D. Karlström, P. Runeson and C. Wohlin, "Aggregating Viewpoints for Strategic Software Process Improvement - A Method and a Case Study", IEE Proceedings - Software Engineering, Vol. 149, No. 5, pp. 143-152, 2002.

Aggregating Viewpoints for Strategic Software Process Improvement – a Method and a Case Study

Daniel Karlström
Daniel.Karlstrom@telecom.lth.se

Dept. Communication Systems,
Lund University,
Box 118, SE-221 00 Lund,
Sweden.

Per Runeson
Per.Runeson@telecom.lth.se

Dept. Communication Systems,
Lund University,
Box 118, SE-221 00 Lund,
Sweden.

Claes Wohlin
Claes.Wohlin@bth.se

Dept. Software Eng. & Comp. Sc. ,
Blekinge Institute of Technology,
Box 520, SE-372 25 Ronneby,
Sweden.

Abstract

Decisions regarding strategic software process improvement (SPI) are generally based on the management's viewpoint of the situation, and in some cases also the viewpoints of some kind of an SPI group. This may result in strategies, which are not accepted throughout the organisation, as the views of how the process is functioning are different throughout the company. This paper describes a method for identifying the major factors affecting a process improvement goal and how the perception of the importance of the factors varies throughout the organisation. The method lets individuals from the whole development organisation rate the expected effect of these factors from their own viewpoint. In this way the strategic SPI decision can be taken using input from the entire organisation, and any discrepancies in the ratings can also give important SPI decision information.

The method is applied in a case study performed at Fuji Xerox, Tokyo, which is reported in this paper. In the case study, significantly different profiles of the factor ratings came from management compared to the ones from the engineering staff. This result can be used to support the strategy decision as such, but also to anchor the decision in the organisation.

1 Introduction

The need for software process improvement (SPI) is well known and also increasingly accepted as a means for a software company to stay competitive in a rapidly changing environment. However, when it comes to which strategy to follow in an SPI program, there are many different opinions in a software organisation. For example, quality assurance staff may tend to stress the importance of the documented processes, while engineers may tend to rely on better tools. One of the contributing factors to such differences is that the documented process is not the same as the actual work performed. Roles and workflows are documented in official descriptions, but the actual roles and work performed are a combination of past practices, each person's interpretation of the official documents and the effects of training programmes.

Decisions made to improve the development processes as a part of a software process improvement initiative, are in most cases taken by a select group of people with their specific views of the entire process. If the different viewpoints in the organisation are not taken into account, there is a risk that the SPI initiative is not sufficiently accepted, and hence the goals will not be effectively implemented.

After the need for an improvement effort is realised and accepted by the company, an initial phase begins when sponsorship, goals and strategies are confirmed for the duration of the improvement effort. The current status of the company's development must then be measured or assessed before any improvement attempts can be made. The steps to be taken for SPI as described by Humphrey [1] can be generalised into the steps shown in Figure 1. These are also main steps of the more advanced SPI method called the IDEAL process [2] introduced by SEI.

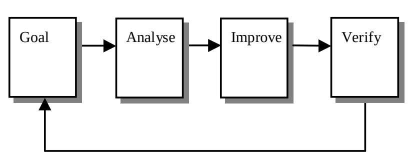

Figure 1: General SPI model.

This paper focuses on the goal and analysis part of the SPI model shown in Figure 1 It presents a method which supports SPI decisions to be taken with knowledge about the different interpretations that each person has made of the current status of the development process and their own role in the process.

The method is divided into five main steps, for which a flow diagram is provided in Figure 2.

A: Determine goal.
B: Identify major factors believed to be affecting the goal.
C: Allow the organisation to rate the factors anonymously based on their personal viewpoints.
D: Analysis of data. Aggregate and analyse the rating results and identify discrepancies.
E: SPI strategy decision.

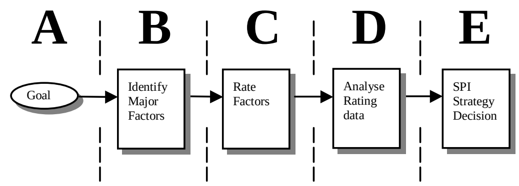

Figure 2: Overview of the method.

The method provides decision support information to SPI responsible individuals by identifying the perceived most effective improvement factors and by analysing discrepancies between different viewpoints throughout the organisation. The advantage is that it provides systematically derived information, which supports decision makers in the decisions on what to improve and how to improve it. Metric analysis can be one method of identifying improvement factors. A general process framework/appraisal method such as the CMM [2] is another. However, factors outside these frameworks may be just as important since they are derived from the individual company. Hence, they should most definitely not be excluded.

The paper is structured as follows: Section 2 presents the general problem of how processes are interpreted from different viewpoints of different persons in the organisation. In Section 3, the method and its constituents are presented and in Section 4, a case study of the application of the method is presented. The case study was performed at Fuji Xerox in Tokyo, Japan.

2 Processes Distortion and Viewpoints

In a process improvement initiative, there may be different opinions on what should be improved in order to reach the improvement goal. To some extent, this depends on the different viewpoints on what the process actually is.

Development processes are documented in official documents, but as they are interpreted, different versions appear depending on the perception of the process. Bandinelli et al. describe these different perceptions of the process and names these as the *official, perceived* and *desired* process [3]. To this list we can add the process that is actually performed, the *actual* process and the observations made by SPI responsible people, the *observed* process. The relations between the processes are illustrated in Figure 3.

