

Factors Determining Long-term Success of a Measurement Program: An Industrial Case Study

Mirosław Staron*, Wilhelm Meding**

**Department of Computer Science and Engineering, Chalmers | University of Gothenburg*

***Ericsson SW Research, Ericsson AB*

mirosław.staron@ituniv.se, wilhelm.meding@ericsson.com

Abstract

Introducing measurement programs into organizations is a lengthy process affected by organizational and technical constraints. There exist several aspects that determine whether a measurement program has the chances of succeeding, like management commitment or existence of proper tool support. The establishing of a program, however, is only a part of the success. As organizations are dynamic entities, the measurement programs should constantly be maintained and adapted in order to cope with changing needs of the organizations. In this paper we study one of the measurement programs at Ericsson AB in Sweden and as a result we identify factors determining successful adoption and use of the measurement program. The results of our research in this paper are intended to support quality managers and project managers in establishing and maintaining successful metrics programs.

1. Introduction

Several authors have already discussed factors that determine successful measurement program adoption at a company, e.g. [1–3]. The results usually are focused on addressing the question “How to establish a measurement program at a company?” which is a prerequisite for the success of the measurement program. Little, however, has been said about the factors that determine if a successfully implemented measurement program “lives” longer than just the first project for which it was established (or until the first re-organization). In this paper we present a study which we conducted at Ericsson AB, which identifies and prioritizes factors important in long-term adoption of a measurement program. Ericsson, being one of the largest telecommunication equipment manufacturers in the world, has a distributed organization and

a whole spectrum of projects (from small to very large).

The main processes are stable in the organization despite re-organizations, process customizations, and usage of various tools is normal situations in the company – conditions which are prevalent in software engineering and uncommon in manufacturing industries. These factors make the needs for measurement programs change constantly and require the program to evolve. In this paper we present results from a survey conducted at the company assessing the success of the measurement program and the measurement systems used in it. The results of this survey are combined with results of interviews with designers of measurement systems in industry to identify the success factors.

In contrast to existing body of knowledge in software engineering, instead of focusing on the establishment of the measurement program,

which most of the articles discuss, we focus on addressing the question of “Keeping the measurement program alive” as identified by Clark [4]. Therefore in our research we address the following research question:

Which are the main factors determining a long-term success of a measurement program?

By using the term “long-term” we mean that the measurement program is used in the organization in more than in a single project, that it gets extended over time, and that it becomes “the new way of working” in the organization (gets integrated in the organization in the everyday work) – the studied measurement program is in existence for 5 years at the time of this study.

The main contribution of our work is identification of four key roles in establishing long-term measurement programs: section manager, stakeholder, quality manager and designer of measurement systems. A number of success factors which are associated with each of the role separately and with several roles together (which is shown through cluster analysis using K-Means tests for clusters). These factors help the roles in being effective and efficient when establishing measurement programs. By efficient we mean that it is possible to run measurement program for an organization of several hundred employees with small resources (ca. 2 full-time employees) dedicated for measure collection, analysis and presentation. We present our factors with short experience reports of how this worked on the case of the studied organization; these guidelines are intended to help other practitioners in realizing measurement programs in other companies.

The paper is structured as follows. Section 2 presents the most related research in the field. Section 3 presents the design of the study and with its subjects, objects, and instruments. Section 4 presents the elicited success factors preceded by the direct results of the case study. Section 5 evaluates validity of our study while section 6 presents the conclusions.

2. Related Work

We investigated the following publications in order to elicit factors important when introducing metric programs into organizations in general, and not to be constrained only to Ericsson’s context:

- Umarji and Emurian [1]: the study describes the use of technology adoption theory when implementing metric programs with focus on social issues. One of the important results from that study was the importance of the factor “ease of use”. When developing our framework we invested in making the framework easy to use and making the presentation of the indicators easy to interpret.
- Gopal et al. [5] and Gopal et al. [6]: these studies present results and conclusions from a survey about metric program implementation conducted with managers at various levels (over 200 data points). The results indicated the importance of such factors as management commitment and the relative low importance of such factors as data collection. In order to check how important the framework is for the managers who we work with, we included the line manager and the project manager in our interviews when evaluating the framework.
- Atkins et al. [2]: among other aspects, this paper discusses how metrics can be reused by projects working on similar things in parallel. We used their experiences when reasoning about the reuse of metrics between different instances of the framework.
- Lawler and Kitchenham [7]: based on the experiences of several case studies, this paper discusses the issues of using metrics at different levels and combining metrics together (e.g. combining metrics from particular designers to provide the status of the whole project). This work affected the design of the framework in such a way that the metrics in the framework can be reused and combined in a way consistent with the study by Lawler and Kitchenham.

- Kilpi [3]: this paper describes how a metric program was implemented at Nokia. We used their experiences when evaluating the framework.
- Niessink and van Vliet [8, 9]: these studies describe external factors important for software metric implementation, including the importance of the goal of software measurement processes. Our experiences support this conclusion, and the need for the monitoring status and progress resulted in finally choosing the ISO/IEC 15939 standard as a basis for our work with metrics.
- de Panfilis et al. [10]: this study describes experiences from introducing a GQM-based metric program. Our experiences showed slightly contradicting picture that one of the most important aspects is not the sole moment of adoption of a program (as advocated by GQM) and possibilities of using subjective metrics, but the use of objective metrics to monitor entities over longer periods of time. A more detailed guidelines supporting the introduction of metric programs can be found in Goodman [11] or [12].
- framework presented by Diaz-Ley et al. [13] can be seen as suitable for smaller enterprises whereas the set of success factors and the framework from Ericsson [14] is targeted mainly for larger enterprises with a number of management levels. The main difference between the large and small-medium enterprises in the context of our work is the fact that the larger enterprises are organized using significantly more levels of management and multiple dimensions of management – e.g. project managers are usually not line managers.

One of the observed issues in program adoption is the reuse of measures. As Jorgensen [15] shows, this is not an easy task due to the potential different definitions of measures. Jorgensen shows contrasting definitions of measures if quality is defined as “a set of quality factors”, “user satisfaction”, and “software quality related to errors”. Our research recognizes the needs for viewing the same aspects (e.g. quality) from different perspectives – depending on the stakeholder.

These needs are also recognized by the Ericsson’s measurement team which we collaborated with.

