An Empirical Study on Views of Importance of Change Impact Analysis Issues

Per Jönsson, Lefteris Angelis, Claes Wohlin

Abstract— **Change impact analysis is a change management activity that previously has been studied much from a technical perspective. For example, much work focuses on methods for determining the impact of a change. In this paper, we present results from a study on the role of impact analysis in the change management process. In the study, impact analysis issues were prioritised with respect to criticality by software professionals from an organisational perspective and a self-perspective. The software professionals belonged to three organisational levels: operative, tactical and strategic. Qualitative and statistical analyses with respect to differences between perspectives as well as levels are presented. The results show that important issues for a particular level are tightly related to how the level is defined. Similarly, issues important from an organisational perspective are more holistic than those important from a self-perspective. However, our data indicate that the self-perspective colours the organisational perspective, meaning that personal opinions and attitudes cannot easily be disregarded. In comparing the perspectives and the levels, we visualise the differences in a way that allow us to discuss two classes of issues: high-priority and medium-priority. The most important issues from this point of view concern fundamental aspects of impact analysis and its execution.**

Index Terms— **Change impact analysis, organisational level, perspective, prioritisation, issues.**

I. Introduction

CHANGE impact analysis is, according to Bohner and Arnold, the activity of "identifying the potential consequences of a change, or estimating what needs to be modified to accomplish a change." [1] It is a crucial software development activity due to the often massive amount of changes a software system is exposed to during its life cycle. Changes can stem from change requests, new requirements, defects, etc. New requirements, and in particular change requests, undergo analysis governed by certain acceptance/inclusion criteria. The focus of this type of analysis is typically on the *consequences* of accepting (or rejecting) the requirement or change. When a proposed change, regardless of source, is going to be inserted into the software system, an analysis of the necessary work should be made. The focus of this type of analysis is necessarily on the *modifications* required by the change.

The gross of impact analysis (IA) research concerns the development of methods and algorithms for supporting and automating the analysis, or the adaptation of existing methods in new contexts. To our knowledge, there is little research about more non-technical aspects of the subject, such as process and organisational aspects. Since IA is a part of the change control process, its success is, in our experience, heavily dependent on organisational support and stakeholder views.

In this paper, we present an explorative, empirical study of the views of IA at Ericsson AB in Sweden. Ericsson develops, among other things, software systems for mobile telecommunications. These systems are by nature both large and complex, and are subject to rapid evolution. In the study, the change source in question was change requests, and IA was seen in a broad sense: both with respect to the analysis of the change requests, and the analysis of the modification induced by the change.

In exploring issues of IA at Ericsson, we identified three *organisational levels* of people with different foci: technical, resource, and product. Looking at management science, we found that these levels mapped well to the decision-making model defined by Anthony (studied previously in a software engineering context in, for example, [2] and [3]), where decisions are categorised as *operative*, *tactical* or *strategic* [4]. We also identified two *perspectives* of views, namely *self* (individual) and *organisational*. The idea is that the organisational perspective aligns with the overall goals of the organisation, whereas the self-perspective aligns with the goals of the individual.

In order to understand how issues associated with IA are seen at the different levels and under the different perspectives, we interviewed 18 employees at the company in their roles as industrial experts. The resulting list of issues was subsequently subjected to prioritisation by the interviewees. Our goal is to investigate whether people see IA differently depending on their level and perspective, and whether these groupings are meaningful in process improvement efforts. This paper extends the analysis of a previous publication that discussed only the organisational levels [5]. The current paper also looks at the aforementioned perspectives (preliminary results of which have been published at a national conference [6]), and discusses the two grouping types in relation to each other. The reason for extending the previous papers is to provide a more in-depth analysis as well as a wider take on the data (level and perspectives). We present a qualitative analysis as well as a detailed statistical analysis of some of the results, and discuss possible improvements pertaining to issues found to be important.

The main contribution of the paper is threefold. First, it presents empirical data from a industrial study on an organisational aspect of impact analysis issues. Second, it provides a rigorous analysis of the data, and a visualisation that is helpful in highlighting critical issues. Third, it proposes a set of process improvements that can help mitigate the critical issues.

The paper is structured as follows. The following section describes the context of the research and presents the research questions for this work. Thereafter, Section III discusses related work. Next, the research method is described in Section IV. The results obtained in the study are presented in Section V, followed

P. Jönsson was at the time of the study with the Blekinge Institute of Technology, Ronneby, Sweden. He is currently with Ericsson AB, Karlskrona, Sweden. E-mail: per.jonsson@ericsson.com.

L. Angelis is with the Aristotle University, Thessaloniki, Greece. E-mail: lef@csd.auth.gr.

C. Wohlin is with the Blekinge Institute of Technology, Ronneby, Sweden. E-mail: claes.wohlin@bth.se.

by analysis and discussion in Section VI. Finally, conclusions and future work are discussed in Section VII.

II. Research Context

Ericsson is a company that operates on a world market, and as such has several large international customers. The population we wish to generalise to is industrial/professional software practitioners in general rather than just within the company. We believe this to be possible due to the fact that the company deals with large-scale software development and is ISO 9000 certified (which means that processes and ways of working are governed by predefined guidelines). It can be assumed that the challenges and characteristics of the development work exist in other companies of the same magnitude as well, and to some extent also in smaller companies.

The research questions for this work are:

- RQ1 How does the organisational level affect one's assessment of importance of IA issues?
- RQ2 What difference does the perspective make in determining the relative importance of IA issues?
- RQ3 Which are important IA issues, and how are these addressed in software process improvement?

Answers to the first two research questions are sought by looking at differences between groups with respect to how people in the groups have judged the relative importance of the issues. For RQ1, the three groups correspond to the three organisational levels. For RQ2, we look at two groups defined by the perspective used when judging the importance of issues. For these two research questions, both a qualitative analysis and a statistical analysis are performed. In addition, the two groupings are studied together. RQ3 is approached in a more holistic manner, with the purpose of weighing together importance from the different levels and perspectives.

The overarching goal of the conducted research is to explore the role of IA in the organisation, in order to be able to improve ways of working with IA. By studying and learning about differences in views of IA, process improvement efforts can be directed towards a certain target group. For example, the requirements on process support tools could differ between different target groups because the groups consider different issues to be of importance.

III. Related Work

Aurum and Wohlin tie requirements engineering (RE) activities to decision-making models, arguing that RE is a decision-intensive process [2]. They suggest that studying decision making within RE helps organisations structure their RE decisions better and, ultimately, produce software with higher quality. We propose that the same argument holds for IA, due to the strong connection between IA and RE.

Several researchers report on differences between managers and engineers in the context of software process improvement (SPI), for example concerning views of software quality [7], use of formal routines to transfer knowledge and experience [8], and how they rate factors affecting an SPI goal [9]. Even though both managers and engineers have personal agendas, our experience is that managers often have a more holistic view of the development work than engineers. Due to this, the personal priorities of managers are likely to align well with those of the organisation, whereas the personal priorities of engineers should be narrower. Thus, differences between managers and engineers could be attributed to differences between the organisational alignment of their default perspectives.

The notion of perspectives can be found in other research, for example on perspective-based reading (PBR). PBR is a scenario-based reading technique introduced by Basili et al., with the purpose of maximising the number of faults found when reading a document by reading it from several different perspectives [10]. The idea is that each perspective finds its own unique faults, and that the union of the perspectives provides a good coverage of the document being read. PBR is a good example of how to take advantage of differences between perspectives, although there is research that indicates that the perspectives need not differ with statistical significance [11].

Finkelstein and Sommerville discuss perspective in relation to viewpoints [12]. A viewpoint consists of an agent (i.e., actor) and a perspective. Large-scale software development requires, due to its complexity, multiple viewpoints and thus multiple perspectives. According to Finkelstein and Sommerville, differences in perspectives stem from the agents' different responsibilities and roles and thus variations in their goals and opinions. Furthermore, it is argued that multiple perspectives need to be integrated if there is a common or shared goal [12]. In SPI, the common goal is to create a successful improvement programme that addresses the the organisation's software processes (see, for example, [13]). This motivates the relevance of looking at multiple perspectives when prioritising impact analysis issues.

Karlsson and Ryan describe a requirements prioritisation approach, in which requirements are prioritised from two different perspectives—a value (customer) perspective and a cost (developer) perspective [14]. Their visualisation of the prioritisations has inspired the visualisation approach used in this paper. A difference between our perspective comparison and their perspective comparison is that in their case there is a preferable situation, i.e. that the value/cost ratio for a requirement should be as high as possible. In our case, a high ratio between the two perspectives (organisational and self) is not "better" than a low ratio. For example, both a high and a low ratio point to a difference between the perspectives, whereas a ratio close to one points to a similarity between the perspectives.

IV. Method

In this section, we describe the research method used in the study.

The study consisted of three main steps: interviews, post-test, and workshop. In the interview step, 18 persons involved in the change management process were asked questions about impact analysis and change management. The roles of the participants ranged from programmers to product managers, but each one had in one way or another a relation to the impact analysis activity, as decision-maker, technical advisor, change request writer, implementer, and so on.