From	**To**	**Mechanism**	**Description**
Reference	Desired	Comprehension	The reference process is interpreted by the person(s) creating the official process.
Desired	Official	Formulating	The official process is formulated from the desired process and written down as the official process.
Official	Perceived	Comprehension	The user based on what he is taught, what he reads and how his co-workers influence him interprets the official process.
Perceived	Actual	Performing	What is actually performed is a result of the perceived new process and the old process and is also dependant on individual opinions, mistakes and external stimuli.
Actual	Observed	Comprehension	The observed process is a result of an interpretation of the actual process.

Table 1: Mechanisms of transfer between types of processes.

Note that of all the processes shown in the figure, only the actual process produces the product and the only absolutely accessible process is the official process.

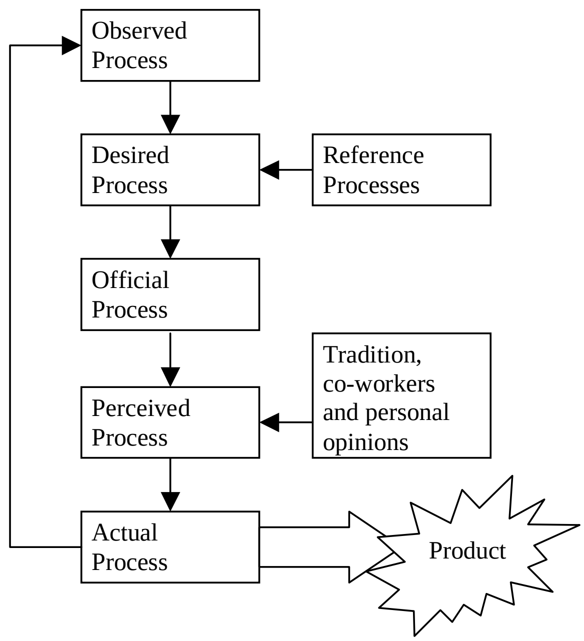

Figure 3: The Process feedback loop.

The mechanisms between the different processes are summarised in table 1. The most complicated process transition is between the official and the perceived process. The individual person's perceived process is affected by an individual interpretation of the official process, both as a result of reading it directly and as a result of some form of training. The perceived process is also affected by how Co-workers communicate their perceptions of the process and by observation of the co-workers' actual execution of the process. Finally the processes that have been in effect in the past also affect the perceived process.

Of the mentioned processes, the observed, desired and perceived processes are subject to viewpoints [4] by each person that is involved with this process. This means that each individual person's perception of this aspect of the process is different. The only externally observable processes are the official process, documented by official process documents, and the actual process, which consists of the actual actions that agents perform to produce the product. As the product is the result of the actual process, the SPI strategy must be to improve this process and make inconsistencies between the actual and the official processes as few and small as possible.

Somerville introduces the concept of viewpoints to software processes [4]. Several of the processes described in this paper thus far are subject to different persons in the company having different viewpoints. The observed process is subject to each SPI responsible persons interpretation, or viewpoint. The desired process is subject to the interpretation and application of reference processes and personal opinions of the SPI people. The perceived process is, as previously described, affected by many factors and each person has their own perceived process.

This paper focuses on using people's perception, or viewpoint, of the official and actual processes to help decision-making in process improvements.

3 Method

In order to support the strategy choice for an SPI initiative, a method is developed which involves the viewpoints of different roles in the software development organisation. In the description of the method, the general model for SPI presented in Figure 1 has been followed, with the additional steps shown in Figure 2, regarding the selection of which factors to implement.

3.1 Determine Goal

The first stage of an SPI initiative is to determine the goal for the process improvement. The goal has to be set in line with the business goals of the organisation. Any method that works in the organisation and produces a list of goals can be used. For instance, the goals of the CMM [2] can be used according to the current maturity level of the company.

3.2 Identify Major Factors

Once a goal has been determined, factors thought to be affecting this goal need to be identified, see Figure 4. The main focus of this paper is not, however, this identification of factors, but the following rating and use of the results of these ratings.

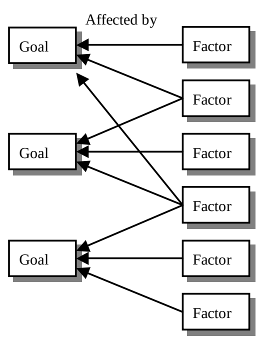

Figure 4: Relationship between goals and factors

Each goal must be possible to realize by factors affecting the development process. If it is not, it is an unrealisable goal, and therefore uninteresting to the SPI work. We define a factor as a possible change in the process, organisation, or environment that is believed to cause a change in the development process towards the goals associated with the factor.

Each factor has an actual effect and an actual cost that cannot be estimated accurately beforehand, if at all. However, the perception of the effect and cost of the introduction of the factors may provide enough information to identify the best factors for implementation. As these factors are to be rated for one goal at a time, the exact correlation between the factors and the goal is not as important as covering all factors that are possible to implement for the goal. This is because factors that are thought not to affect the goal during the rating should be rated low.

