due to his work with the systematic review). This made it possible for him to come up with improvement suggestions for the framework. These improvements were then discussed and accounted for in the framework.

4. RESEARCH FRAMEWORK

This section presents the research framework for requirements prioritization. The framework provides a checklist to facilitate decisions related to how to conduct studies in a suitable way (i.e. what variables to measure and how to design a study), which will in the long run facilitate decisions on what studies to conduct and replicate (i.e. where are the weak spots). In addition (if followed), the long-term effect would be that it facilitates high-level analysis where the results from different studies can be analyzed to determine when an approach is suitable or not. Furthermore, practitioners can study the results in a more structured way and the results could hence be very valuable for decision makers in industry when choosing between different prioritization approaches. Hence, the framework should be seen as recommendations of what to collect and measure and can be used as a checklist to make sure that no crucial variables are omitted.

In Table 1, an overview of the variables in the framework is presented. The framework consist of three the three common kinds of variables [7]: independent, dependent, and context. Even though this framework may seem extensive at a first glance, most variables are easy to report and do not require any additional effort by the subjects.

Table 1 Variables in the Research Framework

Independent Variables	**Dependent Variables**	**Context Variables**
Qualitative Process Description	Time	Environment
Goal	Accuracy	Study Setup
Hierarchy level	Ease of Learning	Subjects
Input	Ease of Use	Requirement
Output	Fault t Tolerance	
	Scalability	
	Understandability of Results	
	Attractiveness	

The independent variables are further discussed in Section 4.1, the dependent variables in Section 4.2 and the context variables in Section 4.3.

4.1 Independent Variables

An independent variable is a variable that cause, influence, or affects outcomes of a study [7]. In prioritization studies, the independent variable is the approaches since changing approach affect the outcome. However, besides just stating which approaches of investigation, it is important to report some additional characteristics about the approaches. This is especially important since authors sometimes use different names for the same approach or the same name for different approaches.

4.1.1 Qualitative Process Description

A prioritization approach is not always applied as described in the original source, even though the researchers may think it is. Therefore, a description of the approach used should be reported as well. When doing so, it is important to describe the implementation of the approach and how the prioritization was actually conducted in the study (e.g. order of activities during the study, instructions to the subjects, if discussions between the subjects were allowed, and how the results were evaluated).

4.1.2 Goal

Explicating the ultimate goal of the prioritization done in the study is important, since the goodness of an approach is dependent on the purpose for what it is needed. For example, sometimes the intention is to find a few killer features, sometimes the intention is to put thousands of requirements in order, while sometimes only two or three piles of requirements is needed. Of course, different approaches are able to handle these different goals differently well and hence it is important to report. Examples of goals could for example be to create a release plan or to get a feeling of which the most important requirements are.

4.1.3 Hierarchy Level

As discussed in Section 2.1, prioritization approaches belong to different hierarchical levels. When presenting a study, it should clearly be pointed out to what level the evaluated approaches belongs. In relation to this, it should be noted that approaches on different levels should not be compared. By following and reporting according to the classification presented in Section 2.1, it is possible for fellow researchers to compare results from different evaluations with approaches at the same level.

4.1.4 Input

When reporting about the approaches under investigation, it is important to report on the inputs that are used for the approach. Here, inputs are divided into two different parts.

Lower-level approaches used: As can be seen in Section 2.1, higher level approaches commonly use one or several lower level approaches as input. For example, the Cost-Value approach uses AHP which is in turn uses pair-wise comparisons. Since high-level approaches can be implemented in many different ways, the usage of lower-level approaches should be reported (e.g. if using HCV instead of AHP in the Cost-Value approach).

Aspects taken into account: Prioritization can be performed with many different aspects (commonly referred to as factors, criteria, etc.) in mind, such as importance, cost, and risk [3]. When reporting a study, it is important to clearly point out what aspects were accounted for by the subjects when conducting the prioritization. It is also important to be clear about the definition of the aspects used. For example, when using value as an aspect, were the participants evaluating user value or the overall value for

the organization? Furthermore, what information did the subjects base their priorities on?

4.1.5 Output

Output refers to the form of results from using the prioritization approach (i.e. measurement scale). The output depends on the approach and could be described for example as “a ranked list”, “grouped in three importance groups” or “percentage of contribution to total value”. Output may have a significant effect on understandability of results and is therefore important information to share.