The concept of a measurement system is not new in engineering or in software engineering – measurement instruments and systems are one of the cornerstones of engineering. In software engineering, we are used to working with metric tools rather than measurement systems. The difference is that metric tools and measurement instruments seem to be very similar, but metric tools and measurement systems are not. Measurement instruments (in other engineering disciplines) are suited for single purposes and usually collect one metric (e.g. voltage) whereas metric tools collect usually a number of metrics at the same time (e.g. length of the program, its complexity). Our framework is placed on top of metric tools with the focus on presenting calculating and presenting indicators rather than collecting metrics and is intended to be composed of multiple measurement instruments (metric tools). Other examples of measurement systems built in the same principles are:

- A measurement system presented by Wisell [16]: where the concept of using multiple measurement instruments to define a measurement system is also used widely at the studied organization.
- Computerized measurement systems in other disciplines facilitating the concept of measuring instruments, as presented in the following papers: [17–24]. All these measurement systems are (i) using the concept of measurement instruments, (ii) used in established engineering fields or physics, (iii) focused on monitoring current value of an attribute (status in our case) not on collecting metrics. Although differing in domains of applications these measurement systems show that concepts which the measurement team adopted from the international standards (like [25]) are successfully used in other engineering disciplines.
- Lawler and Kitchenham [7] present a generic way of modeling measures and building more advanced measures from less complex ones. Their work is linked to the TychoMetric [26] tool. The tool is a very powerful measure-

ment system framework, which has many advanced features not present in the Ericsson's framework (e.g. advanced ways of combining metrics). TychoMetric provides a possibility of setting up advanced and distributed (over several computers) filters and queries for multiple data sources as it is intended to cover all (or at least very many) kinds of metrics and projects.

3. Study Design

In our case study we study the measurement program at Ericsson where several measurement systems are used (over 200 at the time of studying). The concept of a measurement system has been adopted from the existing standards on metrology [25] where it is defined as a set of measuring instruments assembled in order to measure quantities of specific kinds. In the case of software engineering the quantities are dependent on the purpose of measurement and the measured entities. An entity can be a project, process, product, team, etc. and a quantity can be project length, number of activities in the process, lines-of-code in the product, team size, etc. The measurement systems built by the organization are developed according to the ISO/IEC 15939:2007 standard [27]. More details about the measurements are presented in subsection 3.2.

3.1. Sample

The sample in our study was chosen using convenience sampling with blocking: we asked experts with different roles:

- Stakeholder (1 person): A project manager for whom a measurement system was built. The project manager used the measurement system to monitor and control his project during the whole project execution.
- Manager (1 person): A section manager responsible for resources and competence.
- Quality manager (2 persons during 1 interview – i.e. 1 data point): Two quality managers working with measurement in the organization. They do not develop measurement

systems, but are involved in their design and evaluation.

- Designer of measurement systems/quality manager (1 person): A quality manager responsible for designing, developing, and maintaining measurement systems in the organization. This manager was the most insightful into the details of how measurement systems are structured and about their limitations.

These roles covered all persons involved in establishing, development, and maintenance of both measurement programs and measurement systems. All interviewees have several years of experience with working with measurements at Ericsson.

3.2. Objects

The study object in this case study is the measurement program at one of the units of Ericsson which develops large products for the mobile telephony network. The size of the organization is several hundred engineers and the size of the projects can be between 80 and 200 engineers. Projects are more and more often executed according to the principles of Agile software development and Lean production system referred to as Streamline development (SD) within Ericsson [28]. A noteworthy fact is that in SD the releases are frequent and that there is always a release-ready version of the system: referred to as Latest System Version [28]. This means that the measurement program used in the organization was designed to monitor and control software development on a continuous basis as opposed to controlling projects which have beginning and end. The streamline development also posed requirements on measures – they should guide the operation of the Streamline development programs towards improvements during the execution, i.e. without the possibility of doing post-mortem analyses or baselining towards previous projects.

The measurement program was a continuous activity for a number of years and was constantly improved. The last year, however, the organization succeeded in establishing the “measurement culture” in the organization and developed sev-

eral measurement systems according to ISO/IEC 15939 standard [27]. This standard contributed to establishing common measurement processes and vocabulary of indicators, base/derived measures, and information products. The studied organization complemented this standard with the ISO VIM (Vocabulary in Metrology, [25]) which contributed with the definitions and understanding of such concepts as measurement system, measuring instrument, base quantity, measurement process.

ISO/IEC 15939 was used to structure the measurement process at the studied organization and all documentation and information about it. In particular the web pages were named “Indicators”, “Base/derived measures”, “Measurement systems”, etc. This ambient use of ISO/IEC 15939 quickly resulted in spreading the vocabulary of the standard in the organization.

ISO VIM standard was used to structure the information within the measurement systems (i.e. MS Excel files) and to provide definitions of the concepts measured. When possible the measurement team also reused definitions from ISO/IEC 25000 series of standards (Software Quality Requirements and Evaluation) and ISO/IEC 9126 [29].

The goal of the measurement program was to constantly improve the operational excellence of the unit of Ericsson w.r.t. productivity, product and process quality and technology leadership. The measurement program was designed using the ISO/IEC 15939:2002 (and later using: 2007 edition) with the purpose to support stakeholders at multiple levels of organizations, for example:

- Project managers: to support them in monitoring the progress of the project and assisting them in addressing questions like “Will we finish on time?” or “How much resources do we need to maintain/improve the quality of the product?”
- Product managers/owners: to support them in monitoring and improving quality of products, i.e. assisting them in addressing questions like “How to achieve 0-defects at the release date?” or “Will we have good quality at <milestone>?”

- Line managers (at the section, department and unit level): to support them in monitoring the status of the organization and making long-term decisions about products, projects and competence in the organization, i.e. assisting them in addressing questions like: “Will we have enough resources to satisfy needs of <project X>?”

The measures used in the measurement program varied from management measures (e.g. financial) to technical (e.g. number of defects discovered during testing), and used at several levels of abstraction. We were able to study a number of measurement systems, e.g. measurement systems for:

- Measuring reliability of network products in operation for the manager of the product management organization; example measures in this measurement system are:
 - Product downtime per month in minutes
 - Number of nodes in operation
- Measuring project status and progress – for project managers who need to have daily updated information about such areas as requirements coverage in the project, test progress, costs, etc.; example measures in this measurement system are:
 - Number of work packages finished during the current week
 - Number of work packages planned to be finished during the current week
 - Number of test cases executed during the current week
 - Cost of the project up till the current date
- Measuring post-release defect inflow – for product managers who need to have weekly and monthly reports about the number of defects reported from products in field; examples of measures:
 - Number of defects reported from field operation of a product during the last month
 - Number of nodes in operation last month
 - Number of nodes which reported defects
- Summarizing status from several projects – for department manager who needs to have an overview of the status of all projects conducted in the organization, e.g. number of projects with all indicators “green”

These measurement systems were instantiated for a number of projects and products. Each of these instances had a distinct individual as stakeholder (in the role of project manager, product manager, etc.) who used the measurement system regularly.