The number of factors used in the formal evaluation exercise is constrained by the time available for each person to complete the rating procedure. This must be decided taking into account the expected benefit of the whole procedure. In the next method step, pair-wise comparisons between the factors are performed, and an estimation of the time taken for the persons to complete the rating procedure can be performed by estimating a maximum and minimum time for reading an introductory text for the rating procedure and a maximum and minimum time for each comparison. From this a low and high estimation can be calculated for the total completion time and a cost/benefit decision can be taken. If the method has proven successful previously, a larger number of factors may be used the next time. With the rating scheme used in this paper, *n*(n-1)* comparisons are performed to compare *n* factors. If the number of people in the organisation is large, some form of sampling should be undertaken. The subjects were randomly assigned one of four randomly ordered rating pages to ensure that the order of the comparisons did not affect the results.

The goal/factor structure presented here is similar to the structure of goals, questions and metrics in the GQM method [5]. The GQM method, however, uses direct measures available in the actual process to measure aspects thereof in order to achieve the set goal by answering the questions associated with the goal. The method presented in this paper uses the perceptions of the individual persons to indicate the major factors contributing to the improvement goal. The method is, of course, affected by the factors selected for the rating. If a major factor is not identified, or omitted, this will mean that the finally selected factor might not be the ultimate one. The method of selecting factors should therefore be carefully selected and tailored to the situation.

Other methods for identifying factors can be studying past SPI strategies and their effect. Studying the effect of SPI in other companies is a further possibility.

3.3 Rate the Factors

The rating of the factors is to be performed by a suitable sample of all people involved in the development organisation that are active in the process. In a small organisation, a complete sample can be used, but in a large organisation this is not possible. The sample should be designed to ensure that all vertical levels of the organisation (i.e. management levels) and all horizontal groups (i.e. development teams, departments) are represented. The purpose of this is to ensure that all sources of different viewpoints should be represented in the sample.

The method used for rating the factors is called the Analytic Hierarchy Process (AHP). The AHP is a method for evaluating alternatives in a choice situation [6, 7]. It uses pair-wise comparisons and measures each alternative's relative contribution to the rating. Each subjects' individual set of comparisons is put into a judgement matrix. From this matrix the relative weights for the compared items can be calculated for each individual. These relative weights are known as priorities. The process also gives the possibility of calculating how consistent the performed ratings are for each individual as the pair-wise comparisons imply that the alternatives are indirectly compared several times.
The AHP process has been implemented for applications outside the process improvement domain. Karlsson et al. [6], for example, uses the process to select between customer requirements. This method of comparing requirements has been further developed into a commercial tool developed by Focalpoint AB [8]. This tool was not used for the experiment as it does not give full control over the algorithms used for calculating the results.

3.4 Analysis of Data

This step consists of two parts. First the whole data set is analysed and then comparisons are made between groups in the data in order to identify discrepancies within the organisation. To compare the relative importance of each factor, the rating results are aggregated first overall and then in each group. When aggregating results from an AHP comparison process we must decide whether we wish to aggregate the judgements of each individual or the priorities calculated for each individual. This is referred to as aggregation of individual judgements (AIJ) and aggregation of individual priorities (AIP) respectively. The decision is determined by whether the group is assumed to act as a unit or as separate individuals [9].

Secondly it must decided whether to use the arithmetic or geometric mean for the aggregation. For AIJ the geometric mean must be used. For AIP any of the two may be used, but the arithmetic mean has the advantage of being comparable to the original values. We have chosen to use AIP aggregation with arithmetical means.

The statistical methods described in this section are examples of appropriate methods for analysis of the data. There are several other methods that are appropriate and could be used.

3.4.1 Overall Analysis

An overall view of the data can be given by using box plots and descriptive statistics. Further analysis of the ratings can be performed using an ANOVA test [10], if the data is normally distributed and the variances are equal. Significant rating differences can then be established within each group using Fisher's PLSD test [10]. The normal properties of the data is checked using a Kolmogorov-Smirnov test for normal distribution. If the assumptions for ANOVA are not met by the data, a Kruskal-Wallis non-parametric test can be performed instead [11].

3.4.2 Identify Discrepancies

Discrepancies between groups of persons can give vital SPI strategy information. The ratings of one factor by two groups are compared using an unpaired t-test if the ratings are normally distributed. If the ratings are not normally distributed a Mann-Whitney test is performed [11]. A comparison of ratings in each group can be performed using the methods described for the overall results in section 3.4.

3.5 SPI Strategy Decision

The final outcome of the method is an overview of the expected impact of factors to improve in a development organisation and an idea of the difference between the groups in the organisation. The systematic nature of the method implies that we can be fairly sure that the results are valid as decision support material for the SPI strategy. The final decision will of course be taken taking other information into account as well. This is not an automatic decision taking method.

4 Case Study at Fuji Xerox

4.1 The Company

The method presented in Section 3 is applied in a case study at Fuji Xerox in Tokyo, Japan. Fuji Xerox Group is a joint venture owned by Fuji Photo Film Co., Ltd. and Xerox Ltd. of the United Kingdom. Fuji Xerox operates in the Asia-Pacific and Oceania regions as a member of the world wide Xerox Group. The company has approximately 15000 employees in its operating area.

The company's principal business is the manufacturing and selling of office automation equipment such as copiers and low-end laser printers, collectively referred to as the document services business. Other businesses include logistics and educational services. Fuji Xerox also performs research and development, marketing and service activities on behalf of the global Xerox Group.