4.2 Dependent Variables

A dependent variable is a variable that depends on the treatment of independent variables (see Section 4.1), i.e. it is an outcome of the influence of the independent variables [7]. Dependent variables are used to give an answer to when, why, and how an approach is to prefer over another. A dependent variable is the measurement criterion that is chosen when evaluating an approach. In this section, a number of dependent variables that have been considered as valuable in different research studies are presented. Several of the dependent variables can be measured both as perceived (subjective) and actual (objective) and it should be noted that these two different measures often can be combined. Note also that results can be measured both in an absolute (one approach in isolation) and in a relative (comparison between approaches) manner.

4.2.1 Time

Time is commonly crucial in industrial situations, which makes time consumption a very important variable when judging an approach. Time has been measured in several studies (e.g. [1][30]). Further, time is often correlated with scalability (Section 4.2.6), i.e. if the time consumption increases significantly, an approach is probably less scalable. Perceived time consumption is measured by simply asking the subjects about how much time it took to perform the prioritization (i.e. how long time they thought it took). Actual time consumption, on the other hand, is measured by monitoring time manually or automatically. When measuring time consumption, it is of course also possible to correlate the results with, for example, accuracy, to get results on how efficient an approach is.

4.2.2 Accuracy

Accuracy is a measure of how well the results correspond to 'gut feeling' of what the subjects of the study see as their correct priority order. It is important to measure the accuracy of an approach since incorrect results would imply wasted time. Perceived accuracy could be measured by asking the subjects how accurate they thought the result of the approach would be, or how well they thought the resulting priorities corresponded to their view (see e.g. [30]). Actual accuracy would rather be measured by letting the subjects compare the results of different approaches (preferably through a blind test) in order to determine which is most accurate (e.g. [19]).

4.2.3 Ease of Learning

It is important to measure how easy an approach is to learn in order to get an understanding of what is required by the subjects to use the approach efficiently. Measuring this variable is useful in two different ways when determining when to use the approach; 1) the required competence level, and 2) the required familiarity with the approach. Ease of learning is preferably measured through perceived ease of learning, i.e. asking the subjects how easy it was to understand what to do and how to do it (e.g. [19]).

4.2.4 Ease of Use

Ease of use means how easy it is for an experienced user of the approach to use it in every-day-work. By measuring ease of use, it is possible to get an understanding of how convenient the subjects find the approach to be. It is important that an approach is easy to use in order to get people to prioritize, and to get it done in a correct way. Ease of use is preferably measured by how easy the subjects perceive the usage of the approaches in the study (e.g. [18]). It may also be possible to investigate the actual ease of use by utilizing some kind of usability tests, but no such attempts have been reported so far.

4.2.5 Fault Tolerance

Fault tolerance is in some ways related to accuracy (Section 4.2.2) since it measures to what extent judgment errors affects the accuracy (e.g. what effects an incorrect assignment of a priority has for the end results). This measure is preferably measured by perceived fault tolerance where the subject is asked how much he/she thinks that judgment errors affect the end result (e.g. [19]).

4.2.6 Scalability

The number of requirements an approach can handle is of tremendous importance since industrial products/projects often have hundreds, or even thousands of requirements. At the same time, scalability is one of the most commonly discussed problems and several studies have highlighted and discussed scalability problems [3]. When measuring perceived scalability, the subject can be asked for how well he or she thinks that the approach scales (e.g. [1]). Actual scalability, on the other hand, could be measured by letting subjects perform prioritizations with different number of requirements, and then measure how this affects the variables of interest, such as time consumption (e.g. [19]).

4.2.7 Understandability of Results

When making decisions based on the results of a prioritization effort, it is of course very important that the results are easy to understand. If they are not, wrong decisions could easily be made. Hence, it is important to measure how easy it is to understand the results from a prioritization. Measures of understandability could for example regard results presented on different scales, different visualization approaches (e.g. graphs), etc. When measuring perceived understandability, the subject can answer a question about how easy it is to understand/interpret the results (e.g. [14]). When measuring actual understandability, on the other hand, the results from different approaches could be compared (preferably through a blind test) to determine which approach that presents the result in the most understandable way. However, no such measures have been seen in the area so far.