Measures used in these measurement systems were both collected automatically from databases or manually from persons when the data is not stored in databases (e.g. by asking the project manager how many designers are assigned to remove defects from the software in a particular week, with detailed measures are described in [30]). The sources of information were defined in the measures specification and the infrastructure specification for the particular measurement systems (e.g. [31]).

The measures were designed using an in-house developed framework [32] based on the ISO/IEC 15939 standard. The framework was structured around the concepts of information product and indicator; the development of measurement systems started with discussions with stakeholders with two questions: “What do you need to know?” and “Why do you need to know it?” in the context of their management role. Model-Driven-Engineering approach was used when designing, implementing and validating measurement systems [31]. This approach has led to optimizing the number of data collected and the reduction from over 3000 measures to ca. 30 reusable (indicators).

The measurement program was built upon the concept of tools present in every desktop at the company – MS Office. Automated tools were built on top of MS Excel 2003 to collect data, perform measurements, store data, and present the most important information in form of indicators – all according to the ISO/IEC 15939:2007 standard. Detailed description of the technologies used behind this program are described by Staron and Meding [14].

3.3. Instruments

The main instrument used in our study was questionnaire which we used during the interviews. Another instrument was an interview with the

measurement systems designer/quality manager. The questionnaire was originally used by Jeffery and Berry [33] as a means of predicting the success of a measurement program in industry. The analysis of answers to these questions and a further interview result in identifying the main factors which determined successful implementation of measurement program, in a similar way as identifying the factors in other industrial case studies [34,35].

The questionnaires contained a list of questions; each of these was to be evaluated how well it was fulfilled. The evaluation was done by assigning a score on the scale 0–3, where 0 – this requirement is not fulfilled at all, 1 – this requirement is fulfilled to some extent, 2 – this requirement is fulfilled almost fully, and 3 – this requirements is completely fulfilled. This scale was according to the original questionnaire presented by Jeffery and Berry [33]. We modified the scale by adding N/A (Not Applicable) to the scale. An example question is presented in subsection 3.3.

We also added new questions, which were identified as factors important in successful implementation of measurement programs by [36]. All questions, including the ones added, were grouped according to the categories from the original paper [33]:

- Context(C) – questions about the background of the measurement program, the needs for it;
- Inputs(I) – questions about the input to the measurement program and its resources;
- Process – questions about the process of collecting measurements, process responsibilities and measure teams, with subcategories
 - Process motivation and objectives (PM),
 - Process responsibility and metrics team (PR),
 - Process and data collection (PC),
 - Process training and awareness (PT);
- Product (P) – questions about the measurements as products of the measurement process.

The full list of questions from the original questionnaire can be found in [33]. Our complete list of questions is presented below, and the

C1: Were the goals of the measurement program congruent with the goals of the business?				
0	1	2	3	N/A
()	()	()	()	()

Figure 1. Example question in the questionnaire

added questions are annotated with (A) before the question:

- C1: Were the goals of the measurement program congruent with the goals of the business?
- C2: Could the measured staff participate in the development of the measures?
- C3: Had a quality environment been established?
- C4: Were the processes stable?
- C5: Could the required granularity be determined and was the data available?
- C6: Was the measurement program tailored to the needs of the organization?
- C7: Was senior management commitment available?
- C8: Were the objectives and goals clearly stated?
- C9: Were there realistic assessments of pay-back period (e.g. 2 years)?
- (A) C10: Was the process planned to be incrementally implemented?
- I1: Was the program resourced properly?
- I2: Were resources allocated to training?
- I3: Were at least three people assigned to the measurement program?
- I4: Was research done?
- (A) I5: Were existing metrics materials used?
- (A) I6: Was the data seen to have integrity?
- (A) I7: Was the data easy to collect collected?
- (A) I8: Was the data set determined incrementally?
- P1: Were the measures clear and of obvious applicability?
- P2: Did the end result provide clear benefits to the management process at the chosen management audience levels?
- P3: Was feedback on results provided to those being measured?
- P4: Was the measurement system flexible enough to allow for the addition of new techniques?
- P5: Were measures used only for pre-defined objectives?
- PC1: Were the important initial metrics defined?
- PC2: Were tools for automatic data collection and analysis developed?
- PC3: Was a metric database created?
- PC4: Was there a mechanism for changing the measurement system in an orderly way?
- PC5: Was measurement integrated into the process?
- PC6: Were capabilities provided for users to explain events and phenomena associated with the project?
- PC7: Was the data cleaned and used promptly?
- PC8: Did the objective determine the measures?
- (A) PC9: Was the measurement program constantly improved?
- PM1: Was the program promoted through the publication of success stories and encouraging exchange of ideas?
- PM2: Was a firm implementation plan published?
- PM3: Was the program used to assess the individuals?
- PR1: Was the metrics team independent of software developers?
- PR2: Were clear responsibility assigned?
- PR3: Was the initial collection of metrics sold to data collectors?
- PT1: Was adequate training in software metrics carried out?
- PT2: Did everyone know what was being measured and why?

The interviewees were not presented with additional material during the interview, as they understood the measurement program and had extensive experience with it.

As an addition to the questionnaire, we send a question to the designer of measurement systems/quality manager before the interview in order not to influence his answers by the questions in the questionnaire. The question was: “What are the most important factors that determine whether a measurement system is successfully implemented and used in the organization?” We deliberately narrowed the question to measurement system as we wanted to obtain information which covered the issues not addressed by the questionnaire.

In the end we performed also a workshop with the quality managers, section manager, and designer of measurement systems/quality manager where we presented the results and validated our findings.

3.4. Analysis Methods

In the study we use descriptive statistics when analysing the results from the questionnaires. We provide a total percentage of score for each category from Section 3.3. The max score (i.e. 100%) is when all applicable questions are ranked as 3 (requirements are completely fulfilled) by all stakeholders (i.e. $3 * 4 = 12$, and 12 is the 100% score for each questions applicable for all stakeholders). We do not account for non-equal variances in the descriptive statistics as we do not perform hypotheses testing methods that would require doing so.