4.2 Fuji Xerox SPI Background

The SPI process at Fuji Xerox (FX) began in 1995 when a proposition to introduce the SW-CMM won management approval. The company had an initial CMM assessment in April 96 and was found to be operating at level 1.

Improvement Action Teams (IAT) were started in July 96 to prepare for pilot projects operating at level 2. The pilot projects were started in April 97 and were reviewed in June of the same year. FX was assessed to be operating at level 2 in April 98. The SPI work has continued and the organisation was due to start pilot projects for level 3 operations in April 2000.

Occurrences of major events during the process initiative are illustrated in Figure 5. The goal of the program was to increase productivity and quality in the development process by working in the following three areas:

- Increase management visibility into ongoing projects.
- Activate more capability from the engineers using CMM.
- Achieve dynamic resource rotation.

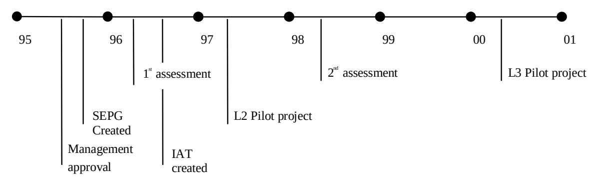

Figure 5: Major events in the Fuji Xerox SPI initiative.

Fuji Xerox was at the time of starting the program experiencing problems in the development of software. Two major factors were identified as causes:

- The size of the code in many projects has increased a factor 10 from 1990 to 1995 (approx. 100 KLOC to approx. 1 MLOC).
- The invisible nature of software.

The size of the code meant that a team consisting of one team leader and a group of developers was not sufficient to develop the software product in reasonable time. Instead the projects needed to be split up into several groups of developers and group leaders. This introduced new management and communication aspects that were hard to solve.

During the spring of 2001 Fuji Xerox has been successfully assessed at CMM level 3 and is operating several projects at level 3. The degree of deployment of level 2 was unknown in the spring of 2000 but it was estimated by members of the process improvement group that among the projects that have started the initiative approximately 80% of the key process areas for level 2 are satisfied.

The Copier Platform and Systems (CP&S) unit, where the case study was performed, has approximately 800 engineers working with software development for the Fuji Xerox document systems. The software is both embedded software for integration in copiers and printers and also independent software products for systems concerned with document handling.

Fuji Xerox employs a matrix organisational structure [12] for development projects. Project managers are responsible for coordinating and completing projects while the group leaders are responsible for providing the functional resources to the projects.

4.3 Determine the Goal

The goal determined at Fuji Xerox was identified by looking at the original intent of the SPI initiative and discussing this with concerned management. One of the important goals of the SPI initiative was to increase management visibility into ongoing projects. It was hence decided that this was one of the main goals to analyse.

Visibility concerns the information flow between the ongoing development projects and the responsible management and can be defined as:

“The ease and accuracy with which it is possible to assess the status of a project's cost, schedule, functionality, or other characteristics.” [13]

Information reaches management in three ways:

1. *Periodical reporting*
   Routine procedures that require reports or other forms of information to be produced and sent to management.
2. *On demand information*
   Management asks for specific information from the project.
3. *General impressions*
   Impressions obtained during day-to-day activities.

In an organisation with low visibility, information is only available and accurate in a close vicinity of the source. Hence if someone outside this area requires information the information has to be collected. This is usually performed in a non-standard fashion in a low visibility organisation. The information is therefore not comparable to information from other parts of the organisation and when it has been collected it may well already be out of date. In an organisation with high visibility, relevant information is collected in a standardised fashion and made available to authorised people, as they need it.

4.4 Identify Major Factors

The next step is to identify the factors, which affect the goal. The factors affecting managements’ visibility into development projects were identified by an observatory study of the SPI work performed at Fuji Xerox. The study was limited to SPI work within the Controller Platform and Service development (CP&S) unit as this unit has been making a software process improvement effort since 1995.

Many of the observations were made in conjunction with the improvement action team (IAT) meetings held under two weeklong periods during the spring of 2000.

The study performed is of a single execution nature and it is of the holistic type as the study concerns the SPI work within one division of Fuji Xerox [14]. During the case study no consideration was given to the situation within specific departments.

The observations made can be divided into 3 different categories depending on the source of the information:

1. General information provided as a background to Fuji Xerox development.
2. Impressions obtained during regular meetings with representatives from Fuji Xerox process improvement group.
3. Impressions obtained during IAT workshop weeks.

General information was collected through impressions during the stay at Fuji Xerox, brochures and internal documents provided by Fuji Xerox. The meetings held every two weeks with Fuji Xerox process experts were intended to keep track of the project, but also provided information about the actual development performed at Fuji Xerox and an opportunity for questions to be answered.

The IAT workshops were the most productive sources of information for process improvement information. This was because of the following reasons:

1. Process improvement is the main focus of these meetings.
2. A CMM expert consultant was present and willing to discuss the Fuji Xerox situation.
3. An interpreter was present and thereby solved any direct communication difficulties.

The IAT workshops were held on two occasions during the course of the case study. Each workshop was a weeklong session dealing with current issues of the SPI work. The main goal of the SPI work was intended to progress the development process on the CMM scale. The following people were present at the workshops:

- IAT team members.
- The CMM consultant and certified lead assessor.
- A Japanese-English interpreter.
- People relevant to the each current topic.