The sampling of the participants can be characterised as *convenience sampling* [15], as no specific system was used in selecting people to interview. Instead, people were selected mainly based on recommendation (and availability), although some were approached at random based on their role. As one of the initial objectives of the study was to look at organisational levels (see Section IV-A), the sampling of participants were largely controlled by the extent to which the levels were thought to be represented.

In the post-test, the participants gave input to the assessment of their organisational levels, and prioritised issues elicited during the interviews. We anticipated that each participant, when asked during the interview, would only mention a limited set of issues, and thus not see (or be aware of) the whole picture. By making the prioritisation a separate step rather than a part of the interviews, all issues could be included in the prioritisation.

The purpose of the workshop was to let the participants discuss the outcome of the prioritisation, and possibly come up with ideas how to address important issues. Furthermore, as several roles were represented by the participants, the workshop could allow people to gain greater understanding of each other's views.

The following sections describe the steps of the study in more detail.

A. Organisational Levels

As mentioned earlier, the three organisational levels we identified map well to the decision-making model defined by Anthony. The model differentiates between decisions at three levels as follows [4]:

- *Strategic* decisions have typically large scope, large impact and long-term perspective. They concern organisational or product-related goals and strategies.
- *Tactical* decisions concern planning of time and resources to reach strategic goals, and are often made by middle management. They have smaller scope and impact, and shorter time horizon, than strategic decisions.
- *Operative* decisions are made when realising the project according to the plan, and are often of technical nature.

In terms of impact analysis and handling of change requests, people at the operative level can be expected to do technical analysis of system impact, and also manage the actual realisation (implementation) of the change when it has been accepted. People on the tactical level can be expected to analyse based on resource and time consumption, and also do planning and coordination of change realisation. Finally, people on the strategic level can be expected to look at how a proposed change aligns with long-term goals, and what the product- or organisation-wide consequences of a change are.

B. Interview Design

The interview instrument contained seven main topics, each of which was associated with one or more open questions. The instrument in its entirety can be obtained from the authors by request. This paper is only concerned with the topic of potential issues of impact analysis and change handling. The question for this topic was (translated from Swedish): "*Which potential issues are associated with performing impact analysis*?"

Note that we asked about potential issues rather than actual ones. The reason for this was to avoid limiting the ability to generalise the results by extracting company-specific issues only.

The remaining topics were more qualitative in their nature, and were intended both for providing a context for and for collecting hidden or implicit knowledge about the issues. We did not intend for the participants to prioritise during the interviews, as we expected each of them to see only a subset of the possible issues.

In order to ensure the appropriateness and clarity of the questions, the interview instrument was developed in close cooperation with the company. Furthermore, we performed a pilot interview in order to find inconsistencies and problems with the interview instrument. The pilot interview resulted in only minor changes to the interview instrument, which lead us to include the interview in the study.

The interviews were *semi-structured*, meaning that it was not necessary to follow the predefined question order strictly, and that the wording of questions was not seen as crucial for the outcome of the interviews (see [15]). The participants could speak rather freely, but we made sure that all questions were answered in one way or another.

The participants were asked if they would agree to join in a follow-up activity, where they should prioritise the issues elicited during the interviews. This was done in order to prepare the participants and increase their commitment towards the follow-up prioritisation.

A great variety of roles were covered in the interviews, including developer, tester, technical coordinator, manager (functional, product and project) and system architect. It should also be noted that the participants in general had been working at the company for a long time, and were thus familiar with processes and routines.

C. Results Triangulation and Filtering

A triangulation and filtering scheme was designed in order to get a complete list of issues. The scheme involved three information sources: (1) the list generated in the interviews based on the topic about issues, (2) qualitative information from the interviews, and (3) information from the literature. The analysis of the interview data was done manually by the researchers.

By using information from all interview topics (i.e., not only the topic about issues), it was possible to capture both explicit and implicit knowledge about issues. For example, it happened on several occasions that participants mentioned issues in passing while talking about processes and tools. These "implicitly" mentioned issues were extracted from the interview data and, as part of the triangulation, merged with the issues explicitly mentioned. Finally, by collecting information from the literature, we were able to add issues of which the participants were not aware.

The purpose of the filtering part was to remove redundancies and inconsistencies in the issue list by merging similar issues together, and by discarding issues that were not directly related to impact analysis.

D. Post-Test with Prioritisation

A post-test was designed as a follow-up to the interviews, for two purposes. The first purpose was to let the participants determine their respective organisational levels, by stating the distribution of their decisions on the decision levels (e.g., 20% decisions are made on the strategic level, 50% on the tactical level, and 30% on the operative level). Based on this, it was possible to deduce their organisational levels. For example, a participant making mostly strategic decisions should be regarded as belonging to the strategic organisational level. This scheme was used since, as Aurum and Wohlin point out, Anthony's decision levels are not entirely orthogonal [2].

The second purpose was to obtain a prioritised list of issues from each participant. The participants should prioritise such that

TABLE I
CUMULATIVE VOTING EXAMPLE.

Item	Priority
Item 1	10
Item 2	50
Item 3	20
Item 4	15
Item 5	5
	$\sum = 100$

the issue with the highest priority would be the one most critical to the organisation if it existed, thereby looking at the issues under the organisational perspective. However, we were also interested in knowing if the participants would prioritise differently under a self-perspective. Thus, we let each participant prioritise also from an individual point of view, assigning high priorities to issues critical to himself or herself if they existed.

To account for the problem that the first prioritisation of issues could affect the second (due to maturation), a two group design was used for the post-test, such that half of the participants should prioritise from the organisational perspective first, and the other half from the self-perspective first.

In the prioritisation, the participants should use the *cumulative voting* (CV) technique [16]. CV works as follows. For each item i being prioritised, the prioritising person assigns a weight, w_i, such that the sum of weights over k items, $\sum_{i=1}^{k} w_i$, equals a constant number, W. The ratio w_i/W corresponds to the importance of item i relative to all other items. Typically, the weight sum W is set to 100, as this facilitates both the assignment of weights and the perception of the relative importance of items. Furthermore, the process of assigning weights to the items is commonly referred to distributing *points* or *money* (which is why the technique is also called the *hundred-dollar test*). In this paper, item weights are referred to as item priorities.

To facilitate for the participants, an Excel-based prioritisation tool was created. The tool helped the participants keep track of the sum of weights, and also gave instant feedback by presenting the resulting list of prioritised issues after the prioritisation.

Advantages of CV are that it is easy to learn and use, and that the resulting priorities are on a ratio scale (i.e., it is meaningful to discuss the ratio between the priorities of two items). Table I gives an example of how the outcome of applying CV to a set of five items could look like. In the example, the relative priorities sum to 100, whereas in the study we used 1 000 due to a larger number of items (otherwise, the participants would have had to deal with very small numbers). It can be seen in the example that the person doing this prioritisation considers item 2 to correspond to 50% of the total importance of all items, being 10 times as important as item 5.

E. Workshop

As mentioned earlier, the workshop was meant to provide a forum where the participants could discuss both the results from the follow-up prioritisation and possible ways to mitigate the important/critical issues. A second purpose was to allow the participants to obtain a better understanding of each other's views.

As not all the participants could meet at one single occasion, two separate workshops were arranged instead of only one. Still, less than a third of the participants were able to take part in the workshops. The workshop results were compiled and sent to all participants afterwards, both to spread the results to those who did not attend, and to merge the results from the two separate workshops. Furthermore, the participants could in this way control that the compiled information was correct.

The set of issues and the prioritisation data were not altered in any way due to the discussions during the workshops. Thus, the low attendance does not pose a threat to the validity of the study.

V. RESULTS

The mapping of participants to the organisational levels (described in Section IV-A) resulted in eight participants at the operative level, five at the tactical level and five at the strategic level. In total, 25 relevant issues could be extracted from the interview material. When doing a thorough literature search, we found six issues relevant to our work (see, for example, [1]). However, these were already among the 25, which indicates that our 25 issues indeed are relevant and have good coverage. The six issues found in the literature are (mapping to issues presented here in parentheses):

- A defined process is missing (i25).
- Tools and automation are missing (i13).
- Manual impact analysis is time-consuming and difficult (e.g., i2, i9, i24).
- The available traceability is insufficient (i11).
- The documentation is not updated properly (i25).
- The available documentation is too coarse (i25).

All issues can be seen in Table II, labelled from i1 to i25. Note that the original issues were in Swedish and have been translated into English here. In the translation, we have strived to keep the main point of each issue.

The 25 issues vary in closeness to impact analysis. For example, some issues concern change requests, some concern the actual analysis activity, some concern process details, and others concern the analysis results. This is of course less than optimal from a prioritisation point of view. The alternative would however be to divide the issues into groups and prioritise the groups separately, which would complicate the prioritisation further.

To illustrate the distribution of priorities over issues, Fig. 1 shows the box plot for the resulting priorities for the operative level under the individual perspective. It can be seen that for many issues, the majority of the priorities are low, while only a few issues have higher peaks. The situation is similar for the other levels. This is a consequence of using CV with many items.