To test for significant differences between roles, we use also the Friedman test [37]. Our hypotheses are:

- $H0$: There is no difference between roles.
- $H1$: There is a difference between roles.

Testing these hypotheses allows for assessing whether the different respondents perceived (assessed) the measurement program differently, or whether there is a consensus on how the program is implemented.

In order to further test for which questions the respondents were uniform and for which their

answers were disperse, we use the hierarchical cluster analysis for between-variable (roles) and between-treatment (questions) clusters [38]. We use dendrograms for visualizing the results.

Using the cluster analysis provided us with the statistical means of suggesting groups of success factors. The suggested groups were then evaluated together with the study subjects whether they should be grouped into a more compound success factor.

4. Results and Analysis

The results are presented in the following parts: (i) results from questionnaires, (ii) success factors identified by the designer of measurement systems/quality manager, and (iii) the list of success factors identified and generalized from both (i) and (ii).

4.1. Questionnaire Results

The percentage of requirements fulfilled for each category is presented in Table 1.

Table 1. Percentage of requirements fulfilled

Category	Number of questions	Score
Context	10	79%
Input	8	80%
Process	17	64%
Product	5	76%

The table shows that the input and context are categories with the requirements fulfilled to the largest extent. The process is the category with requirements fulfilled to the least extent. This seems to be natural as the organization and its measurement program constantly evolves, and so do the measurement processes. The summarizing descriptive statistics per respondent are presented in Table 2.

The descriptive statistics show that stakeholder was the most positive respondent, which was a desired effect (since the “survival” of the measurement program depends on stakeholders using the measurement systems). After the presentation of these results the designer of measure-

Table 2. Descriptive statistics per respondent

Respondent	Median	Number of 3's	Number of 0's
Designer of measurement systems/Quality manager	2	19	7
Quality manger	2	14	6
Stakeholder	3	25	2
Section manager	3	21	1

ment systems/quality manager provided us with feedback on his low assessment results. The results were caused by the designer of measurement system/quality manager having a complete picture of the further work to improve the existing measurement program in the company.

The Friedman test resulted in rejecting the null hypothesis with the p-value of 0.00042. With the total number of questions over 30, the β -value was below 0.05. Having rejected the null hypothesis we can conclude that the respondents had different view on the measurement program and perform the hierarchical cluster analysis.

The hierarchical cluster analysis for between-variables (roles) clusters results in the dendrogram presented in Figure 2.

The dendrogram shows that the quality manger(s) and the section manager have the most similar opinions. The stakeholder's opinion was the least similar to the rest of the respondents. A closer analysis (indicated in Table 2) showed that the stakeholder was more positive than other respondents to the measurement program and its fulfilment of requirements. This, in turn indicated that the organization was successful in spreading the measurement systems and establishing the measurement program.

The hierarchical cluster analysis for between-treatments (questions) cluster results in the dendrogram presented in Figure 3.

The results show that there are questions where the different respondents do not agree – e.g. question 21. After a closer analysis we found that these are the questions about aspects not familiar to some of the respondents – e.g. stakeholder (project manager) was not aware that we have a large metrics database. An example of a group of questions where the respondents agreed is: PR1, PR2, C5, C6, I5, I6, and I8 (in the middle of the figure). A closer analysis revealed that these were the questions which scored 3 (the top rank) by all stakeholders.

4.2. Measurement Systems Designer's Perception: Success Factors

The list of factors which are identified as important by the designer concerned the way in which measurement systems are developed and deployed in the organization. These factors were not added to the questionnaire, because they were at a much lower level than the questionnaire – they concerned technical aspects of building measurement systems and measuring instruments rather than establishing a measurement program in the organization.

The measurement systems designer/quality manager identified the following factors (without prioritizing them):

1. Work according to the standards (also identified in [39]), which is important as it ensures that:
 - a) all measurement systems are built and presented in the same way,
 - b) there is a well known nomenclature regarding measurement systems,
 - c) all steps regarding building and maintaining of measurement systems are well defined,
 - d) ISO/IEC 15939 is a very solid standard that is recommended for Software Engineering.
2. Always providing certain base measures, e.g. defect statistics for projects and products.
 - a) Using standards like ISO/IEC 25000 (SQUARE) is recommended.
3. Definition and use of a known process to get information about all main elements of a measurement system (e.g. stakeholder, information need, indicators). In particular there should always be a stakeholder for the measurement system.
 - a) The stakeholder should have a real and legitimate power in the project – e.g. project