4.2.8 Attractiveness

Although accuracy, scalability, etc. are measured, it does not matter how well the approach behaves if it is really boring or complex (even though this variable and other variables probably are correlated positively most of the time). Hence, it is also important to measure the attractiveness of an approach. This is preferably measured by simply asking the subjects how they liked the approaches of interest (e.g. [14]).

4.3 Context Variables

As indicated earlier in this paper, the result of a study can depend on many different contextual factors and it is seldom possible to generalize findings from one isolated study. In order to know when an approach is suitable or not, and being able to generalize results, it is hence important to monitor under which circumstances an approach are evaluated. This section presents a number of variables that are considered as important to monitor with regards to the context of the study. These kinds of variables are sometimes also referred to as intervening or mediating variables [7].

4.3.1 Environment

Especially in studies conducted in industry, basic information about the organizational characteristics is needed to understand the practical context and constraints in which prioritization is performed. The following variables are recommended to report:

Type of market: Market situation refers to how the organization's market look like, i.e. if it is a bespoke or market-driven situation and if the customers are professionals or not. Since the suitability of an approach may differ depending on market situation, and because a market situation can look in many different ways [3], it is important to explain the situation.

Process model: The process model an organization uses may influence the suitability of a prioritization approach. For example, prioritizing requirements upfront in a waterfall situation where the full scope of the project shall be captured is very different from prioritizing in an agile environment where several increments shall be planned and where change is considered as part of the process. Since implementation of different processes varies greatly, it is important to not only mention which process that is used but also describe how it is implemented in the organization.

Phase of prioritization: The phase of the product development in which the requirements are prioritized may influence suitability since the purpose and scope of the prioritization differ in different phases. Hence, the phase where the prioritization is performed should be reported.

Size of the project and organization: Information about project and organization size is useful to determine if the approach is most suitable in large or small settings. Reporting the number of persons involved, size from a technical point of view (e.g. LoC, complexity), etc. is recommended.

Application domain: Since different application domains (e.g. game industry, nuclear systems) require a different amount of control, stability, etc. and have to follow different rules and regulations, reporting about application domain may be important in order to know when an approach is suitable.

4.3.2 Study Setup

When reporting from a research study, it is important to give information about the study setup. This information is valuable both in terms of how to interpret the results and also to get additional knowledge about how things were performed, especially if the study should be possible to replicate. The following variables are recommended to be reported.

Prioritization Tools: The tools used to prioritize the requirements during the study are important to describe since tool usage may affect the results (see e.g. [20]). If using commercial tools, a reference to the tool should be given, together with possible configurations. If using a tool that is not publicly available, or a tool that is developed with the study in mind, a description of the tool should be reported. It may also be interesting to report about other tools used within an organization (e.g. requirements management tools) but such reporting is yet considered as optional. If no tools are used at all, it is of course still important to describe how the prioritization was conducted (e.g. [19]).

Work mode: When prioritizing requirements, some approaches may be better to use in a team-based prioritization while others are more useful in individual prioritization. To be able to get knowledge of what approaches that are useful for different ways of prioritizing, it is important to report how the prioritization was conducted. When reporting about team-based approaches, it is of course important to state how many were involved, and how the assignments of priorities were decided in the group.

Location/Amount of Control: When performing studies, the location of where the subjects performed the prioritization can affect the results greatly. For example, if doing the prioritization in their ordinary work places, they might answer phone calls, get e-mails, etc. that may affect time consumption, concentration, etc. Such disturbances may of course not only be bad (since this is the normal situation) but it could give different results than if doing it in isolation. Because of this, it is important that researchers report the environment where the prioritization was made, to make it possible to get an understanding about differences between different studies.

Duration of Study: In contrast to time (see Section 4.2.1), duration does not only regard the actual time for prioritization, but includes the time for the whole study. For example, if the subjects do several consecutive prioritizations, the subjects (and hence the results) may be affected.

Selection Strategy for Prioritization Approach: The subjects may perform differently if they get to choose/invent the prioritization approach by themselves, or if a boss or researcher has chosen it. The motivation is probably affected by the choice and it should hence be reported.

Role of the Researcher: The degree of involvement by the researcher may affect the results of a study since the amount of steering, information given, trust, etc. will affect the subjects' possibilities to perform the prioritization. For example, if using students as subjects, the subjects may be affected if their tutor (researcher) arranges the study since they have a special relationship to him/her.