During each IAT workshop a special session was allocated to discussions regarding the SPI work described in this paper. A summary of the seven factors finally used in the AHP rating are presented in table 2. The factors affect the software process throughout the whole company.

The 7 major factors affecting the SPI goal identified at Fuji Xerox:

1. **Deployment:** Increase the number of projects using new processes in the organisation and optimise for those already using the standard process.
2. **Maturity:** Increase process maturity level as rapidly as possible in projects that are already in the improvement program.
3. **Tools:** Introduce new more powerful tools to aid the process, for instance automatic reporting tools and development tools.
4. **Training:** Ensure quality and acceptance of CMM training and integrate into all levels of the FX training program.
5. **Support-group process introduction:** Introduce process oriented thinking to operations outside the engineering function such as management, marketing, planning and production.
6. **Culture change:** Adapting the current CMM level 2 process to be more company specific.
7. **Standardising data:** Introducing standards to increase homogeneity of information and numerical data used throughout the organisation.

Table 2: The SPI Factors Identified in the Case Study

4.5 Rate the Factors

Senior Fuji Xerox management authorised a survey scope of approximately 160 subjects based on the estimated completion range of 24 to 52 minutes per subject. The subjects were to be chosen from within the CP&S unit consisting of 800 people in total.

After consulting with Fuji Xerox representatives, the strategy for choosing subjects was determined as follows:

1. Subjects should be chosen from projects that are a part of the process improvement initiative.
2. Project traits should be comparable using the survey (i.e. a sufficient number of people from each project should be sampled to establish an impression of differences between projects).
3. Organisation level traits should be comparable (i.e. differences between management and engineers should be apparent if present).

The AHP rating was performed using a web based form. The web pages were created in English and then translated to Japanese by a translator. The Japanese content of the web pages was reviewed by Fuji Xerox process experts together with the author of the form so that the content was as correct as possible in the following aspects:

- Content is in correct Japanese.
- Content is correct in Software Engineering terminology.
- Content is presented in a Japanese style.

This translation strategy implies that the content of the Japanese version is not a word for word exact translation of the English original. This was deliberately avoided in order to make the survey as effective as possible as the terminology involved is very different in the two languages.

The web pages first presents the purpose of the rating and instructions for the rating system. Then the concept of visibility used in this study is explained. This is followed by a description of all the factors to be compared. Finally the actual rating form is presented. The rating form itself is divided into two parts. The first part is intended to gather information from the subject in order to classify the subject. Classification is performed using the subject’s role in the functional organisation, either *management* or *engineer*. Management is defined as the three levels called division manager, group leader and function manager in the Fuji Xerox functional organisation. Engineer is defined as the two levels called sub leader and engineer.

The second part of the survey is the rating part. Each rating is presented with the factors to be rated on each side of a series of buttons that each represents a degree of choice in the AHP scale. All keywords are hyper-linked to short descriptions. The rating sheet is randomly chosen as one of four different randomly ordered rating sheets to address validity issues. An example of a rating is shown in Figure 6. When all the ratings have been completed the form is submitted and a thank you page is shown. The results are written to a database on the web server.

Test runs were performed on two occasions with two subjects on each occasion. The subjects’ impressions from these test runs were used to improve the contents of the survey. The test runs also verified that the server was reachable from all relevant development facilities around Tokyo, many at significant distances from the physical location of the server.

The actual rating started with an email sent to all 148 selected subjects explaining the purpose and reason for the rating and how to access the server. The server was available for a period of 11 days. During this time 75 subjects completed the rating procedure. Of these, 8 subjects were later removed due to lack of consistency in their ratings. The consistency of each individual’s ratings was appraised using the consistency ratio calculated according to the AHP method. The consistency ratio limit was set to 0.15, well below the limit of 0.2 that has been deemed acceptable in practice [6], which is good.

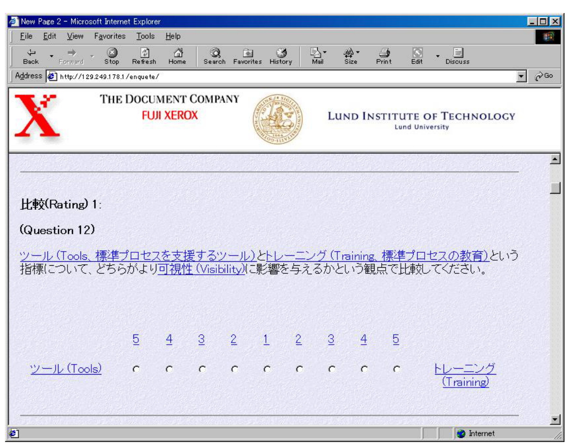

Figure 6: A rating from the web rating system.

4.6 Analysis

The analysis of the rating data is divided into two steps as described in the method section. First we analyse all the data to get an overall view of the trends in the data. Then we check for significant differences between the groups.

4.6.1 Overall Analysis

All the ratings were plotted in a box plot in order to get an overview of the whole data set. The box plot is shown in Figure 7.

The results were first aggregated using the arithmetic mean of the ratings for each factor, after outliers had been removed based on their consistency ratio, as described in section 3.4. The means are presented in the 'mean' column in Table 3. A Kolmogorov-Smirnov test was then performed on the data to check if it was normally distributed. It was found that the data for four out of the seven factors was normally distributed. In the case of the other three, the test could not reject that the data was normally distributed at a 0.05 significance level.