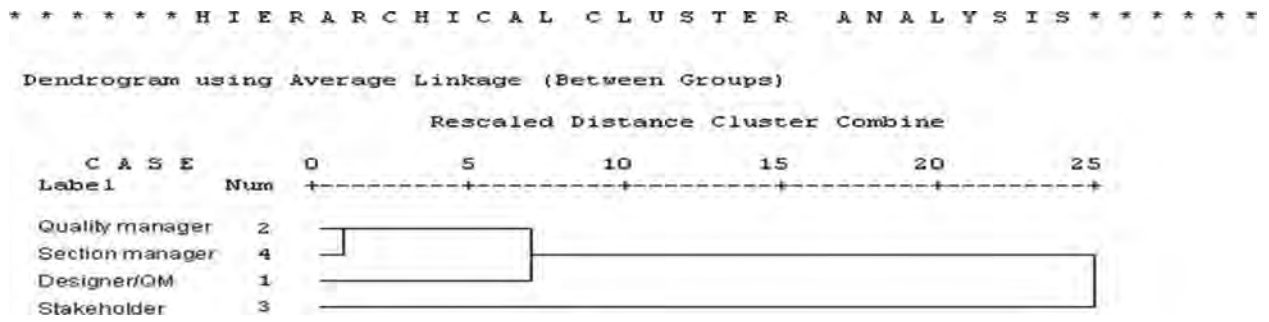


Figure 2. Dendrogram for between-variables clusters

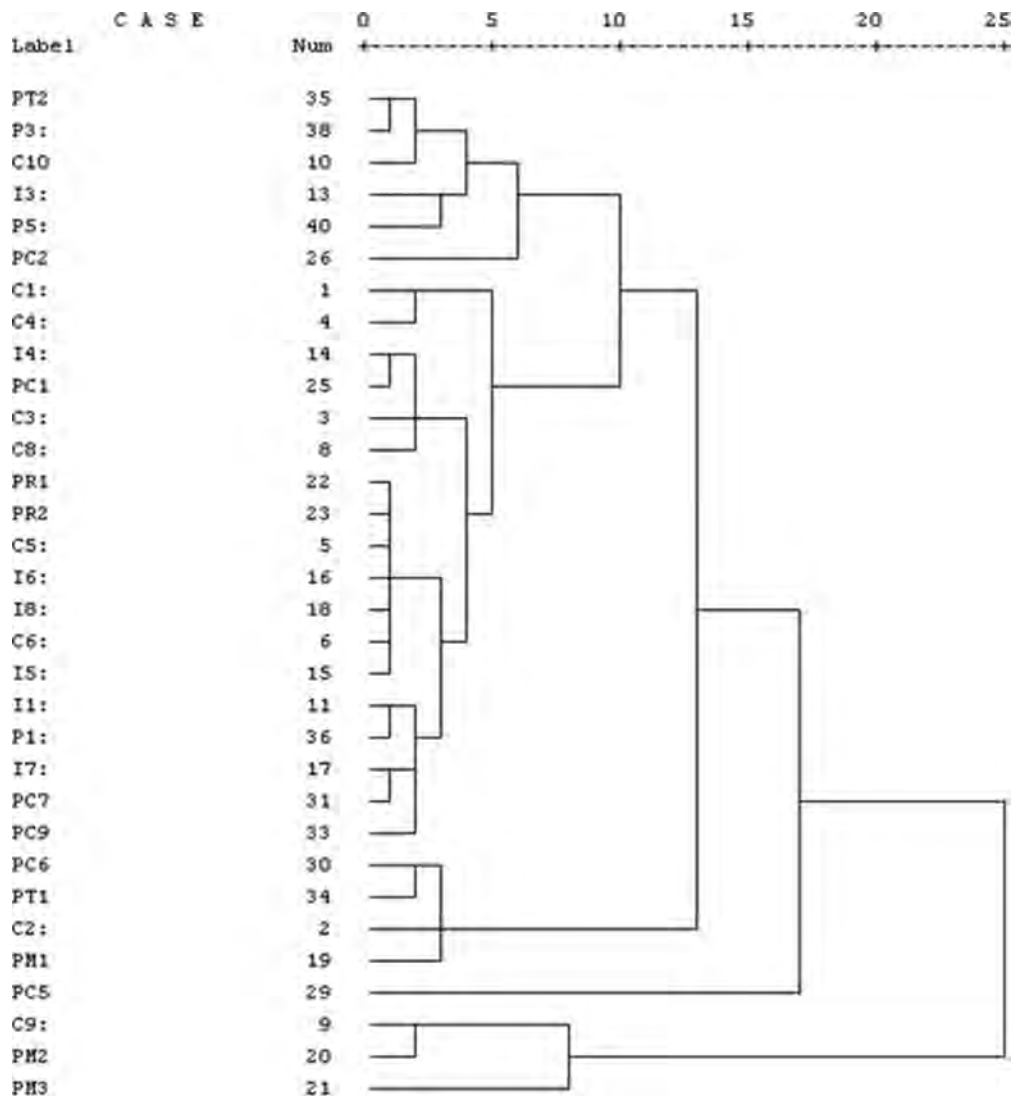


Figure 3. Dendrogram for between-treatment clusters

- manager or section manager. Otherwise there is a real risk of waste, i.e. measurements are not used for decision making.
4. Specify and implement measurement systems in a constant way, e.g. logical and physical views of architectural design, implementation technology, and/or knowledge base.
 - a) Maintain the infrastructure and measures so that it can be deployed on large scale.
 5. Use pre-defined infrastructure and allocated areas for storing measures and information about the measures (define the measurements database).
 - a) It is important to keep the values of measures for future use and future analyses. Using simple databases with structure of information in accordance to ISO/IEC 15939 is recommended.
 6. Present the main information (e.g. indicators) in a simple, non-ambiguous, and succinct manner.
 - a) Present details in another place, which is linked from the main information presentation.
 - b) Gadget in MS Windows Vista/7 or Widgets for MacOS are recommended since they provide the stakeholders with information without the need for them to be active (for example, please see [30]).
 7. Ensure reliability of the measurement system – provided information should be reliable and up-to-date.
 - a) We recommend using indicators of information quality [40].
 8. Ensure that the necessary knowledge is in place (for details see also [14]).
 - a) Stakeholders should know how to interpret the information and make adjustments to measurement systems.
 - b) Designers of measurement systems should know the standards and implementation technology for the measurement systems.

The above factors are related to how measurement systems are built and deployed in the organization. They have an effect on the measurement program, to which other factors apply as well.

4.3. Success Factors

In this section we focus on the factors, which have not been identified previously, and do not re-consider the importance of such factors as:

- Management commitment [6]: Measurement program as a “shadow” activity of employees without management support stand no chances of success as it is the managers who decide whether new methods/tools/ways of working are introduced or not. When we designed the first measurement systems the commitment was rather hard to obtain. The turning point came when we showed the results of our predictions to one of the project managers and his response was “If these predictions are correct, then we cannot let this happen”; this was followed by his actions to adjust resources and avoiding problems in the project. This first “success” helped us to get strong commitment from the project manager and in turn (gradually) from other project managers and line managers.
- Team commitment [6]: Without the commitment from the team being measured the information quality might be low, which jeopardizes the reliability of the data. In the case of the studied organization the team commitment was obtained after about 1 year of using measurement systems for making decisions for one project. The team has realized that the measurements help them to visualize the goal and achieve it.
- Making measurements part of processes [41]: Putting new burdens on persons in the organization is never popular and should be avoided. It is much better to use “probes” which measure in-process data from the tools already used at the organization. This minimizes the threat that other activities are prioritized over measuring for the persons being measured. In our case this was reduced by using automation based on MS Excel. Since everyone in the organization knew MS Excel virtually no learning was involved; automation reduced even the burden of processing and presenting the information (see [14]).

We see the above factors being prerequisites for a successful program and these factors were present in the studied organization. What we have observed in the organization was the gradual (over ca. 2 years) change of culture. The concept of “main measures” was discussed in the organization at the beginning whereas in the end only the indicators were considered.