4.3.3 Subjects

Information about the subjects involved in a study is important to report because their background, demographics, commitment, etc. probably will affect the results in different ways. As with study setup, this information can be very valuable when interpreting the results and when understanding how the study was performed. The following variables are recommended to be reported.

Roles/Perspectives: Besides just presenting the number of subjects, it is good to report what roles/perspectives the ones that prioritize have since some approaches may be more suitable for certain kinds of roles/perspectives. Previous experience on prioritizations performed in industry show that subjects with different roles/perspectives prioritize differently, and it is likely that they also favor different approaches.

Commitment: When conducting studies on requirements prioritization, it has been shown that commitment is one of the major factors that determine if the subjects are suitable or not [2]. In Höst *et al.* a classification scheme for degree of commitment is proposed, where four different classes are presented [11]. In addition to such a classification, it would be good to also express the commitment to the product as such, for example if it is a toy system (that may be developed in a project), a product ordered by a customer, or a product that the subjects themselves care about. In addition to this kind of commitment, the suitability can also be influenced by how the subjects were recruited or what rewards they get through being part of the prioritization. However, this kind of commitment is seen as of secondary importance, and is hence considered as optional.

Experience: Experience is another variable related to the suitability of subjects [11]. Höst *et al.* present a classification scheme for subjects with regards to their experience based on education and recent and relevant industrial experience [11]. To make this classification scheme more fine-grained, it is possible to ask for education and recent and relevant industrial experience in two separate questions. Besides only asking for recent and relevant industrial experience, it is also preferable to ask for experience with the approaches included in the study.

View on Software Development: One variable that may affect the results with regards to what the subjects think about an approach is the view they currently have on software development, i.e. if they favor agile or plan-driven processes. By measuring this variable, it is possible to get a feeling for if some approaches are more suitable with a certain process model.

Gender and Age: Gender, age, or similar demographics are possible to measure in order to get an understanding of to what extent an approach suits a particular type of person better.

4.3.4 Requirements

Information about the requirements to prioritize is important to report since different approaches may be suitable for different types of requirements. Below follows a number of characteristics of requirements that are recommended to be monitored. It should be noted that it may be a good idea to present examples of requirements included in the study to make it possible to ensure similar interpretation of type, abstraction level, and structure of the requirements.

Number of requirements: The number of requirements prioritized greatly influences the results regarding a prioritization approach. For example, an approach may be seen as easy to use, gives good overview, etc. with a few requirements but as the number of requirements increases, the usefulness decreases. This makes it very important to report the number of requirements, especially from scalability (Section 4.2.6) point of view.

Type of requirements: There exist several different types of requirements, e.g. functional requirements, non-functional requirements, and constraints [31]. The suitability of an approach may be influenced by which type of requirements that are prioritized. The reason for differences in suitability may be due to, for example, extent and kind of dependencies between different types of requirements.

Abstraction level: The abstraction level of requirements may also influence suitability to a large extent since, for example, the extent and kind of dependencies, and the number of requirements may differ greatly between high level (e.g. features) and low level (e.g. deep technical) requirements. Since there do not exist any commonly accepted standards or guidelines about abstraction levels yet (see e.g. [4] for differences), reporting about abstraction levels are seen as optional.

Structure: Requirements structures are experiences in many different ways, e.g. amount of overview, dependencies, and complexity. These structures may very well influence suitability very much but as with abstraction level, there exist no rules or guidelines yet in relation to structure. Therefore, it is considered as optional to report on such structures.

5. FULFILLMENT OF FRAMEWORK

As indicated in Section 2.2, one of the main reasons of constructing the research framework was due to make different researchers report information about the same variables. In order to evaluate how well the current studies fulfill the proposed framework, an attempt has been made to classify the papers from the systematic review (see Section 2.2) according to the proposed framework. It should be noted that the systematic review as such is presented elsewhere (see [22]) and that the presentation below mainly serves as an evaluation of how well existing studies fulfill the framework.

Based on the papers from the systematic review as input, an evaluation has been performed to investigate how they fulfill the suggested framework. The initial idea was to present a table where the fulfillment of each variable could be presented. However, when constructing such a table, it was evident that the presented information in the different studies was too diverse to quantitatively analyze the fulfillment. Instead, this section presents a qualitative evaluation of how the papers in general measure and report the different variables presented in the framework. The papers that were included in the systematic review are presented in Table 2 and the analysis is based on the results from the systematic review.