Fig. 2 displays bar charts showing the mean priorities for all issues (sorted in descending order of importance) under each of the two perspectives. Error bars showing plus one standard deviation are provided to again illustrate the distribution of priorities. An interesting feature of the bar chart for the self-perspective is that the issues with high mean priorities are fewer and more prominent than the remaining issues. This means that the participants allocated more points to a smaller set of important issues under the self-perspective than under the organisational perspective.

When it comes to investigating the agreement (or disagreement) among the different levels and perspectives, the original data have some characteristics that complicate visualisation and qualitative interpretation: (1) The CV technique augments disagreement between two prioritising persons. Numerically speaking, a disagreement on one issue necessarily leads to a disagreement also

TABLE II
ISSUES ELICITED IN THE INTERVIEWS.

id	Issue
i1	It is difficult to find resources for performing impact analysis.
i2	There is not enough time to perform impact analysis.
i3	System impact is underestimated or overlooked.
i4	Change requests are unclear.
i5	Responsibility and product/project balance are difficult to handle for analyses that span several systems.
i6	Analyses are incomplete or delayed.
i7	Analyses require much expertise and experience.
i8	Analyses are too coarse or uncertain.
i9	It is difficult to handle conflicting and synergetic change requests.
i10	Analyses are not prevented from being disregarded.
i11	Existing traceability is manual and cumbersome.
i12	It is difficult to see trends and statistics for collective impact.
i13	Tools to support the analysis are missing.
i14	Affected parties are overlooked.
i15	Analyses are performed by the wrong persons.
i16	Change request decisions are based on interest.
i17	Requirements and baseline are missing for early change requests.
i18	Analyses and change implementation evoke stress.
i19	It is not possible to see the outcome of a change request.
i20	It is difficult to see status and updates for a change request.
i21	Different change request have different levels of complexity, and there is no single method for handling all levels.
i22	Cheap, short-term solutions win over good, long-term solutions.
i23	Solutions are specified with too much detail by high-level analysts.
i24	Hardware and protocol dependencies are difficult to handle for late change requests.
i25	Relevant structure and documentation to support the analysis are missing.

on at least one other, since the sum of the relative priorities assigned to the issues is constant. The practical significance of the augmented disagreement may not be that large, since two persons could disagree much on one single issue, and then disagree equally much but very little on the remaining issues which would constitute agreement in practise (e.g., $\{520, 20, \ldots, 20\}$ vs. $\{400, 25, \ldots, 25\}$ for 25 issues). Still, such a situation is not very likely to occur, and unless the disagreement is on very few issues, and is closely mirrored (such that the disagreements cancel each other out), overall disagreement seems inevitable. (2) Due to the large number of issues, persons might assign non-zero priorities only to a limited subset of issues. For different persons, these subsets may be of different sizes and even non-overlapping. (3) Because of the low number of participants within each level, disagreement between two individuals contribute much to the overall disagreement in the level. It can also be argued that with few participants per level, the high level of precision in the priorities is of limited use anyway.

Given the reasons outlined above, we decided to classify the priorities assigned to issues into importance classes, labelled A to D, to facilitate the qualitative analysis. This is equivalent to looking at the *magnitude* of importance rather than just the actual importance. The ultimate purpose of the classification is to increase the ability to visualise and thereby interpret the results qualitatively. The classes and cutoff limits are described below, where w_i denotes the priority of issue i:

A. ($w_i < 40$) With 1 000 points to distribute over 25 issues, the average priority is 40. We argue that priorities below the average are unimportant. The only way for a priority to be below average is for another one to be above average, in which case the prioritising person rejected the former in favour of the latter.

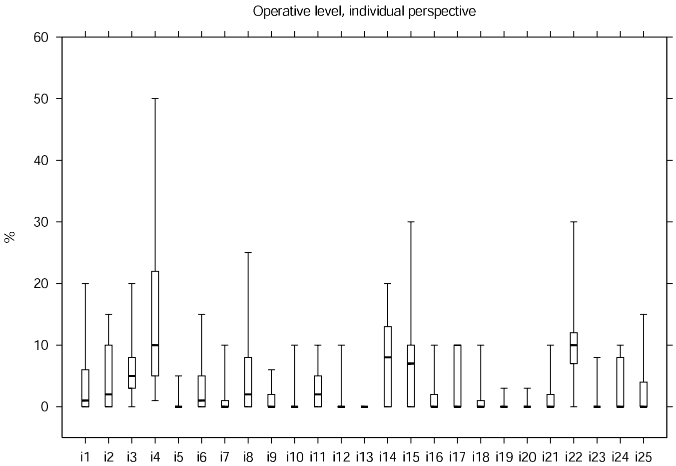


Fig. 1. Example of distribution of relative priorities (*y axis*; expressed as percentages of the total amount distributed in the CV prioritisation) for each issue (*x axis*).

B. ($40 \leq w_i < 100$) In terms of importance, this class corresponds to priorities that are above average, but below 10% of the total importance of all issues.

C. ($100 \leq w_i < 200$) This class corresponds to priorities that are between 10% and 20% of the total importance of all issues.

D. ($200 \leq w_i$) This class corresponds to priorities that are above 20% of the total importance of all issues. We argue that priorities in this class signify high importance.

The limits were chosen to distinguish among the issues, many of which have priorities that lie in the interval from 0 to 20% importance. The limits may of course have been chosen differently, but the purpose is not to provide a model for classification of importance, but rather to be able to compare how issues were prioritised at the different organisational levels and from the different perspectives.

Classifying the priorities means altering the data, which has a number of consequences:

- A certain degree of information loss occurs, as ratio-scale data is transformed into ordinal-scale data. However, as we wish to increase the ability to interpret the results, rather than being able to distinguish between slight variations in the priorities (which can depend on a number of factors, as pointed out earlier), this is acceptable.
- The amount of agreement on importance among participants becomes higher as the granularity decreases.
- Outliers can be kept and be allowed to influence the analysis and conclusions. All outliers are grouped together with high (or low) priorities in any case. See also Section VI-D.
- The non-statistical assessment of disagreement becomes decoupled from descriptive statistics such as the mean or median. This is beneficial as the mean is sensitive to outliers, while the median on the other hand might hide high priorities due to the low number of participants.

Fig. 3 shows an overview of the importance of issues for each level. For each issue, a score of importance has been calculated by weighing and summing the priorities in each class, giving class A a weight of 0, class B a weight of 1, C a weight of 2, and D a weight of 3. The scores for the tactical and strategic levels have been scaled up with a factor of 1.6, and rounded to the nearest whole number, to compensate for the fact that these levels contain

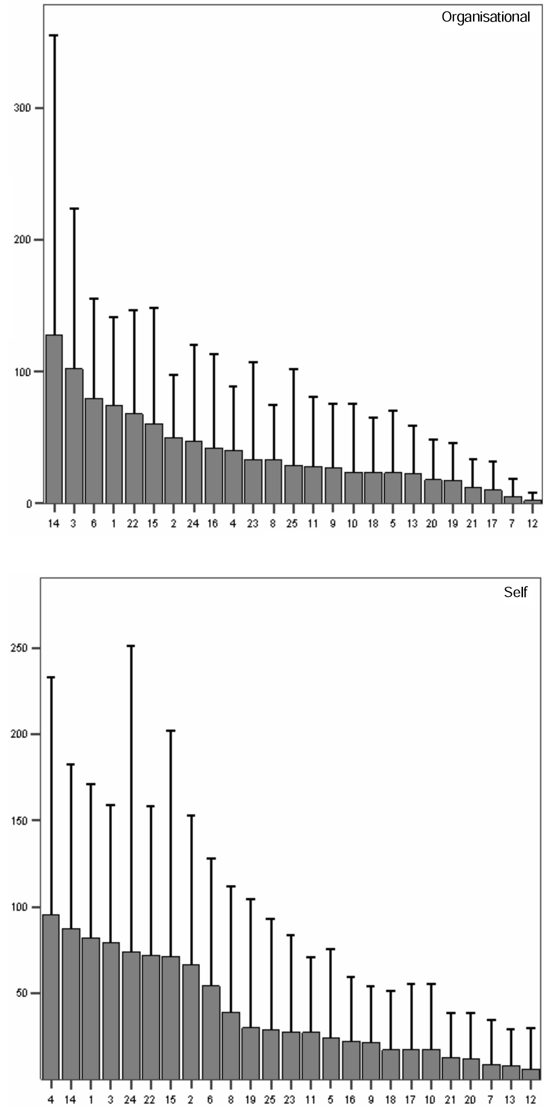


Fig. 2. Mean priorities (*y axis*) with error bars for each issue (*x axis*) under the two perspectives.