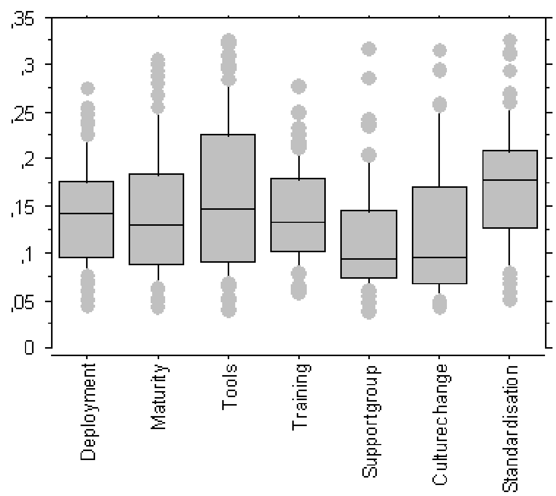

Figure 7: Box plot of all ratings

The next stage of analysis was to perform an ANOVA test. This test is normally used on a ratio scale and as the results from the AHP are of a ratio nature compared to the ordinal nature of the ratings. This test showed that there is a significant difference in the data set at the 0.05 significance level. As we could not prove the normal distribution of all the factors we also performed a Kruskal-Wallis test. This test gave the same result as the ANOVA test. The final test that we would like to perform is the Fisher PLSD test to investigate the individual relationships in the data, but this requires that the preconditions for the ANOVA are satisfied. As the ANOVA provided the same result at 0.05 significance level as the Kruskal-Wallis test and taking into account the robustness of the ANOVA test towards fit of normality, the Fisher PLSD test was used nonetheless [15]. The whole analysis procedure is illustrated in Figure 8.

	Overall Rating	Mean	Std Dev	Std Err
1	Standardisation	0.1733	0.0668	0.007
2	Tools	0.1614	0.0801	0.009
3	Deployment	0.1429	0.0556	0.010
4	Maturity	0.1422	0.0700	0.006
5	Training	0.1408	0.0513	0.008
6	Culture change	0.1238	0.0776	0.010
7	Support group	0.1152	0.0612	0.008

Table 3: Aggregated Results from the AHP calculations

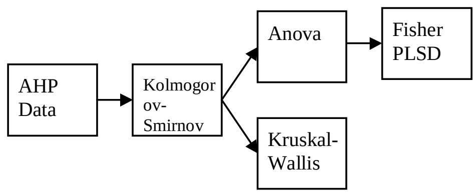

Figure 8: Data analysis procedure for all subjects

The results of the PLSD show that the Standardisation factor is rated significantly higher than Training, Deployment, Maturity, Support group and Culture change. Tools is rated significantly higher than Support group and Culture change. Finally Training, Deployment and Maturity are rated significantly higher than Support group. These relationships are illustrated in Figure 9. In the figure a significant relationship at the 0.05 level is denoted by a ‘S’ at the point of intersection between two factors. The factors are ordered according to their arithmetic means.

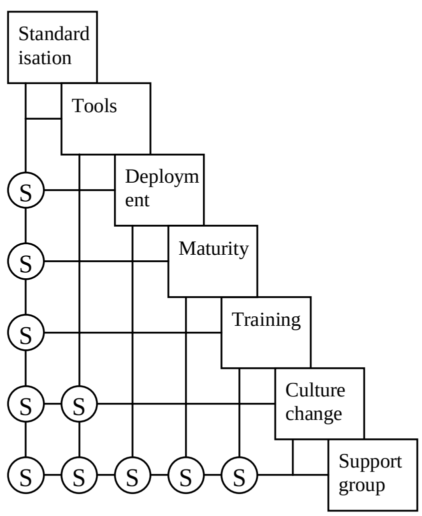

Figure 9: Significant relationships between factors according to Fisher’s PLSD.

The results indicate that the factors Standardisation and Tools seem to be the most firmly anchored in the organisation. To investigate if this finding is tied to a specific group, the analysis is continued by dividing the data into the two different groups.

4.6.2 *Identify Discrepancies*

For the purpose of identifying discrepancies, the data was divided into two groups according to the role of the subjects, the management group and the engineer group. The hierarchical nature of the organisation implies that there are more engineers available to perform the comparisons. In total 10 managers performed the comparisons, versus 56 engineers. A Kolmogorov-Smirnov test for normal distribution was performed on the data in each group as in the case with the whole data set. Each factor was then compared between the groups using a Mann-Whitney test. If no significant relationship (0.05 level) was discovered by this and both groups were normally distributed for the factor a t-test was performed instead. This analysis procedure is illustrated in Figure 10.

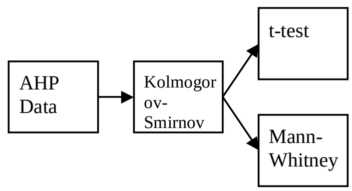

Figure 10: Data analysis procedure for comparison between groups

Significant differences were found between the two groups for factors Maturity (using the Mann-Whitney test) and Standardisation (using the t-test). Next, the analysis procedure performed for the entire data set was repeated to investigate the rated relationships between the factors within each group. The results for this analysis are summarised in Figures 11 and 12 for engineers and management respectively. It can be seen that the order of factors is different between the two groups.