Table 2 Papers included in the systematic review

Study	Techniques
[1]	Analytical Hierarchy Process (AHP), Binary Search Tree (BST), Planning Game (PG), $100-test (Cumulative Voting), PG&AHP
[13]	AHP
[14]	Numeral Assignment, AHP
[15]	AHP
[16]	AHP
[18]	AHP(*2), Minimal Spanning Tree, Bubble Sort, BST, Priority Groups
[19]	PG, AHP
[30]	$100-test

As can be seen in Table 2, AHP is the technique that has received most attention in the past. However, it should be noted that there is an overlap of studies in [14], [15], and [16]. Paper [18], on the other hand, presents two studies but only the main study is included in the evaluation here since the second study was given too little attention in the paper. A similar situation is present in

[17] where a study of tool-supported-AHP is mentioned very briefly and hence is not included in the evaluation. Further, an extension of the study presented in [19] is presented in [20]. However, the study presented in [20] was not published at the time when the systematic review was conducted. It should also be noted that several of the papers presented in Table 2 are almost ten years old and it can not be expected that they are as well planned, conducted, and reported as more recent papers (due to maturation in the field). This maturation effect is visible in the papers which is a good sign that the field are getting closer to a more mature discipline; year by year and study by study.

In Section 5.1, the fulfillment of the independent variables is discussed while Sections 5.2 and 5.3 present the fulfillment for dependent and context variables respectively.

5.1 Independent Variables

In general, the papers fulfill the requirements of the framework quite well when it comes to the independent variables. In most of the papers, information about "qualitative process description", "goal", "input", and "output" are provided. Since hierarchical levels have not been introduced before, no paper explicitly identifies which level the technique they discuss about belongs to. However, the papers often describe the techniques well enough to identify the level. Based on this analysis, it seems that papers fulfill the requirements well already even though it may become more explicit if following the suggested framework.

5.2 Dependent Variables

When it comes to the dependent variables, there is a rather big difference between different papers on what variables they report and how they report on these variables. One reason for why different variables are reported is of course that the different studies investigate different issues. However, if being aware of what other variables to measure, many of the studies could easily collect and present the results of those variables as well. There are big differences between different variables on how often they are collected and reported. Understandability, for example, is measured in only one study (qualitatively) while for example time is measured in most of the studies.

Nevertheless, if looking at the variable "time", it is also evident that the way of reporting about this variable differs very much between different studies. For example, some studies measure time objectively (by using a stop watch), others report on how much time the subjects perceived it to take, and still others present time consumption on an ordinal scale in relation to the other techniques evaluated.

Another observation with regard to the dependent variables is that the same or similar variables are presented under different names. Accuracy, for example, has been investigated using accuracy, 'gut feeling', certainty, and reliability.

The above analysis of the reporting shows that there is a need for a research framework that makes the reporting and analysis of dependent variables more coherent in the field.

5.3 Context Variables

Context variables are important to get an understanding of the suitability of the approach in different situations. Nevertheless, when looking at the reporting of context variables in the papers analyzed, it seems that this is the kind of variables that are most often omitted. This does not mean that all the investigated papers are bad on reporting on all context variables. Instead, it seems that some papers focus on reporting on some variables and others on some other variables.

If looking at the environmental variables (primarily interesting for industrial studies) the application domain is commonly reported while the others often lack description. One reason may be due to company regulations but the information of interest is not of critical nature for a company (which can be seen in some papers that gives very good descriptions). Instead, the reason for why not reporting extensively on environmental variables may rather be because of not knowing about what is interesting to present.

When it comes to study setup variables, the information is also commonly a little bit thin. Most papers explain very well on some variables but the problem is that different authors seem to favor different variables. This means that some report on tools used, some on work mode, and yet others on location. The situation with subjects is similar even though it is slightly better. Most papers report on the number of subjects and also try to explain what roles/perspectives the subjects have. The variables concerning commitment and experience, on the other hand, are commonly weakly reported in the papers even though it has been shown that this is important to measure. The other context variables are also commonly left out or presented sparingly.

The description of the requirements prioritized in the study is also very different between different papers. Some papers report very well while others barely mention the number of requirements. One good practice that can be seen in some papers is that they present the requirements used (or at least some descriptive examples). This is very good since there does not exist unambiguous agreements about abstraction levels and structure of requirements, which would make it possible for others to see and value the requirements.