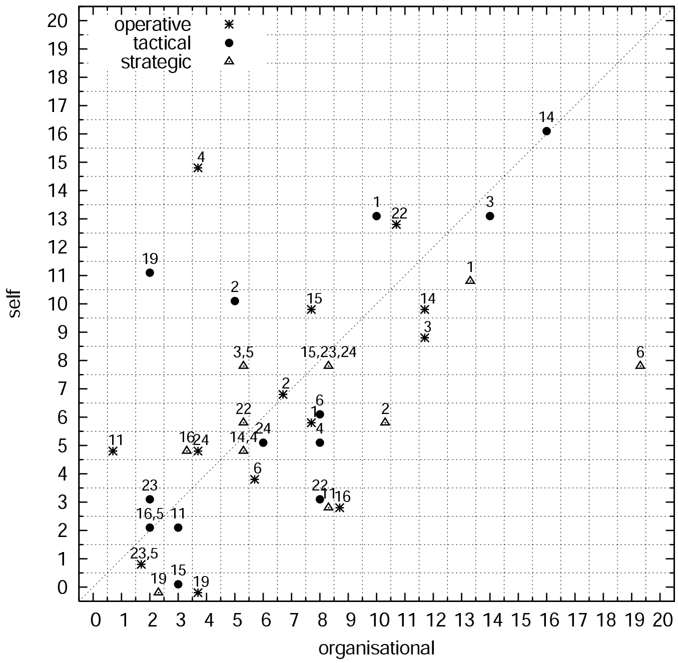


Fig. 3. Importance of issues for each level, and comparison of perspectives.

five participants each, whereas the operative level contains eight participants. The issue scores correlate well with the mean values of the relative priorities assigned through CV, but by basing the scores on the classes instead, they become more coherent. The figure only includes those issues that appear among the top five (which may be more than five issues because of ties in the scores) for any level and any perspective. In addition to showing the importance of issues, the figure also provides a comparison of the perspectives for each level. Note that the data points have been offset slightly in relation to each other to minimise overlap among the levels, but all scores are whole numbers. In other words, data points for different levels that are located in the same "grid square" have the same scores.

Two main conclusions can be drawn from studying Fig. 3. The first concerns the analysis of how the importance of issues differ among the different levels. The second concerns the analysis of how the importance of an issue is affected by the perspective. The significance of the data shown in the figure is discussed in Section VI-B.

VI. Analysis and Discussion

In this section, we discuss the analysis of the results presented in the previous section. Both a quantitative (statistical) and a qualitative analysis have been performed, the former on the original data and the latter on the classified data. The quantitative analysis targets research question 2 (RQ2), whereas the qualitative analysis targets all three research questions. The outline of the qualitative analysis with respect to the research questions is provided in Section VI-B.

A. Quantitative Analysis

The purpose of the statistical analysis was to investigate how the opinions of the participants, expressed as numerical prioritisation values, differ according to the prioritisation perspective. It is important to emphasise that the type of numerical data we obtained from the 1 000 points CV method presents many peculiarities. This is due to the inherited inter-dependence of the values summing to a constant value (1 000 in our case) and the large number of zeros appearing in the responses. This is the reason for using several statistical tests, in order to study the data in the most thorough and complete manner possible.

Traditional parametric tests like the t-test and the ANOVA cannot be applied to the original data since the normality assumptions are not valid for these data. Non-parametric tests like the Wilcoxon signed rank test are also problematic due to the large number of ties in the ranked data (see [17]) that follow from a large number of zeros or equal priorities (e.g., it can be expected that people are inclined to put 50–50–50 rather than 49–50–51 on three issues of similar importance).

The *coefficient of variation* (CoV), defined as the ratio of the standard deviation to the mean, can be used as a measure of disagreement [18], in the sense that a high CoV would be an indicator of disagreement. However, a mean value close to zero inflates the measure, thereby limiting its usefulness. For example, if a number of priorities for an issue are zero or close to zero, their mean will be close to zero. This would result in a high CoV, though it can be argued that the agreement on the unimportance of the issues is high. To make the metric useful, it would have to be judged along with the median, the mean, or preferably both. Furthermore, CoV is a univariate test, whereas a multivariate test should be used here.

1) Internal disagreement within each perspective: We first look at the internal disagreement in each perspective. Since our data are numeric, it is reasonable to express the abstract notion of "disagreement" by a dissimilarity measure computed from the numerical responses. The most interesting question is whether participants agree more under the organisational perspective than under the self-perspective. In order to test this, a multivariate measure of dissimilarity should be used. For the constant-sum data resulting from the prioritisations, an appropriate measure is the Chi-square statistic. Suppose that the responses of participant $i \in \{1, \ldots, 18\}$ under the perspective $P \in \{O, S\}$ (organisational and self) is a vector:

$$(p_{i1}^{(P)}, \ldots, p_{i25}^{(P)}), \text{where } p_{ik}^{(P)} \geq 0, \sum_k p_{ik}^{(P)} = 1000$$

Then, the dissimilarity of participants i and j is measured as follows under any of the two perspectives:

$$C_{(i,j)}^{(P)} = \sqrt{(\chi^2)_{ij}^{(P)}} = \sqrt{\sum_k \frac{(p_{ik}^{(P)} - E(p_{ik}^{(P)}))^2}{E(p_{ik}^{(P)})} + \sum_k \frac{(p_{jk}^{(P)} - E(p_{jk}^{(P)}))^2}{E(p_{jk}^{(P)})}}$$

where

$$E(p_{ik}^{(P)}) = E(p_{jk}^{(P)}) = \frac{p_{ik}^{(P)} + p_{jk}^{(P)}}{2}$$

Note that the summation above is performed for all issues $k \in \{1, \ldots, 25\}$ that have $p_{ik}^{(P)} + p_{jk}^{(P)} \neq 0$.

We used this dissimilarity measure which is known to be appropriate for proportional data, since our data can be expressed very easily as proportions if they are simply divided by 1 000. The application of the Chi-square measure to each of the two perspectives produces two dissimilarity matrices which are symmetrical. These matrices are provided with rounded values in Table III as a combined matrix. The values in the lower triangular part of the matrix are the distances between the participants under the self-perspective, while the values in the upper triangular part are the distances under the organisational perspective. The diagonal contains the distances of the answers of each of the participants under the two perspectives.

It can be observed in Table III that there is no apparent trend in the distances between participants—some of them are almost the same under the two perspectives (i.e., the participants differ equally much irrespective of perspective), while others differ quite much. Fig. 4 shows box plots and descriptive statistics for the two distributions of distances. We can see that the dissimilarity, and consequently the disagreement, is increasing in the self-perspective.

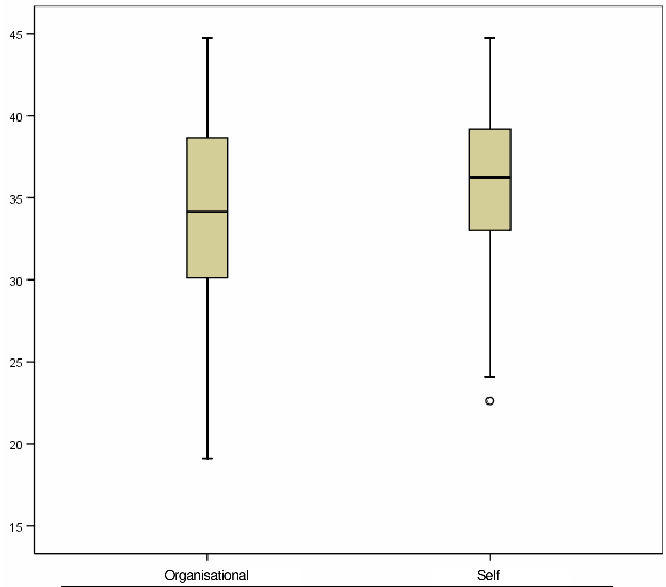
2) Comparison of disagreement between the perspectives: The question is how to test if there is a difference in the disagreement of the two perspectives. If we consider all the distances of both perspectives as pairs and use some statistical test, then the assumption that the observations used for the test are independent is violated (since, for example, the distance between participants 1 and 2 is not independent from 1 and 3). However, if we choose to ignore this assumption, the paired t-test shows a significant difference ($p < 0.001$) between the perspectives, which is confirmed by the Wilcoxon test ($p < 0.001$). Also, in 91 out of 153 cases the distance between two participants is greater under the self-perspective than under the organisational perspective.

All of the above results indicate that the disagreement tends to grow under the self-perspective. It should be noted that the square of each distance can be used to test if any two of the

TABLE III

DISTANCES OF PARTICIPANTS PER PERSPECTIVE, USING THE χ^2 DISSIMILARITY MEASURE.

		Organisational perspective																	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Self-perspective	1	**19**	34	31	36	40	37	27	41	39	34	33	42	27	32	43	26	27	37
	2	33	**34**	36	34	39	34	31	41	39	35	34	37	31	31	45	32	35	41
	3	37	35	**30**	28	30	34	20	35	37	32	27	38	29	23	40	22	34	29
	4	38	40	32	**22**	33	29	30	38	34	34	22	39	29	22	39	24	32	25
	5	33	34	37	32	**33**	41	36	32	40	36	35	42	35	32	37	36	38	27
	6	35	33	30	26	36	**19**	32	35	42	38	27	39	35	30	45	31	31	36
	7	25	30	27	33	30	32	**15**	34	36	31	21	35	27	22	41	20	26	32
	8	34	35	39	40	31	41	34	**35**	45	42	34	39	43	36	37	35	33	34
	9	37	33	30	39	39	36	34	45	**17**	30	36	39	32	33	45	39	39	38
	10	34	41	33	35	40	36	35	45	34	**29**	32	40	28	27	45	33	33	36
	11	29	32	36	38	37	35	29	34	45	32	**30**	30	32	23	40	19	26	27
	12	38	39	35	45	45	42	33	35	42	43	35	**22**	40	37	45	34	43	37
	13	44	37	28	37	37	37	36	42	38	45	45	39	**33**	25	43	28	34	31
	14	36	38	23	37	37	37	29	42	27	34	41	41	32	**21**	43	24	31	27
	15	42	39	33	37	36	37	35	41	37	41	45	42	27	27	**33**	39	40	40
	16	36	38	32	34	32	34	27	42	37	38	39	37	32	28	24	**22**	29	28
	17	34	37	40	42	41	41	34	43	41	38	37	39	43	39	43	41	**31**	35
	18	38	36	28	30	29	33	32	35	34	33	38	41	32	32	35	33	41	**0**



Persp.	min	mean	median	st.dev	max
Org.	19	34	34	6	45
Self	23	36	36	5	45

Fig. 4. The distribution of distances between participants in the two perspectives, box plots and descriptive statistics.

