Aggregating viewpoints for strategic software process improvement—a method and a case study

D. Karlström, P. Runeson and C. Wohlin

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. 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 are described 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 to a case study performed at Fuji Xerox, Tokyo. In the case study, significantly different profiles of the factor ratings came from management compared with those 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 programme, 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 implemented effectively.

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 Fig. 1. These are also the main steps of the more advanced SPI method called the IDEAL process [2] introduced by SEI.

This paper focuses on the goal and analysis part of the SPI model shown in Fig. 1. It presents a method that 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 Fig. 2.

Step A: determine goal;
Step B: identify major factors believed to be affecting the goal;
Step C: allow the organisation to rate the factors anonymously based on their personal viewpoints;
Step D: analysis of data. Aggregate and analyse the rating results and identify discrepancies; and
Step E: SPI strategy decision.

The method provides decision-support information to SPI-responsible individuals by identifying the perceived most effective improvement factors and by analysing discrepancies

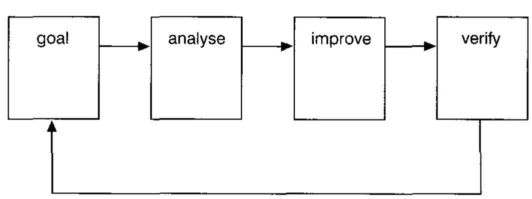

Fig. 1 *General SPI model*

IEE Proceedings online no. 20020696
DOI: 10.1049/ip-sen:20020696
Paper first received 19th April and in revised form 30th July 2002

D. Karlström and P. Runeson are with the Department of Communication Systems, Lund University, Box 118, SE-221 00 Lund, Sweden
C. Wohlin is with the Department of Software Engineering & Computer Science, Blekinge Institute of Technology, Box 520, SE-372 25 Ronneby, Sweden

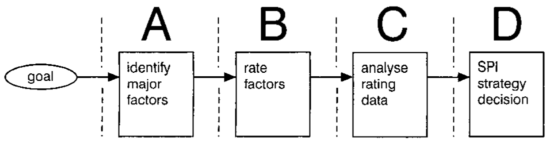

Fig. 2 *Overview of the method*

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. An example of such factors is provided by Cattaneo *et al.* [3] by demonstrating the need for organisational restructuring within development companies prior to, or in combination with, CMM-based SPI work.

Herbsleb *et al.* show the importance of management involvement in successful CMM-based SPI in their survey of companies using the framework [4]. By making information available about the views on process-improvement alternatives, the method presented in this paper is thought to be able to improve management involvement in organisations.

2 Process distortion and viewpoints

In a process-improvement initiative, there may be different opinions on what should be improved 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 [5]. To this list can be added 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 Fig. 3. Note that of all the processes shown in Fig. 3, only the actual process produces the product and the only absolutely accessible process is the official process.

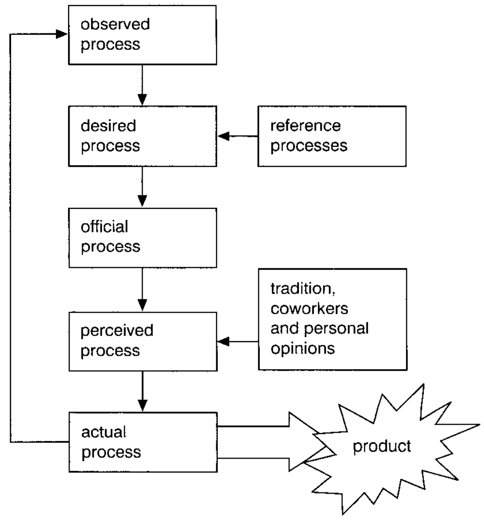

Fig. 3 *Process-feedback loop*

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 coworkers communicate their perceptions of the process and by observation of the coworkers' actual execution of the process. Finally, processes that have been in effect in the past also affect the perceived process.

Table 1: Mechanisms of transfer between types of process

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

Of the processes mentioned, the observed, desired and perceived processes are subject to viewpoints [6] 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.

Sommerville *et al.* introduce the concept of viewpoints to software processes [6]. 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 person's 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.

Herbsleb *et al.* [4] analyse responses from respondents with differing roles within the companies using CMM-based SPI in their survey. No statistical differences are shown between different roles, but only one person in each role is used from each company.

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

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 Fig. 1 has been followed, with the additional steps shown in Fig. 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 Fig. 4). The main focus of this paper is not, however, this identification of factors, but the subsequent rating and use of the results of these ratings.

Each goal must be possible to realise by factors affecting the development process. If it is not, it is an unrealisable goal, and therefore uninteresting to the SPI work. A factor is defined 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.