As can be seen in this section about context variables, there is a great need for studies to report on the same variables. This is needed to build up a knowledge base about when to use certain approaches. If not reporting on the same context variables, it is impossible to determine when one approach is to favor over another.

6. DISCUSSION AND CONCLUSIONS

Requirements prioritization is recognized as a challenging decision-making activity that requires support. Many approaches for prioritization of software requirements are introduced in the literature. Despite of several empirical studies, there is still a lack of evidence of which prioritization approaches that are to prefer, since different studies have resulted in different conclusions. One reason for this may be that studies are done in different contexts, measure and report different variables, and use different data sets.

By conducting a systematic review, the authors of this paper recognized that there is a need for a research framework that could guide researchers on what to measure and report when conducting studies within the area. Based on this need, such a framework is proposed based on the result of the systematic review. With the systematic review as main input together with input from papers presenting similar frameworks in other areas, a two-day workshop was held where the variables recognized were discussed and the content of the framework was decided.

The framework proposed in this paper should, however, be seen as a first step to towards a more standardized framework. Similarly to similar frameworks (e.g. [12]), the framework needs to be further refined based on experience gained from evaluation (see e.g. evaluation of [12] in [24]) and usage of the framework. It should also be noted that following and reporting the issues given in this framework does not replace the need of following and reporting on good practices regarding research methodology issues. This means that it is important to report about what kind of study was performed, how validity threats were handled, how the study was designed, etc. For example, it is advised that the maturing guidelines presented in [12] are followed in parallel when reporting requirements prioritization experiments. At the same time, the framework can and should be used as input when designing studies, i.e. a study should be designed with the framework in mind and enable collection of the proposed variables. Even though the framework may seem extensive at a first glance, most variables are easy to report and do not require any additional effort by the subjects.

Threats to validity regarding this research are mostly due to the fact that the presented framework has not been evaluated in practice yet. To get the framework really useful, researchers must report according to the framework and hence validate that it is correct (and if not, improve it further). Additionally, the fact that three researchers constructed the framework may imply a threat that the framework represents only their view and opinions. However, the framework originated from a systematic review in the requirements prioritization area and from other similar frameworks in other areas. Furthermore, the three authors come from different environments and have different backgrounds, which mean that they account for different perspectives. In addition, two of the authors have conducted an extensive amount of research in the area and were able to use their experience concerning the validation of variables when constructing the framework. At the same time, the third researcher was able to take an outsider's perspective to not make the framework colored only by experience from the prioritization area. Based on these facts, the framework can be regarded as a good starting point when it comes to what and how to design and report results in the area. However, the need for empirical evaluation and further refinements of the framework is essential.

7. ACKNOWLEDGMENTS

We would like to thank Lena Karlsson for many interesting discussions and inputs in the early parts of this work. We would also like to thank Per Jönsson for valuable comments when reviewing the paper in later stages.

This work was partly funded by the Knowledge Foundation in Sweden under a research grant for the project "Blekinge — Engineering Software Qualities (BESQ)" (http://www.bth.se/besq) and by the National Technology Agency of Finland (TEKES) under a research grant for the project CORE (Competitive Advantage through Stakeholder-Driven Requirements Engineering).

8. REFERENCES

[1] Ahl, A. (2005) *An Experimental Comparison of Five Prioritization Techniques - Investigating Ease of Use, Accuracy, and Scalability*, Master Thesis No. MSE-2005-11, School of Engineering, Blekinge Institute of Technology.

[2] Berander, P. (2004) 'Using Students as Subjects in Requirements Prioritization', *Proceedings of the 2004 International Symposium on Empirical Software Engineering (ISESE'04)*, Redondo Beach, CA, pp. 167-176.

[3] Berander, P. and Andrews, A. (2005) 'Requirements Prioritization', in *Engineering and Managing Software Requirements*, ed. Aurum, A., and Wohlin, C., Springer Verlag, Berlin, Germany, pp. 69-94.

[4] Berander, P., and Jönsson, P. (2006) 'Hierarchical Cumulative Voting (HCV) - Prioritization of Requirements in Hierarchies', accepted for publication in *International Journal of Software Engineering and Knowledge Engineering (IJSEKE) - Special Issue on Requirements Engineering Decision Support*, December 2006, World Scientific Publishing Company, Singapore.