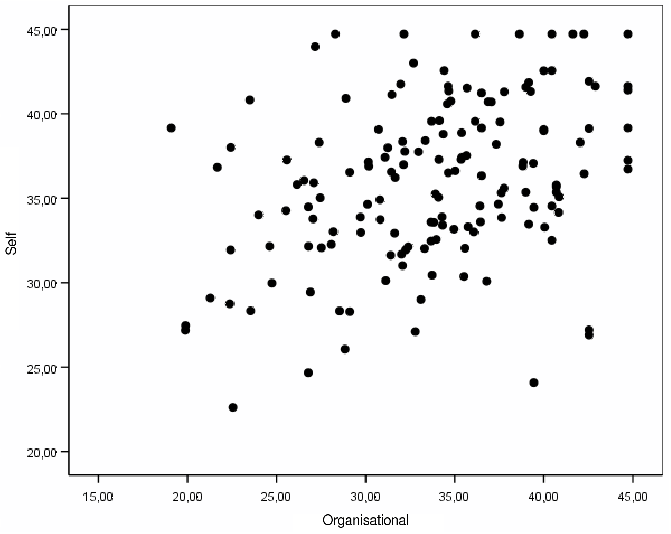


Fig. 5. Correlation between the distances under the two perspectives.

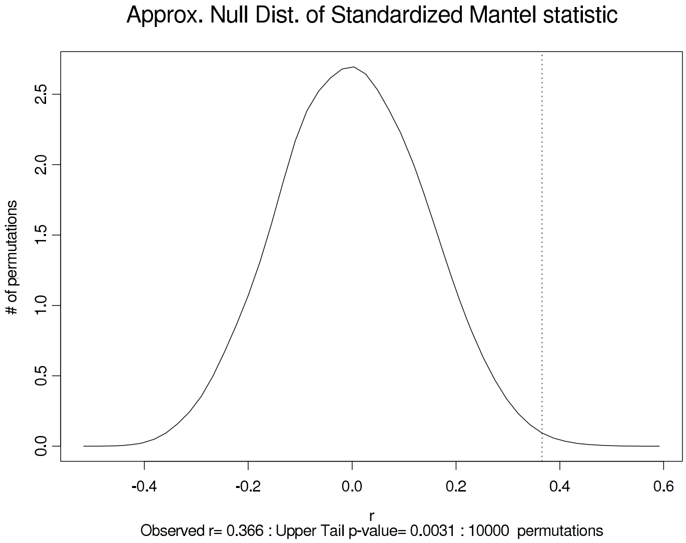


Fig. 6. Results of the Mantel test for the two distance matrices.

prioritisation vectors obtained from the participants come from the same distribution. The smallest square from the two triangular parts in Table III (excluding the diagonal) is $19^2 = 361$, while the 95% critical value of the Chi-square statistic is 36.42 for 24 degrees of freedom. This in practice means that for any two participants, their disagreement is statistically significant in both perspectives. It is also useful to observe the diagonal of Table III where we can see large disagreement of the answers of each participant separately between the two perspectives. Only participant 18 prioritised in exactly the same way under both perspectives. All the others give significantly different answers (the smallest diagonal square value is $15^2 = 225$).

Next, we consider the correlation between the two distance matrices. Fig. 5 presents a scatter plot of the distances, which shows a clear positive correlation between the perspectives. If we again ignore the independence assumption and compute the most known correlation coefficients (Pearson's, Kendall's and Spearman's), we find that Pearson's is 0.366, Kendall's is 0.254 and Spearman's is 0.355. All three show a significant correlation ($p < 0.001$).

In order to test the correlation of the two matrices with a more valid test which overcomes the assumption of independence, we use the *Mantel test* which is especially designed for testing the null hypothesis that two distance matrices of the same cases are uncorrelated [19]. The test is based on one of the usual correlation coefficients (e.g., Pearson's) but uses for the computation of the significance a randomisation technique to permute the rows and columns of one of the distance matrices a large number of times. For the two distance matrices, the application of 10 000 permutations resulted in an empirical distribution which shows that the observed correlation coefficient ($r = 0.366$) has significance $p = 0.0031$. This in turn means that the correlation of dissimilarities between the two perspectives is highly significant (see Fig. 6). For the application of the Mantel test, the Splus function *mantel.fcn*() was used, provided by J. Reynolds [20].

As a conclusion of all the above we can say that:

- There is a large amount of disagreement between all the participants under both perspectives,
- the disagreement on average grows under the self-perspective, and
- the disagreement in both perspectives are significantly positively correlated.

TABLE IV

DESCRIPTIVE STATISTICS FOR THE DISTANCES IN THE THREE LEVELS UNDER EACH PERSPECTIVE.

Group	N	min	mean	med.	st.dev	max
Org., O	28	19.88	33.85	34.13	4.75	40.83
T	10	23.54	33.57	32.70	6.80	42.54
S	10	28.30	34.00	32.27	4.46	40.00
Self, O	28	24.67	33.58	33.65	3.94	40.58
T	10	24.08	34.29	33.78	6.55	42.56
S	10	32.12	39.65	40.42	4.80	44.72

The third bullet is important as it means that the disagreement of participants under the organisational perspective is not independent of the disagreement under the self-perspective. In practical terms, any two participants that appear to highly disagree under the self-perspective, are expected to highly disagree under the organisational perspective too.

It is interesting to search further for the reason behind the observed disagreement growth in the self-perspective. We can see from Table IV and Fig. 7 that this is partly due to the growth in the strategic level where the difference is the only significant one (paired t-test gives $p = 0.010$ and Wilcoxon gives $p = 0.013$).

3) Correlation between perspectives: Moreover, we have looked at the correlation between the perspectives for each issue. The non-parametric Spearman correlation coefficient was used for this purpose. Table V displays the results only for the issues with significant correlations at the 0.10 level.

The significant correlation of certain issues shows that the ranking attitude of the participants towards these issues is not changing under the different perspectives. For example, those who allocate high amounts to i23 under the organisational perspective, tend to do the same under the self-perspective. This result is in accordance to the previous analysis where the dependence of the two perspectives was shown in terms of the agreement.

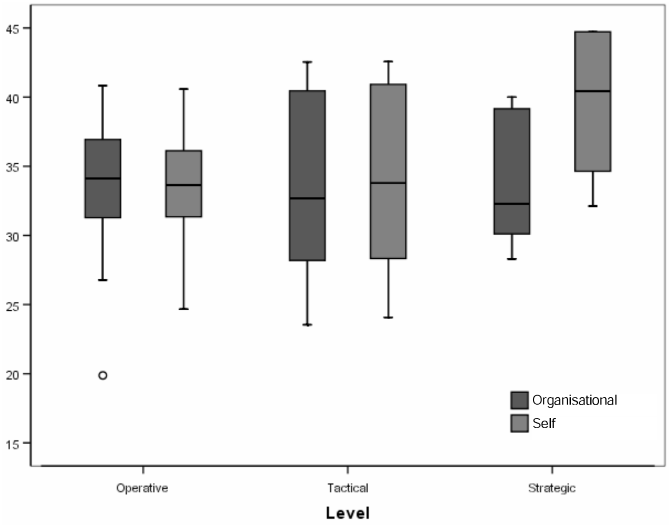


Fig. 7. The distribution of distances for the three levels under each perspective.

TABLE V

SPEARMAN'S RHO AND SIGNIFICANCE FOR ISSUES WITH SIGNIFICANT CORRELATION BETWEEN PERSPECTIVES.

	i5	**i6**	**i8**	**i9**	**i10**	**i11**	**i14**
ρ	.470	.450	.500	.431	.697	.474	.742
Sig.	.049	.061	.035	.074	.001	.047	.000
	i15	**i18**	**i20**	**i22**	**i23**	**i24**	**i25**
ρ	.584	.677	.467	.721	.845	.666	.578
Sig.	.011	.002	.051	.001	.000	.003	.012

B. Qualitative Analysis

The qualitative analysis is based on the arrangement of priorities into importance classes as discussed in Section V, and only on those issues that are shown in Fig. 3. We first discuss how issue importance is affected by the organisational level (research question RQ1), then how it is affected by the perspective (RQ2). Thereafter, we discuss issues of general importance, and possible improvements for these (RQ3).