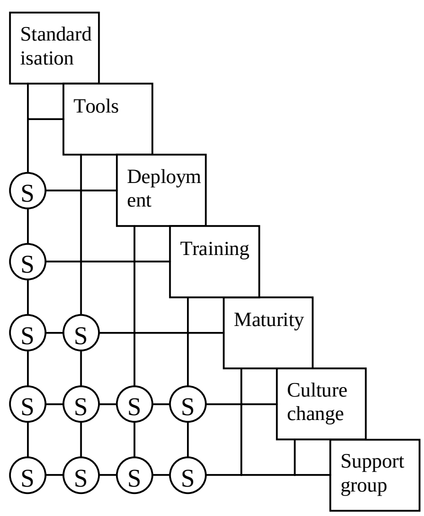

Figure 11: Significant relationships between factors in the 'Engineering' group according to Fisher's PLSD.

The results indicate that there is a difference between the groups regarding Maturity and Standardisation. It is interesting to note though that the groups agree on the importance of Tools.

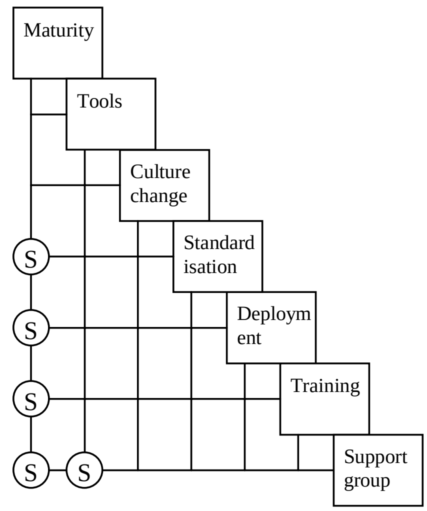

Figure 12: Significant relationships between factors in the 'Management' group according to Fisher's PLSD.

4.7 Validity

The threats to the validity of the survey have been evaluated according to the lists presented by Wohlin et al. [16]. Only significant threats to validity are discussed.

4.7.1 Conclusion Validity

Reliability of measures is a validity concern during the translation of the survey form from English to Japanese. It is not certain that the translator understands the context of the English version and therefore may provide a translation that misleads the subjects. This validity concern was reduced by arranging review sessions with the author of the English text and English-speaking process experts from Fuji Xerox.

Cultural differences were addressed by performing the case study over the relatively long period of six months and by reviewing observations with English-speaking members of Fuji Xerox staff.

The formulation of the factor descriptions and the formulation of the instructions will affect the ways in which the subjects complete the survey. The effects of these validity concerns where reduced by putting a lot of time into the construction of the survey and by test running the form twice. Good consistency values during the test runs and during the actual runs suggest that the subjects have understood both the instructions for performing the survey and the meaning of each factor.

The order of the ratings may affect the results from the rating procedure. The randomisation of the rating order is an attempt to reduce this factor.

Reliability of treatment implementation is a validity concern for the survey as the form was applied over a web interface there was no control of the environment during the duration of the study. The subjects were free to fill out the form in any situation they chose and it is assumed that most of the subjects filled out the form in his/her normal working place.

Random irrelevancies in experimental setting: The effects of the subjects' normal working environment cannot be eliminated as a validity threat as there was no control over this.

Threats to the conclusion validity concerning statistical tests are under control. The data is tested for normal properties, and the tests applied (ANOVA and t-test) are standard test for the analysis. In the cases where normality cannot be established, non-parametric tests are used instead (Kruskal-Wallis and

Mann-Whitney tests).

4.7.2 Internal Validity

Maturation: The subjects could become more acquainted with the different rating factors as the survey progresses. The randomisation of the order of the ratings should reduce this threat.
Instrumentation: Wordings in the web form affect the grounds for the rating choices made and will therefore be a validity threat. Wordings in the instructions to the rating will affect the way the subjects perform the study and may therefore threaten the validity. The relatively high level of consistency and the general homogeneity of the results suggests that the effects of these two threats are low, unless the descriptions explicitly make a certain rating order favourable.

Selection: The basis on which the subjects were selected does not provide a correct sample from the group as Fuji Xerox representatives chose projects that they were interested in. The low response rate caused by the fact that there was no method of forcing form completion means that only subjects willing to fill out the form did so.

4.7.3 Construct Validity

Inadequate pre-operational explication of constructs: Although great care was taken in defining the constructs the limited total time of the study implied that there came a time when there simply was no more time to work on these and the study had to be performed.

Evaluation apprehension: Subjects can be purposely not answering the survey in order to demonstrate how busy they are and how devoted they are to working with their primary tasks.

4.7.4 External Validity

The nature of the case study method usually implies low external reliability [14]. As the factors were identified using observations made at Fuji Xerox the study has external limitations. Fuji Xerox however faces the same challenges as most other large corporation developing software. Making changes in the way the company functions takes an extremely long time and standardising is difficult due to the large number of people and opinions involved. The case study does, however, show that the method presented in the paper is useful in one software engineering organisation and it provides information useful to other companies interested in trying the method.