[5] Biolchini, J., Mian, P. G., Natali, A. C., Cruz, T. and Guilherme, H. (2005) *Systematic Review In Software Engineering*, System Engineering and Computer Science Department COPPE/ UFRJ, Technical Report ES 679/05.

[6] Carlshamre, P. (2002) 'Release Planning in Market-driven Software Product Development - Provoking an Understanding', *Requirements Engineering Journal* **7**(3), pp. 139-151.

[7] Creswell, J. W. (2003) *Research Design - Qualitative, Quantitative, and Mixed Methods Approaches*, 2nd Edition, Sage Publications, Thousand Oaks, CA.

[8] Dybå, T., Kampenes, V. B., and Sjøberg, D. (2005) 'A Systematic Review of Statistical Power in Software Engineering Experiments', Accepted for publication in *Journal of Information & Software Technology*.

[9] Gallis, H., Arisholm, E., and Dybå, T. (2003) 'An Initial Framework for Research on Pair Programming', *Proceedings of the 2003 International Symposium on Empirical Software Engineering (ISESE'03)*, Rome, Italy, pp. 132-142.

[10] Greer, D., and Ruhe, G. (2004) 'Software Release Planning: an Evolutionary and Iterative Approach', *Information and Software Technology*, **46**(4), pp. 243-253.

[11] Höst, M., Wohlin, C., and Thelin, T. (2005) 'Experimental Context Classification: Incentives and Experience of Subjects', *Proceedings of the 27th* International Conference on Software Engineering (ICSE 2005), St Louis, MO, pp. 470-478.

[12] Jedlitschka, A., and Pfahl, D. (2005) *Reporting Guidelines for Controlled Experiments in Software Engineering*, Report IESE-035.5/E.

[13] Karatzas, K., Dioudi, E., and Moussiopoulos (2003) 'Identification of Major Components for Integrated Urban Air Quality Management and Information Systems via User Requirements Prioritisation', *Environmental Modelling and Software*, **18**(2), pp. 173-178.

[14] Karlsson, J. (1996) 'Software Requirements Prioritizing', *Proceedings of the 2nd* International Conference on Requirements Engineering (ICRE'96), Colorado Springs, CO, pp. 110-116.

[15] Karlsson, J., and Ryan, K. (1996) 'Supporting Selection of Software Requirements', *Proceedings of the 8th International*

Workshop on Software Specification and Design (IWSSD'96), Schloss Velen, Germany, pp146-149.

[16] Karlsson, J., and Ryan, K. (1997) 'A Cost-Value Approach for Prioritizing Requirements', *IEEE Software* **14**(5), pp. 67-74.

[17] Karlsson, J., Olsson, S., and Ryan, K. (1997) 'Improved Practical Support for Large-Scale Requirements Prioritizing', *Requirements Engineering Journal* 2(1), pp. 51-60.

[18] Karlsson, J., Wohlin, C., and Regnell, B. (1998) 'An Evaluation of Methods for Prioritizing Software Requirements', *Information and Software Technology* **39**(14-15), pp. 939-947.

[19] Karlsson, L., Berander, P., Regnell, B., and Wohlin, C. (2004) 'Requirements Prioritisation: An Experiment on Exhaustive Pair-Wise Comparisons versus Planning Game Partitioning', *Proceedings of the 8*th International Conference on Empirical Assessment in Software Engineering (EASE 2004), Edinburgh, Scotland, pp. 145-154.

[20] Karlsson, L., Thelin, T., Regnell, B., Berander, P., and Wohlin, C. (2006) 'Pair-Wise Comparisons versus Planning Game Partitioning – Experiments on Requirements Prioritisation Techniques', *Empirical Software Engineering (EMSE)*, available from Internet <http://dx.doi.org/10.1007/s10664-006-7240-4> (14 June 2006).

[21] Khan, K. S., Ter, R. G., Glanville, J., Sowden, A. J., and Kleijnen, Jo. (2001) *Undertaking Systematic Review of Research on Effectiveness. CRD's Guidance for those Carrying Out or Commissioning Reviews*, CRD Report Number 4 (2nd Edition), NHS Centre for Reviews and Dissemination, University of York, ISBN 1 900640 20 1.

[22] Khan, K. A. (2006) *Systematic Review of Requirements Prioritization – A Research Framework*, Master Thesis, Blekinge Institute of Technology, to be presented and defended 2006-10-23.