1) Level Differences: The issues with high disagreement among organisational levels are those for which the data points are far apart in Fig. 3. Note that this definition of level disagreement also takes into account disagreement between perspectives; two levels are said to disagree if an issue is important under one perspective for the first level, but under the other perspective for the second level. The perimeter of the triangle defined by the level data points can be used as a measure of the overall disagreement among the levels for a particular issue; the term *perimeter distance* is used onwards. The perimeter distance (D_k) for issue k corresponds to the sum of the Euclidean distances between each pair of issue data points. Let $X = \{O, T, S\}$ be the set of organisational levels, and $Y = \{(a, b) \in X \times X | a \neq b\}$, then:

$$D_k = \frac{1}{2} \sum_{(L,M) \in Y} \sqrt{(q_{ks}^{(L)} - q_{ks}^{(M)})^2 + (q_{ko}^{(L)} - q_{ko}^{(M)})^2}$$

where q_{ks} is the score of issue k under the self-perspective, and q_{ko} is the score under the organisational perspective. The division by 2 must be performed since each pair of levels occurs twice in Y. The five issues with the largest perimeter disagreement are listed in Table VI and discussed below:

TABLE VI

TOP FIVE ISSUES BASED ON PERIMETER DISAGREEMENT (D_k), INCLUDING (SCALED) SCORES FOR EACH LEVEL AND PERSPECTIVE.

	Operative		**Tactical**		**Strategic**		
Issue	**org.**	**self**	**org.**	**self**	**org.**	**self**	D_k
i14	12	10	16	16	5	5	31.4
i6	6	4	8	6	19	8	27.6
i19	4	0	2	11	2	0	24.2
i22	11	13	8	3	5	6	23.9
i4	4	15	8	5	5	5	23.8

- Issue **i14** is most important at the **tactical level**, but also to some degree at the **operative level**. Its importance at the tactical level most likely stems from the fact that resources, coordination, and planning are important factors to take into consideration at this level, all of which may be disturbed by the omission of some party in the analysis. At the operative level, we believe that the importance comes from the fact that the omission of a party in the analysis can result in solutions that are suboptimal and/or not general enough.
- Issue **i6** is most important at the **strategic level**. The reason for this may be that the effects are longer change request lead times and possibly decisions made on loose terms. In the end, this may affect the end product negatively (in terms of feature set). Note that the difference among levels lies mainly in the organisational perspective; the spread for the self-perspective is much less.
- Issue **i19** is most important at the **tactical level**. A possible explanation for this is that follow-up on estimates can be used to improve future planning, the ability of which is inhibited by this issue. The difference lies almost only in the self-perspective, while all levels agree on a low importance from the organisational perspective.
- Issue **i22** is most important at the **operative level**. It is reasonable to believe that short-term solutions generate rework and overhead later on (as they, by definition, are not designed to last), and that this has to be dealt with by people at the operative level. The difference among the levels comes mainly from the self-perspective.
- Issue **i4** is also most important at the **operative level**. While unclear change requests eventually will be handled, they cause annoyance, especially for those at the operative level who will implement the changes.

A pattern can be discerned in the list above. The issues that are most important at the operative level are related to the realisation of a change or the solution for a requested change. The issues most important at the tactical level are related to project and planning. Finally, the issue most important at the strategic level is related to process and end product. This is in line with what can be expected in terms of how the organisational levels are defined. An interesting observation is also that when the level difference is due to a perspective difference (issues i6, i19, and i22), the strategic level stands out because of the organisational perspective, whereas the other levels stand out because of the self-perspective.

2) Perspective Differences: In Fig. 3, issues on the line $y = x$ are equally important under both the organisational perspective and the self-perspective. The farther away from this line an

TABLE VII

THE ISSUES WITH LARGEST DIFFERENCES BETWEEN PERSPECTIVE SCORES, FOR EACH LEVEL (DOMINATING PERSPECTIVE WITHIN PARENTHESES).

Operative	Tactical	Strategic
i4 (self)	i19 (self)	i6 (org.)
i16 (org.)	i22 (org.), i2 (self)	i11 (org.)

issue is, the larger the difference is between its score for the organisational perspective and its score for the self-perspective. Initially, we expected the perspectives to differ quite much for two reasons. First, we believed that the fact that people have personal agendas should result in a prioritisation where a few issues were given high weights from the self-perspective (as, in fact, Fig. 2 indicates). Second, and related, we argued that certain issues ought to be more germane to the organisational perspective and some to the self-perspective. The quantitative analysis presented in Section VI-A supports our expectations, but it can be seen in Fig. 3 that many issues are close to the $y = x$ line irrespective of organisational level. This indicates that the perspectives in many cases are not distinguishing, if we consider the classified priorities. Table VII shows for each level the two issues (three for the tactical level because of ties) with largest difference between the perspective scores. The cutoff is chosen here for the sake of brevity. The differences are discussed below.

- Operative level, issue **i4**: this issue is seen as very important from the **self-perspective**, but not so much from the organisational perspective. We believe this is because, as noted above, unclear change requests will eventually be dealt with, but they hamper the activity of analysing the impact of a change, generating additional work and consuming time.
- Operative level, issue **i16**: this issue is seen as more important from the **organisational perspective**, most likely because this means that decisions are not in line with what is best for the organisation, whereas the realisation of a change is indifferent to the reason for its inclusion (from a technical point of view).
- Tactical level, issue **i19**: this issue is more important from the **self-perspective**. This indicates that the organisation does not suffer from the inability to do follow-ups; given skilled staff, the change is properly dealt with in any case. Note that for a small organisation, the situation could be different.
- Tactical level, issue **i22**: this issue is more important from the **organisational perspective**. It is likely that this is because the short-term solutions will have to be replaced later on, thus causing planning problems, additional required resources, etc.
- Tactical level, issue **i2**: this issue is more important from the **self-perspective**. People at the tactical level need to balance many factors regarding resources, tasks, and time constraints. A high normal workload leaves little room for analysing proposed changes.
- Strategic level, issue **i6**: this issue is most important from the **organisational perspective**. It is most likely so, that the organisation as a whole suffers from negative end product effects (as discussed previously), whereas individuals manage to do their work anyway.
- Strategic level, issue **i11**: this issue is most important from the **organisational perspective** (but notably less important than issue i6). A reason for this can be that when it is difficult to ensure traceability, the level of transparency decreases, and so also the ability to check that all development is in line with the goals of the organisation.

The pattern here is that the issues most important from the organisational perspective concern process, project, or end-product aspects, whereas the issues most important from the self-perspective rather concern the execution of the analysis task, and/or the realisation of a change.

3) Overall Importance: Intuitively, considering Fig. 3, an issue has high overall importance if the points for all levels lie in the upper right corner of the chart, and close to the $y = x$ line. Then, participants on all levels agree on the (high) importance of the issue, and do so irrespective of perspective. However, it can also be argued that an issue is important to take care of if it is of high importance on any level, from any perspective. This reasoning can be used to divide issues into priority classes. Issues of high priority are the ones that have high *overall* importance, whereas issues of medium priority are the ones that have high importance for at least one group.

We have done this division as follows. The overall score of issue $k \in \{1, \ldots, 25\}$ is the average distance from the origin to the issue data point (Fig. 3) over all levels $L \in \{O, T, S\}$ (operative, tactical, and strategic):

$$Q_k = \frac{1}{3} \sum_L \sqrt{(q_{ks}^{(L)})^2 + (q_{ko}^{(L)})^2}$$

where q_{ks} is the score of issue k under the self-perspective, and q_{ko} is the score under the organisational perspective. The summation is over all three levels.

The cutoff limit for the selection of high-priority issues was set to $Q_k > 10$. This limit was chosen in order to include an issue so important that more than five out of eight[1] of the priorities are in class C (the class weights are discussed in Section V) and the remaining in class A (e.g., $6 \times 2 + 2 \times 0 = 12$). Of course, there are other sets of priorities equivalent to that situation. For example, an issue with three class D priorities, two class B priorities, and three class A priorities would also be included ($3 \times 3 + 2 \times 1 + 3 \times 0 = 11$). Because of the way the class weights relate to each other, we might see some peculiar priority sets that make it over the limit, but we reckon that the limit makes sense in most cases.

Using similar reasoning, medium-priority issues are considered to be those that have a score of more than 15 for at least one of the levels under at least one of the perspectives (i.e., $\max(q_{ks}, q_{ko}) > 15$ for any level). This cutoff limit was set higher, as this criterion is more inclusive than the one for determining high-priority issues. This limit was chosen in order to include an issue so important that more than five out of eight of the priorities are in class D and the remaining in class A (e.g., $6 \times 3 + 2 \times 0 = 18$), or with a set of priorities equivalent to that situation.

Based on the division just described, the high-priority issues are **i1** and **i3**, while the sole medium-priority issue is **i6** (also note that i6 is the only of these issues that has surfaced in the previous discussions about differences, showing that it indeed is of local importance). It can be seen that the high-priority issues are concerned with the execution of the analysis task and the result of the analysis, respectively, which arguably are fundamental aspect of the change management process. The medium-priority issue

[1] After scaling the scores, all levels can be said to contain eight participants.

is concerned with both of these aspects. This result indicates a need for support in the process, in order to alleviate the effort of dealing with and analysing changes, and thereby increasing the likelihood of a good result.