Table 3 presents factors which we identified as important when implementing measurement programs when performing the program evaluation at Ericsson. These factors are important for different roles, which is indicated by a cross in the column denoting particular stakeholder (D/QM – Designer/Quality manager; QM – Quality manager; SH – Stakeholder, SM – Section manager).

The above factors have already been identified and they are mostly related to the process of establishing the measurement program. After being established, the program needs to be maintained in order not to be dropped. Therefore we identify the following:

1. *Working according to the ISO/IEC 15939 standard:* A standardized nomenclature (ISO/IEC 15939 [27] and ISO/IEC Vocabulary on Metrology [25]), terminology and proven processes are key factors in the long-term adoption. Using standards make the effort less person-dependent and interpretation dependent. It makes reuse across organizations easier, as also indicated in [43]. In our case we follow: ISO/IEC 15939:2007, ISO/IEC Vocabulary on Metrology, and ISO/IEC 9126.
2. *Providing information quality indicators:* Information is as good as it is reliable and up-to-date. Providing information, especially automatically should also indicate the quality of the information provided. An existing model can be used (e.g. [44, 45]) or a dedicated one can be developed. The issues to address when indicating information quality are: providing the data which is up-to-date, correctly processed, complete, and unbiased. In our work we use the following indicators of information quality:
 - a) Timeliness (the information presented to the stakeholder is up-to-date, e.g. from today, this month, or current – depending on the purpose),
 - b) Completeness (the information contains no missing values),
 - c) Correctness (there were no errors in calculation),
 - d) Accuracy (the data sources contain the updated information).
3. *Automated data collection based on simple software tools* (also identified in [46]): measures should be collected automatically to minimize the burden of data collection to the (usually) already busy organization. If not automated the program will eventually be rejected. In our work we use MS Excel and Visual Basic for Applications to automate the data collection and processing. By developing measurement systems, the organization gains competence on working with measures and does not rely on external entities when building and maintaining the measures.
4. *Individual stakeholders for each measurement system:* (related to “Use in decision making” from [6]): there is one role/individual in the organization whose information need is satisfied with the measurement system (a.k.a. producing data inside their range of validity as identified in [46]; identified also in [47] as using different strokes for different people). If this is not the case, then the measurements are not used in the decision process and thus become ineffective. Stakeholders should be able to adjust the measurements to the situations that can happen over time (e.g. by adjusting decision criteria for indicators).
5. *Direct benefits to the organization:* The results from the measurement program should be applicable in the organization “now” and not after a period of time. The most current activities are usually prioritized, and benefiting from measures in decision process depends on using current data to satisfy current information needs.
6. *Devoted measurement team:* the measurements are collected throughout the organization, but there is a team of specialists who help to define and introduce measurements. These specialists are also responsi-

Table 3. Factors important for long-term success identified in our study

Factor	D/QM	QM	SH	SM
Congruence of measurement goals with business goals	x	x		
Incremental implementation of the program	x			x
Participation of measured staff in program development	x		x	
Quality environment	x		x	
Process stability	x			
Availability of data at the required granularity	x	x	x	x
Tailoring measurements to organization needs	x	x	x	x
Clear objectives and goals (also in [1])	x		x	x
Proper program resourcing (proper metric team)		x	x	
Conducting research prior to/during measurement program development	x		x	x
Using existing metric materials	x	x	x	x
Integrity of the data	x	x	x	x
Using existing data for processing		x	x	x
Data set determined incrementally	x	x	x	x
Clear measures of obvious applicability			x	x
Clear benefits for the management process	x			
Providing feedback to those being measured			x	
Flexibility for adding new measurements (also in [1])	x	x		x
Pre-defined objectives for the measures			x	
Initial definition of important metrics (also in [42])	x		x	x
Automatic data collection and processing (also in [1, 42])	x	x	x	x
Metric database	x	x		x
Mechanisms for adjusting measurement systems to changing needs	x			
Integrating measurement into the process	x			
Stakeholders are able to explain the meaning of metrics values	x		x	
Using data in clean and prompt way (also in [42])	x	x	x	x
Measures are determined by objectives	x		x	x
Constant improvement of the measurement program	x	x	x	x
Independent metric team (from developers)	x	x	x	x
Clear assignment of responsibilities	x	x	x	x
Adequate training in software metrics	x		x	

ble for maintenance of the measurement program. Evidence of such a team being a positive factor has also been found when introducing modelling into large organizations [48], which, although seems unrelated, is similar to introducing measures (as a new way of working). In the case of the studied organization the measurement team consists of quality managers, section managers, technology specialists and researchers – which is similar to the team of specialists when introducing models – modelling specialists, technology specialists and researchers.

7. *Measurement collection effort should be minimal:* (also identified in [46, 47]), which means that using already collected data (at least

initially) is a good point. Every organization collects data from their processes (e.g. such high level data as project cost), and such data should be used when the measurement program is being established to show that measurement programs provide positive support. After the measurement program has been adopted, the measures should be refined to optimize the data collection and fulfilment of stakeholders' information needs.

8. *Providing standard base measures:* Certain base measures, e.g. product performance, should always be provided if applicable to support benchmarking and reuse. However, the number of measures provided in this way should be optimized w.r.t. needs and costs for collecting them. Example base mea-

asures provided in the studied organization are: In-Service-Performance, resource allocation, number of work packages completed.

9. *Reusing base measures*: Costs of measures collection should be optimized and measures should be aimed at being reused. Therefore the measures should be specified, described, and stored in repositories which would enable reusing them – e.g. for benchmarking or measuring improvements over time.
10. *Using measures specifications and specification of their instantiation*: The measures are specified in relation to the kinds of measured entities – e.g. measures of a project (one single project), like number of designers. These measures are then instantiated for different projects. The distinction is important since measurement systems might be different because they measure different projects (entities) or measure projects and processes (different kinds of entities).
11. *Do not use the program to assess individuals*: It is important not to create negative attitude to the program (a.k.a. Fear of adverse consequences in [1,47]) by creating situations that measurements are to assess the work/performance of individuals.

The above factors are ordered according to their importance – factors 1 being the most important one.

5. Validity Evaluation

We identify the threats to validity of our study using the categories presented by Wohlin et al. [49].

The main external validity threat of our study is the fact that we studied only a single organization. However, the found success factors are consistent with the trends observed in literature and do not seem to be organization or process specific. The underlying technology for implementing automation is based on MS Excel which is used in almost every company and is not an Ericsson-specific tool. The add-ons for Excel with measurement instruments are specific, but these do not influence the generalizability of the results.