4.7.5 Validity Summary

The lack of control of the environment where the subjects performed the tests because of the web interface is a major validity threat together with translation and formulation effects in the form itself. The main threat to internal validity in this study is the selection of the subjects. Even though the results are specific to Fuji Xerox they can provide useful information to other organisations in similar situations.

4.8 Case Study Summary

The case study shows that the factors can be prioritised and that differences exist between different groups in the company. Especially the study shows that there are several differences between subjects at the management level and at the engineering level in the organisation.

The methods used in the case study prove adequate for identifying the factors and performing a prioritisation between the factors. The analytical hierarchy process worked extremely well for rating the factors provided from the case study. Especially, the ability to calculate the consistency of the subjects gave a good idea of the quality of the results obtained and also provided a means to remove inconsistent subjects.

It is interesting to note that the comparison between the groups showed a significant difference in one of the most highly rated factors overall, Standardisation. This should have a major impact on future SPI work. The Fuji Xerox management group decided to investigate introducing standardisation of low-level work processes through the use of some kind of reference model such as the Personal Software Process [17] after the results of the investigation were presented.

5 Summary and Conclusions

In a software improvement program, there is a risk that the improvement strategy is not well anchored in the organisation. Different persons in the organisation have different viewpoints on what the process is, and how it should be improved.

To support the establishment of an improvement program strategy, which takes these different viewpoints into account, a method is developed and evaluated in a case study at Fuji Xerox, in Tokyo, Japan. The method supports the improvement program by providing decision support information for SPI work.

Based on a determined goal and a set of factors that affect the goal, people from the organisation rate the factors, using the analytic hierarchy process (AHP), which basically means pair-wise comparisons of the factors. The AHP provides ranking of the factors as well as a measure of how consistent the rankings are. There is also an opportunity to compare the ranking throughout the organisation and thereby identify any discrepancies. With the information collected in this way, management is expected to be able to make better decisions concerning SPI strategy.

With the information at hand, which was derived by using the method, it was shown in the case study that management was able to make better decisions on which strategy to choose, and how to create a more homogenous process perception throughout the organisation. The method provided information on the viewpoints of different stakeholders in the organisation, which was a support in the selection of improvement factors as well as in the identification of change support needed. Further, the involvement of people at different levels in the organisation provides in itself a more firmly anchored improvement program.

6 Acknowledgements

The authors would like to thank Fuji Xerox for their extensive cooperation and support for the project. Especially Dr. Jun Miyazaki and Mr. Atsushi Nakamura. Mr. Mark Amaya of Software Process People and Technology provided many illuminating insights into the process improvement work at Fuji Xerox. Finally we would like to thank the Sweden-Japan Foundation for the funding of the research performed in Japan associated with the project.

7 References

[1] Humphrey, W. S., "*Managing the Software Process*", Addison-Wesley, 1989.

[2] Paulk, M.,C., Curtis, B., Chrissis, M., B., and Weber, C., V., "*Capability Maturity Model for Software, Version 1.1*", Software Engineering Institute, CMU/SEI-93-TR-24, February 1993.

[3] Bandinelli, S., Fuggetta, A., Lavazza, L., Loi, M., Picco, G. P. "*Modelling and Improving an Industrial Software Process*", IEEE Transactions on software Engineering, Vol. 21, No. 5, 1995, pp. 440-454.

[4] Sommerville, I., "*Managing Process Inconsistencies Using Viewpoints*", IEEE Transactions on Software Engineering vol. 25, No. 6, 1999, pp. 784-799.

[5] van Solingen, R., Berghout, E., *"The Goal/Question/Metric Method"*, McGraw-Hill International (UK) Limited, 1999.

[6] Karlsson, J., Ryan, K., *"A Cost-Value Approach for Prioritising Requirements"*, IEEE Software September/October, 1997, pp. 67-74.

[7] Saaty, T. L., *"The Analytic Hierarchy Process"* McGraw-Hill, New York, 1980.

[8] Home page of Focal Point AB, accessed 02/03/05, http://www.focalpoint.se/index_eng.html.

[9] Forman, E., Peniwati, K., "*Aggregating Individual Judgements and Priorities with the Analytic Hierarchy Process"*, European Journal of Operational Research, 1998, pp. 165-169.

[10] Montgomery, D. C., *"Design and Analysis of Experiments"*, John Wiley and sons, *1997.*

[11] Siegel, S., Castellan Jr., N. J., *"Nonparametric Statistics"*, McGraw-Hill International, 1998.

[12] Thayer, R., H., "*Software Engineering Project Management"*, from "*Software Engineering Project Management"*, edited by Thayer, R., H., pp. 72-104, IEEE Computer Society, 2000.

[13] McConnell, S., "*Software Project Survival Guide"*, Microsoft Press, 1998.

[14] Yin, R. K., "*Case Study Research Design and Methods"*, Sage Publications, Beverly Hills, California, 1994.

[15] Bratthall, L., Wohlin, C., "*Understanding Some Software Quality Aspects from Architecture and Design Descriptions"*, International Workshop on Program Comprehension, 2000, pp.27-36.

[16] Wohlin, C., Runeson, P., Höst, M., Ohlsson, M., Regnell, B., Wesslén, A., "*Introduction to Experimentation in Software Engineering "*, Kluwer Academic Publishers, 2000.

[17] Humphrey, W.S., "*A Discipline for Software Engineering"*, Addison-Wesley, 1995.