A goal of process improvement must necessarily be to address as many issues as possible. However, seeing to the cost aspects, it is of course more optimal to focus on issues that are of overall high criticality. Having said that, it is important to note that addressing an issue has a positive effect on people experiencing the issue, regardless of the criticality they have assigned to it. Differentiating between levels and perspectives is a way to make more informed decisions about which issues to deal with first.

For medium-priority issues, there is the question of determining which perspective and/or level to satisfy (in terms of trying to address the issue) first. For example, addressing issues important under the organisational perspective is of course in the best interest of the organisation, while addressing issues important under the self-perspective should facilitate the daily work of the individual software developers. As we see it, the selection of issues to try to address should include both criticality and cost. Three aspects of dealing with issues are important:

1) Focus on issues that have high criticality, in particular high-priority ones.
2) Also focus on issues that can be addressed in inexpensive ways.
3) Finally, consider improvements that affect as many issues as possible (discussed later).

We believe that decisions based on perspective and/or level are specific for the particular situation and context, and we do not feel confident in speculating about how these aspects should be handled.

C. Improvement Proposals

During the workshops, a number of possible improvements to help address the issues at hand were discussed. These improvements are presented in this section. It is of course desirable to select those improvements that can address as many issues as possible at the same time, to a reasonable cost. As many of the issues are related in one way or another (e.g., i3 and i14 both have to do with missed impact, albeit of different types), improvements are likely to have multi-issue impact. However, a multi-issue improvement may be very expensive, in which case other, possibly narrower, improvements have to be considered instead.

The following list shows the improvements that were identified during the workshops:

S1 Limit the number of outstanding change requests at any time and use a selection process for taking in change requests.

S2 Introduce different ways of handling different types of change requests. For example, use different change request "tracks" (e.g., based on change request priority).

S3 Involve the design organisation more in the requirements specification work to avoid change requests related to requirements problems.

S4 Plan early for the fact that there will be many change requests in the project.

S5 Introduce a database for storing old impact analysis results to be used as a knowledge base for future analyses.

TABLE VIII

MAPPING FROM IDENTIFIED IMPROVEMENTS TO HIGH- (*top*) AND MEDIUM-PRIORITY (*bottom*) ISSUES.

	S1	S2	S3	S4	S5	S6	S7	S8	S9
i1	x	x	x	x					
i3					x	x	x	x	x
i6	x	x		x	x				x

S6 Introduce meetings where different development teams and sub-projects discuss joint impact analysis efforts.

S7 Involve the support organisation more in the impact analysis work to ensure focus on post-delivery aspects.

S8 Missionise the target architecture more within projects to make sure that everyone knows about it.

S9 Introduce tool and method support (such as checklists) to help individuals perform better and more robust analyses.

S10 Act more towards the customer—meet the customer more often to discuss solutions. Listen to several customers instead of only one to avoid solutions that just a few need/want.

S11 Anchor release intent in the design organisation to make sure that decisions go towards a common goal.

S12 Do not allow change requests late in the project (close the change control board as early as possible).

Each improvement was subsequently mapped to one or more issues. The mapping was done manually by cross-comparing each improvement with each issue, discussing whether or not the improvement should be effective in mitigating the issue. Thus, the mapping involved experience, information from relevant literature, and knowledge about the company. An improvement package was formed for each issue, and most improvements were shared among several packages. Table VIII shows the mapping between the issues that are of high overall importance and the improvement proposals (S10–S12 are excluded as they do not map to any of the issues shown in the table). An x in a cell means that the improvement in the current column should be effective in addressing the issue in the current row. It should be noted that the mapping of improvements to issues mainly took place during the workshops, although it was reviewed and in some cases completed afterwards.

As pointed out, improvements affecting several issues should be favoured over more localised improvements. Thus, improvements S1, S2, S4, S5, and S9, which all affect two issues, are good candidates for a process improvement programme. Of course, one single improvement is unlikely to completely resolve an issue—it is better to use a combination of multiple improvements.

In addition to the ones presented above, a number of improvements were tailored specifically for the studied company. As these are not generic, they have been left out here.

D. Treatment of Outliers

A participant who has prioritised completely different than other participants, for example by giving extremely high weights to certain issues, could be seen as an outlier. Outliers disturb statistical measures, such as the mean value, and may have severe impact on statistical tests. However, in our statistical analysis we use the non-parametric Chi-square test, which makes no assumption about the distribution of the data. Fig. 1 illustrates some of the data from the prioritisations, including outliers, for the sake of showing the distribution of priorities. The classification of priorities into importance levels allowed us to retain the importance of the outliers, while reducing their extremeness to normal levels.

We argue that dealing with outliers is of particular importance when the data consists of measurements of performance or effort, in which case outliers could be seen as the result of measurement or sampling problems. In our case, however, each data point is a subjective opinion, meaning that removing outliers would disqualify uncommon opinions. A way of dealing with outliers resulting from uncommon or extreme opinions is to go back to the source and discuss the reason behind the outlier. This approach has, however, not been pursued in this study.

E. Threats to Validity

The validity of a study is often divided into four types. *Construct validity* is concerned with the design of the main study instrument and that it measures what it is intended to measure [15]. A study has *internal validity* if there is a causal relationship between the treatment and the outcome [15]. *External validity* is concerned with the generalisability of the results [21]. Finally, *conclusion validity* has to do with being able to draw correct conclusions about how the treatment has affected the outcome [21].

Two general arguments against the threats discussed here are: (a) the research is mainly exploratory in the sense that it contributes in understanding how experts tend to prioritise IA issues, and (b) the data are not responses from randomly selected individuals but from responsible experts interested in research and are for this reason invaluable.

1) External validity: The small sample size and the fact that we sampled based on convenience are threats, as the participants may not have been representative for the population. However, as the participants were selected mainly based on recommendations by several persons, we believe they were good representatives both of their roles and of professional software developers in general. Furthermore, the fact that we focused on potential issues rather than actual ones should increase the external validity. Also, the participants covered all issues from the literature, which indicates that their views of impact analysis were not company specific. These threats affect the internal validity as well.

2) Internal Validity: A threat is *maturation*, in our case that the first prioritisation of issues could have affected how the same issues were prioritised a second time (although from a different perspective). To counter this threat, we used a two-group design where participants did the prioritisations in different order (see Section IV-D).

Another threat is *instrumentation*, which means that the instruments used in the study could have been badly designed. This threat affects conclusion validity as well, and possibly construct validity. We believe the threat is reduced due to the facts that we conducted a pilot interview to test the interview instrument (see Section IV-B), and that the prioritisation tool (see Section IV-D) was designed to alleviate the prioritisation effort of the participants as much as possible.

3) Construct Validity: A threat is that the participants may not have had the desired mind set when prioritising issues. As stated in Section IV-D, we asked the participants to prioritise as if the issues currently were not present, but we could not verify if they adhered to our request.

Another threat is that the participants may not have been entirely neutral when prioritising under the organisational perspective. This argument is supported by the finding that there is a correlation between the perspectives. It is probably difficult to push the self-perspective aside, so there is a risk that this perspective (here, we refer to state of mind rather than prioritisation because, as noted before, we used a two-group design to avoid maturation) has coloured the organisational perspective in the prioritisations. A solution for reducing this threat would have been to let some participants prioritise under one perspective only, and the others under the other perspective only. However, we chose not to do this due to the small sample size.

4) Conclusion Validity: A threat is that the participants may have been disturbed when doing the prioritisation. As they prioritised individually in their work place, we could not control this threat. Since we find it unlikely that any disturbance would prevent a participant from returning to his or her "prioritisation mindset," or if necessary start over with the prioritisation, we do not believe this threat to be very significant.

Another threat is the selection of limits (cutoff values) in the classification of priorities, weighing of classes, and identification of top issues. We do not claim that our choices are optimal, but it should be noted that the purpose of this paper is not to provide a specific truth about the absolute importance of issues, but rather to study the issues relative to each other. Thus, we claim that the results and analyses presented are valid with respect to the research questions.

VII. Conclusions

In this paper, we have presented results from a study where impact analysis issues were prioritised by software professionals at three different organisational levels (*operative*, *tactical*, and *strategic*) and under two different perspectives (*organisational* and *self*). The organisational perspective was used to learn about which issues were critical for the organisation, while the self-perspective was used to learn about issues critical for the software professionals themselves. The following paragraphs discuss the research questions posed in Section II.

RQ1: How does the organisational level affect one's assessment of importance of IA issues? From the point of view of organisational levels, we have discussed a number of issues for which there is disagreement among the levels on importance. While the data are not enough to provide statistical evidence of the cause for disagreement, it appears as if the important issues for a particular level are those that can be expected based on how the level is defined. At the operative level, which is concerned with realising the project according to plan, important issues are related to the realisation of a change, or the chosen solution. At the tactical level, which is concerned with planning of time and resources, important issues are related to project and planning aspects. Finally, at the strategic level, which addresses long-term goals and product aspects, important issues are related to process and product quality. We have also found that when the level difference depends on a perspective difference, the strategic level differs because of an organisational focus, whereas the operative and tactical levels differ because of a self-focus.