[23] Kitchenham, B. A. (2004) *Procedures for performing systematic reviews*, Joint Technical Report: Keele University Technical report TR/SE-0401, ISSN: 1353-7776, and NICTA Technical Report 0400011T.1.

[24] Kitchenham, B., Al-Khilidar, H., Ali Babar, M., Berry, M., Cox, K., Keung, J., Kurniawati, F., Staples, M., Zhang, H., and Zhu, Liming (2006) 'Evaluating Guidelines for Empirical Software Engineering Studies', *Proceedings of the 5th ACM-IEEE International Symposium on Empirical Software Engineering (ISESE'06)*, Rio de Janeiro, Brazil, pp. 38-47.

[25] Lehtola, L., and Kauppinen, M. (2004) 'Empirical Evaluation of Two Requirements Prioritization Methods in Product Development Projects', *Proceedings of the European Software Process Improvement Conference (EuroSPI 2004)*, Trondheim, Norway, pp. 161-170.

[26] Lehtola, L., Kauppinen, M., and Kujala, S. (2004) 'Requirements Prioritization Challenges in Practice', *Proceedings of* 5th International Conference on Product Focused Software Process Improvement, Kansai Science City, Japan, pp. 497-508.

[27] Lubars, M., Potts, C., and Richter, C. (1993) 'A Review of the State of the Practice in Requirements Modelling', *Proceedings of First IEEE Symposium on Requirements Engineering (RE´93)*, San Diego, California, USA, pp. 2-14.

[28] Mendes, E. A. (2005) 'Systematic Review of Web Engineering Research', Proceedings of the 2005 International Symposium on Empirical Software Engineering (ISESE 2005), Noosa Heads, Australia, pp. 481-490.

[29] Ngo-The, A., and Ruhe, G. (2005) 'Decision Support in Requirements Engineering', in *Engineering and Managing Software Requirements*, ed. Aurum, A., and Wohlin, C., Springer Verlag, Berlin, Germany, pp. 267-286.

[30] Regnell, B., Höst, M., Natt och Dag, J., Beremark, P., and Hjelm, T. (2001) 'An Industrial Case Study on Distributed Prioritisation in Market-Driven Requirements Engineering for Packaged Software', *Requirements Engineering*, **6**(1), pp. 51-62.

[31] Robertson, S., and Robertson, J. (1999) *Mastering the Requirements Process*, Addison-Wesley, Harlow, England.

[32] Ruhe, G. (2002) 'Software Engineering Decision Support', *Proceedings of the 4th International Workshop of Advances in Learning Software Organizations*, Chicago, IL, pp. 104-113.

[33] Ruhe, G., Eberlein, A., and Pfahl, D. (2002) 'Quantitative WinWin – A New Method for Decision Support in Requirements Negotiation', *Proceedings of the 14th International Conference on Software Engineering and Knowledge Engineering (SEKE'02)*, Ichia, Italy, pp. 159-166.

[34] Runeson, R., Ohlsson, M. C., and Wohlin, C. (2001) 'A Classification Scheme for Studies on Fault-Prone Components', *Proceedings of the International Conference on Product Focused Software Process Improvement (PROFES01)*, Kaiserslautern, Germany, pp. 341-355.

[35] Saaty, T. L. (1980) *The Analytic Hierarchy Process*, McGraw-Hill, New York.

[36] Saaty, T. L., and Vargas, L. G. (2001) *Models, Methods, Concepts & Applications of the Analytic Hierarchy Process*, Kluwer Academic Publishers, Norwell.

[37] Siddiqi, J., and Shekaran, M. (1996) 'Requirements Engineering: The Emerging Wisdom', *IEEE Software* **13**(2), pp. 15-19.

[38] Sommerville I. (1996) *Software Engineering*, 5th Edition Addison-Wesley, Wokingham, England.

[39] Vegas, S., Juristo, N., Moreno, A., Solari, M., and Letelier, P. (2006) 'Analysis of the Influence of Communication between Researchers on Experiment Replication', *Proceedings of the 5th ACM-IEEE International Symposium on Empirical Software Engineering (ISESE'06)*, Rio de Janeiro, Brazil, pp. 28-37.

[40] Wiegers, K. (2003) *Software Requirements*, 2nd Edition, Microsoft Press, Redmond, WA.