The main construct validity threat is related to mono-operation bias, which is a bias introduced by observing a single phenomenon at a single point of time and thus not capturing the full breadth of the phenomenon. This is a typical threat to operationalizations in single-case case studies. Our research is a summary of a 2 year action research project research and the respondents in the study were involved in measurement activities for a number of years.

The main threat to the internal validity of our findings is the maturation effect as it was a 2 year project. Naturally this is a threat, but to some extent the maturity effect is desired in studies like this. The primary goal of our action research project was not to observe whether the measurement program was correct, but to establish and maintain a measurement program. In this manner, the maturity effect is a desired “cultural change” effect in the organization.

Finally, the main threat to conclusion validity is related to the fact that we have not used grounded theory to analyze interview material, but rather asked direct questions to the respondents and the interviewee. It was a deliberate choice since the authors were part of the team establishing the measurement program and we had this opportunity to reduce the “noise” in the interview data by asking direct questions and using experience to reason about the answers. We use the statistical analysis when possible to evaluate the significance of some of the claims we made.

6. Conclusions

Software development projects are entities where change is prevalent and constant adaptations are predominant – especially if the projects are to meet their goals and deliver quality software. A long-term success of a measurement program requires its constant adaptation towards the change in software projects, a situation unlike in manufacturing industries. The studied organization has chosen not to use GQM in order to be more flexible when adopting their measurement program and take advantage of adjusting

interpretations of measures (embedded in the concept of indicator) and to be able to combine the ISO/IEC 15939 standard with measurement theory from other engineering disciplines. The decision to remain independent from tool vendors and do not purchase off-the-shelf solution provided the organization with ability to remain the core measurement competence in-house, and hence be more reactive to changing needs of the organization.

The organization combined three key elements when establishing and maintaining the measurement program: the use of international standards, significant experience base, and research activities. This combination contributed to the success of a measurement program constantly grows in the organization. By including researchers in the process of developing, establishing, and maintaining both the measurement program and the measurement systems, the company benefited from external competence, but did not rely on external entities to establish the program. This elevated the competence of the measurement team and resulted in publications related to measures, e.g. [50].

In this paper we described factors contributing positively to the success of a long-term measurement program. These factors are based on the experience of the team working with the measurement program and have been obtained through interviews and surveys.

Our further work is focused on observing threats to the working measurement program and identifying these threats over a longer period of time (at least 3 years). Identifying such threats would help to prevent withdrawing from the measurement program in the organizations.

7. Acknowledgements

The project has been partially sponsored by the Swedish Strategic Research Foundation (www.stratresearch.se) under the program Mobility in IT. It was also partially sponsored by Ericsson Software Architecture Quality Center and Ericsson Software Research.

We would like to thank the involved managers and engineers at Ericsson for their support in this study.

References

- [1] M. Umarji and H. Emurian, "Acceptance issues in metrics program implementation," in *11th IEEE International Symposium Software Metrics*, H. Emurian, Ed., 2005, pp. 10–29.
- [2] K. L. Atkins, B. D. Martin, J. M. Vellinga, and R. A. Price, "STARDUST: implementing a new manage-to-budget paradigm," *Acta Astronautica*, Vol. 52, No. 2–6, pp. 87–97, 2003.
- [3] T. Kilpi, "Implementing a software metrics program at Nokia," *IEEE Software*, Vol. 18, No. 6, pp. 72–77, 2001.
- [4] B. Clark, "Eight secrets of software measurement," *IEEE Software*, Vol. 19, No. 5, pp. 12–14, 2002.
- [5] A. Gopal, T. Mukhopadhyay, and M. S. Krishnan, "The impact of institutional forces on software metrics programs," *IEEE Transactions on Software Engineering*, Vol. 31, No. 8, pp. 679–694, 2005.
- [6] A. Gopal, M. S. Krishnan, T. Mukhopadhyay, and D. R. Goldenson, "Measurement programs in software development: determinants of success," *IEEE Transactions on Software Engineering*, Vol. 28, No. 9, pp. 863–875, 2002.
- [7] J. Lawler and B. Kitchenham, "Measurement modeling technology," *IEEE Software*, Vol. 20, No. 3, pp. 68–75, 2003.
- [8] F. Niessink and H. van Vliet, "Measurement program success factors revisited," *Information and Software Technology*, Vol. 43, No. 10, pp. 617–628, 2001.
- [9] —, "Measurements should generate value, rather than data," in *6th International Software Metrics Symposium*, 2000, pp. 31–38.
- [10] S. De Panfilis, B. Kitchenham, and N. Morfuni, "Experiences introducing a measurement program," *Information and Software Technology*, Vol. 39, No. 11, pp. 745–754, 1997.
- [11] P. Goodman, *Practical implementation of software metrics*, ser. International software quality assurance series. London: McGraw-Hill, 1993.
- [12] K. H. Moeller, *Software metrics: a practitioner's guide to improved product development*. London: Chapman-Hall, 1993.
- [13] M. Diaz-Ley, F. Garcia, and M. Piattini, "Implementing a software measurement program in small and medium enterprises: a suitable frame-