RQ2: What difference does the perspective make in determining the relative importance of IA issues? We have seen that the characteristics of an issue seem to affect how it is prioritised from the two different perspectives—issues that are most important from the organisational perspective typically concern process, project, or end-product aspects, and thus are more holistic in their nature. On the other hand, issues most important from the self-perspective rather concern the execution of the analysis task, and/or the realisation of a change. These issues are of more local nature. A finding is that the participants focused more on fewer issues under the self-perspective than under the organisational perspective. The statistical analysis shows that the agreement among participants within the organisational perspective is larger than within the self-perspective; however, it also reveals that the participants disagree with each other regardless of perspective. The disagreement within perspectives are positively correlated, meaning that the organisational perspective is not free from disagreement originating in the self-perspective. We conclude that personal opinions and attitudes cannot easily be disregarded from.

RQ3: Which are important IA issues, and how are these addressed in software process improvement? In comparing the perspectives and the levels, we visualised the differences in a way that allowed us to discuss two classes of issues: high-priority and medium-priority. The priority classes can serve as a means to optimising a process improvement effort, as focus should be put mainly on the high-priority issues. These are issues that have overall high importance (criticality), and addressing them would benefit a large group of people. Medium-priority issues are important at least at some level, under some perspective. Addressing them is beneficial for a smaller group of people, but it does nevertheless make a difference. All issues found to have high or medium priority concern the execution of the analysis task or the outcome thereof: *it is difficult to find resources for performing impact analysis*, *system impact is underestimated or overlooked*, and *analyses are incomplete or delayed*. This is not surprising, as these are fundamental aspects of the change management process.

A number of improvements for addressing the most important issues have been proposed. Each of the following five affects two out of three of the high- and medium-priority issues, and can therefore be considered highly relevant:

1) Limit the number of outstanding change requests at any time and use a selection process for taking in change requests.
2) Introduce different ways of handling different types of change requests. For example, use different change request "tracks" (e.g., based on change request priority).
3) Plan early for the fact that there will be many change requests in the project.
4) Introduce a database for storing old impact analysis results to be used as a knowledge base for future analyses.
5) Introduce tool and method support (such as checklists) to help individuals perform better and more robust analyses.

In an improvement effort, these improvements should be combined, although the cost aspect needs to be taken into consideration, to maximise the outcome with respect to addressing existing issues. The issues and improvements presented in this paper are of generic nature, and should apply to other companies than the studied one. Thus, they could be used in future research about impact analysis as a change management activity.

In summary, we want to highlight the following two points:

(1) Studying issues from multiple perspectives and levels is rewarding and entails certain benefits from a process improvement point of view. To answer two of our research questions, we have found that both level (RQ1) and perspective (RQ2) affect the assessment of importance of impact analysis issues.

(2) The most important issues (RQ3) concern fundamental aspects of impact analysis and its execution. This underlines that impact analysis needs to be addressed as a crucial activity in the change management process. More specifically, it is imperative to realise that processes need to not only enforce the existence of proper impact analysis, but also to prescribe how impact analysis should be carried out in order to achieve satisfactory and timely results.

Acknowledgements

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`). The authors wish to thank the anonymous reviewers for the constructive comments which helped to improve the paper. We also thank the people at Ericsson who have been involved in the study.

References

[1] S. A. Bohner and R. S. Arnold, *Software Change Impact Analysis*. Los Alamitos, CA, USA: IEEE Computer Society Press, 1996.

[2] A. Aurum and C. Wohlin, "The fundamental nature of requirements engineering activities as a decision-making process," *Information and Software Technology*, vol. 45, no. 14, pp. 945–954, 2003.

[3] A. Ngo-The and G. Ruhe, "Decision support in requirements engineering," in *Engineering and Managing Software Requirements*, A. Aurum and C. Wohlin, Eds. Springer-Verlag, 2005, ch. 12, pp. 267–286.

[4] R. N. Anthony, *Planning and Control Systems: A Framework for Analysis*. Boston, MA, USA: Harvard University Press, 1965.

[5] P. Jönsson and C. Wohlin, "Understanding impact analysis: an empirical study to capture knowledge on different organisational levels," in *Proceedings of the 17th International Conference on Software Engineering and Knowledge Engineering*, Taipei, Taiwan, Jul. 2005, pp. 707–712.

[6] ——, "A study on prioritisation of impact analysis issues: A comparison between perspectives," in *Proceedings of the 5th Conference on Software Engineering Research and Practice in Sweden*, Västerås, Sweden, October 2005, pp. 11–19.

[7] T. Hall and D. Wilson, "Views of software quality: a field report," *IEE Proceedings on Software Engineering*, vol. 144, no. 2, pp. 111–118, 1997.

[8] R. Conradi and T. Dybå, "An empirical study on the utility of formal routines to transfer knowledge and experience," in *ESEC/SIGSOFT FSE*, Vienna, Austria, September 2001, pp. 268–276.

[9] D. Karlström, P. Runeson, and C. Wohlin, "Aggregating viewpoints for strategic software process improvement—a method and a case study," *IEE Proceedings Software*, vol. 149, no. 5, pp. 143–152, 2002.

[10] V. R. Basili, S. Green, O. Laitenberger, F. Lanubile, F. Shull, S. Sørumgård, and M. V. Zelkowitz, "The empirical investigation of perspective-based reading," *Empirical Software Engineering: An International Journal*, vol. 1, no. 2, pp. 133–164, 1996.

[11] B. Regnell, P. Runeson, and T. Thelin, "Are the perspectives really different?—further experimentation on scenario-based reading of requirements," *Empirical Software Engineering*, vol. 5, no. 4, pp. 331–356, 2000.

[12] A. Finkelstein and I. Sommerville, "The viewpoints FAQ," *Software Engineering Journal*, vol. 11, no. 1, pp. 2–4, 1996.

[13] S. Zahran, *Software process improvement: practical guidelines for business success*. Boston, MA, USA: Addison-Wesley Professional, 1998.

[14] J. Karlsson and K. Ryan, "A cost-value approach for prioritizing requirements," *IEEE Software*, vol. 14, no. 5, pp. 67–74, 1997.

[15] C. Robson, *Real World Research*. Blackwell Publishing, 2002.

[16] D. Leffingwell and D. Widrig, *Managing Software Requirements—A Unified Approach*. Boston, MA, USA: Addison-Wesley Professional, 1999.

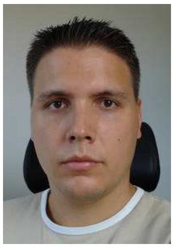
[17] D. J. Sheskin, *Handbook of parametric and non-parametric statistical procedures*. Chapman & Hall/CRC, 2004.

[18] B. Regnell, M. Höst, J. Natt och Dag, P. Beremark, and T. Hjelm, "Visualization of agreement and satisfaction in distributed prioritization of market requirements," in *6th International Workshop on Requirements Engineering: Foundation for Software Quality*, Stockholm, Sweden, Jun. 2000.

[19] N. Mantel, "The detection of disease clustering and a generalized regression approach," *Cancer Research*, vol. 27, pp. 209–220, 1967.

[20] J. Reynolds, "Mantel test code," http://www.biostat.wustl.edu/archives/html/s-news/2001-03/msg00154.html, 2001, last checked 2008-04-06.

[21] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in Software Engineering: An Introduction*. Norwell, MA, USA: Kluwer Academic Publishers, 2000.



Dr. Per Jönsson is currently with Ericsson AB in Sweden. He received his M.Sc. and Ph.D. degrees in Software Engineering from Blekinge Institute of Technology in Sweden. His research interests include change impact analysis, requirements engineering, and software architecture, primarily in the context of software process improvement.

Dr. Lefteris Angelis received his B.Sc. and Ph.D. degrees in Mathematics from Aristotle University of Thessaloniki (A.U.Th.). He is currently an Assistant Professor at the Department of Informatics of A.U.Th. His research interests involve statistical methods with applications in information systems and software engineering, computational methods in mathematics and statistics, planning of experiments and simulation techniques.

Dr. Claes Wohlin is a professor of software engineering at Blekinge Institute of Technology in Sweden and provost of the Institute. He is currently also a visiting professor at Chalmers University of Technology. Prior to this, he has held professor chairs at Lund University and Linköping University. He has a Ph.D. in Communication Systems from Lund University and five years of industrial experience. His research interests include empirical methods in software engineering, software metrics, software quality, requirements engineering and systematic improvement in software engineering. Claes Wohlin is the principal author of the book "Experimentation in Software Engineering—An Introduction" published by Kluwer Academic Publishers in 2000. He has published over 150 articles internationally. He is editor-in-chief of the journal of Information and Software Technology published by Elsevier. Dr. Wohlin is on the editorial boards of Empirical Software Engineering: An International Journal, Software Quality Journal and Requirements Engineering Journal. Dr. Wohlin was the recipient of the Telenor Nordic Research Prize in 2004 for his achievements in software engineering and improvement of reliability in telecommunication systems.