- work,” *IET Software*, Vol. 2, No. 5, pp. 417–436, 2008.
- [14] M. Staron, W. Meding, and C. Nilsson, “A framework for developing measurement systems and its industrial evaluation,” *Information and Software Technology*, Vol. 51, No. 4, pp. 721–737, 2008.
- [15] M. Jorgensen, “Software quality measurement,” *Advances in Engineering Software*, Vol. 30, No. 12, pp. 907–912, 1999.
- [16] D. Wisell, P. Stenvard, A. Hansebacke, and N. Kesitalo, “Considerations when designing and using virtual instruments as building blocks in flexible measurement system solutions,” in *IEEE Instrumentation and Measurement Technology Conference*, P. Stenvard, Ed., 2007, pp. 1–5.
- [17] A. Sehmi, N. Jones, S. Wang, and G. Loudon, “Knowledge-based systems for neuroelectric signal processing,” *IEE Proceedings-Science, Measurement and Technology*, Vol. 141, No. 3, pp. 215–23, 2003.
- [18] J. Feigin and K. Pahlavan, “Measurement of characteristics of voice over IP in a wireless LAN environment,” in *IEEE International Workshop on Mobile Multimedia Communications*, K. Pahlavan, Ed., 2003, pp. 236–240.
- [19] M. Foote and D. Horn, “Video measurement of swash zone hydrodynamics,” *Geomorphology*, Vol. 29, No. 1–2, pp. 59–76, 1999.
- [20] N. P. Kolev, S. T. Yordanova, and P. M. Tzvetkov, “Computerized investigation of robust measurement systems,” *Instrumentation and Measurement, IEEE Transactions on*, Vol. 51, No. 2, pp. 207–210, 2002.
- [21] R. F. Kunz, G. F. Kasmala, J. H. Mahaffy, and C. J. Murray, “On the automated assessment of nuclear reactor systems code accuracy,” *Nuclear Engineering and Design*, Vol. 211, No. 2–3, pp. 245–272, 2002.
- [22] A. N. Zaborovsky, D. O. Danilov, G. V. Leonov, and R. V. Mescheriakov, “Software and hardware for measurements systems,” in *The IEEE-Siberian Conference on Electron Devices and Materials*, D. O. Danilov, Ed. IEEE, 2007, pp. 53–57.
- [23] H. Zhiyao, W. Baoliang, and L. Haiqing, “An intelligent measurement system for powder flowrate measurement in pneumatic conveying system,” *IEEE Transactions on Instrumentation and Measurement*, Vol. 51, No. 4, pp. 700–703, 2002.
- [24] G. Kai, “Virtual measurement system for muzzle velocity and firing frequency,” in *8th International Conference on Electronic Measurement and Instruments*, 2001, pp. 176–179.
- [25] *International vocabulary of basic and general terms in metrology = Vocabulaire international des termes fondamentaux et généraux de métrologie*, 2nd ed. Genève, Switzerland: International Organization for Standardization, 1993.
- [26] “TychoMetrics,” Predicate Logic, 2007.
- [27] *ISO/IEC 15939:2007 Systems and software engineering – Measurement process*, International Standard Organization and International Electrotechnical Commission Std., 2007.
- [28] P. Tomaszewski, P. Berander, and L.-O. Damm, “From traditional to streamline development – opportunities and challenges,” *Software Process Improvement and Practice*, Vol. 2007, No. 1, pp. 1–20, 2007.
- [29] *ISO IEC 9126, Software engineering, Product quality Part: 1. Quality model*, International Standard Organization / International Electrotechnical Commission Std., 2001.
- [30] M. Staron and W. Meding, “Defect inflow prediction in large software projects,” *e-Informatica Software Engineering Journal*, Vol. 4, No. 1, pp. 1–23, 2010.
- [31] —, “Using models to develop measurement systems: A method and its industrial use,” Vol. 5891, pp. 212–226, 2009.
- [32] M. Staron, W. Meding, G. Karlsson, and C. Nilsson, “Developing measurement systems: an industrial case study,” *Journal of Software Maintenance and Evolution: Research and Practice*, 2010.
- [33] R. Jeffery and M. Berry, “A framework for evaluation and prediction of metrics program success,” pp. 28–39, 1993.
- [34] M. Staron, L. Kuzniarz, and L. Wallin, “Factors determining effective realization of MDA in industry,” in *2nd Nordic Workshop on the Unified Modeling Language*, K. Koskimies, L. Kuzniarz, J. Lilius, and I. Porres, Eds., Vol. 35. Abo Akademi, 2004, pp. 79–91.
- [35] —, “A case study on industrial MDA realization – determinants of effectiveness,” *Nordic Journal of Computing*, Vol. 11, No. 3, pp. 254–278, 2004.
- [36] T. Hall and N. Fenton, “Implementing effective software metrics programs,” *Software, IEEE*, Vol. 14, No. 2, pp. 55–65, 1997.

- [37] D. Altman, *Practical Statistics for Medical Research*. Chapman-Hall, 1991.
- [38] D. F. Morrison, *Multivariate statistical methods*, 3rd ed., ser. McGraw-Hill series in probability and statistics. New York: McGraw-Hill, 1990.
- [39] L. Westfall, "Are we doing well, or are we doing poorly?" p. 20, 2005.
- [40] M. Staron and W. Meding, "Ensuring reliability of information provided by measurement systems," Vol. 5891, pp. 1–16, 2009.
- [41] C. A. Dekkers and P. A. McQuaid, "The dangers of using software metrics to (mis)manage," *IT Professional*, Vol. 4, No. 2, pp. 24–30, 2002.
- [42] J. D. Herbsleb and R. E. Grinter, "Conceptual simplicity meets organizational complexity: case study of a corporate metrics program," in *Software Engineering, 1998. Proceedings of the 1998 International Conference on*, R. E. Grinter, Ed., 2003, pp. 271–280.
- [43] F. Garcia, M. F. Bertoa, C. Calero, A. Vallecillo, F. Ruiz, M. Piattini, and M. Genero, "Towards a consistent terminology for software measurement," *Information and Software Technology*, Vol. 48, No. 8, pp. 631–644, 2006.
- [44] Z. R. Pendic, L. Kovacevic, and J. Stupar, "An approach to evaluation of quality of integrated information systems," *Annual Review in Automatic Programming*, Vol. 14, No. Part 2, pp. 63–68, 1988.
- [45] B. K. Kahn, D. M. Strong, and R. Y. Wang, "Information quality benchmarks: Product and service performance," *Communications of the ACM*, Vol. 45, No. 5, pp. 184–192, 2002.
- [46] F. J. Buckley, "Standards-establishing a standard metrics program," *Computer*, Vol. 23, No. 6, pp. 85–86, 1990.
- [47] S. L. Pfleeger, "Lessons learned in building a corporate metrics program," *Software, IEEE*, Vol. 10, No. 3, pp. 67–74, 1993.
- [48] P. Baker, S. Loh, and F. Weil, "Model-driven engineering in a large industrial context – a motorola case study," in *Model Driven Engineering Languages and Systems*, ser. Lecture Notes in Computer Science, Vol. 3713, 2002, pp. 476–491.
- [49] C. Wohlin, P. Runeson, M. Host, M. C. Ohlsson, B. Regnell, and A. Wesslèn, *Experimentation in Software Engineering: An Introduction*. Boston MA: Kluwer Academic Publisher, 2000.
- [50] M. Staron and W. Meding, "Predicting weekly defect inflow in large software projects based on project planning and test status," *Information and Software Technology*, 2007, (available online).