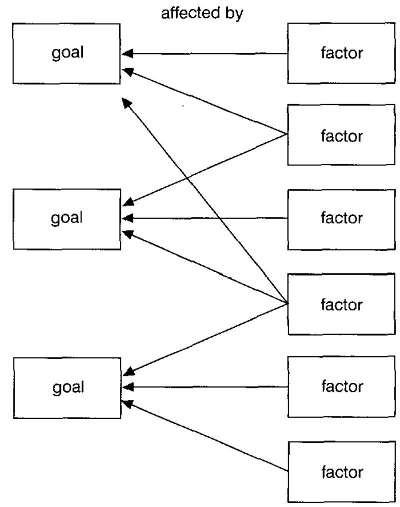

Fig. 4 *Relationship between goals and factors*

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, pairwise 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, $n*(n-1)/2$ 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 [7]. The GQM method, however, uses direct measures available in the actual process to measure aspects thereof 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 factor finally selected 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 options in a choice situation [8, 9]. It uses pairwise comparisons and measures each option's relative contribution to the rating. Each subjects' individual set of comparisons is put into a judgment matrix. From this matrix the relative weights for the items compared 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 pairwise comparisons imply that the options are indirectly compared several times.

The AHP process has been implemented for applications outside the process-improvement domain. Karlsson *et al.* [8], 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 [10]. 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 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 one must decide whether one wishes to aggregate the judgments of each individual or the priorities calculated for each individual. This is referred to as aggregation of individual judgments (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 [11]. Secondly it must be 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 with the original values. AIP has been chosen in this paper for 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 [12], 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 [12]. The normal properties of the data are checked using a Kolmogorov–Smirnov test for normal distribution. If the assumptions for ANOVA are not met by the data, a Kruskal–Wallis nonparametric test can be performed instead [13].

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 [13]. 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 one can be fairly sure that the results are valid as decision-support material for the SPI strategy. The final decision will, of course, be 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 UK. Fuji Xerox operates in the Asia–Pacific and Oceania regions as a member of the world wide Xerox Group. The company has approximately 15 000 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 1996 and was found to be operating at level 1.

Improvement-action teams (IAT) were started in July 1996 to prepare for pilot projects operating at level 2. The pilot projects were started in April 1997 and were reviewed in June of the same year. FX was assessed to be operating at level 2 in April 1998. 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 Fig. 5. The goal of the program was to increase productivity and quality in the development process by working in the following three areas:

(*a*) increase management visibility into ongoing projects;
(*b*) activate more capability from the engineers using CMM; and
(*c*) achieve dynamic resource rotation.

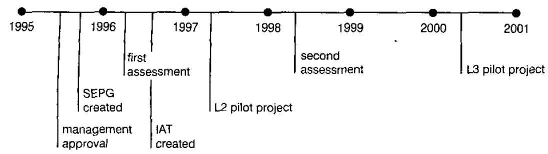

Fig. 5 *Major events in the SPI initiative*

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

(i) an increase in the size of the code in many projects by a factor 10 from 1990 to 1995 (from approximately 100 KLOC to approximately 1 MLOC); and
(ii) 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 a 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 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 [14] for development projects. Project managers are responsible for co-ordinating 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 the management concerned. 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'. [15]

Information reaches management in three ways:

(i) *periodical reporting*: routine procedures that require reports or other forms of information to be produced and sent to management;
(ii) *On-demand information*: management asks for specific information from the project; and
(iii) *General impressions*: impressions obtained during day-to-day activities.

In an organisation with low visibility, information is only available and accurate in the 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 nonstandard fashion in a low-visibility organisation. The information is therefore not comparable with 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 management's visibility into development projects were identified by a qualitative study, observing the SPI work performed at Fuji Xerox. The study was limited to SPI work within the 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 week-long 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 during one single period of time [16]. During the case study, no consideration was given to the situation within specific departments.

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

(*a*) general information provided as a background to Fuji Xerox development;
(*b*) impressions obtained during regular meetings with representatives from Fuji Xerox process-improvement group; and
(*c*) 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 for the following reasons:

(i) Process improvement is the main focus of these meetings.
(ii) A CMM expert consultant was present and willing to discuss the Fuji Xerox situation.
(iii) 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 week-long 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, and
- 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 is presented in Table 2. These factors affect the software process throughout the whole company.

4.5 Rate the factors

Senior Fuji Xerox management authorised a survey scope of approximately 160 subjects based on the estimated completion range of 24–52 min 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:

(i) Subjects should be chosen from projects that were a part of the process-improvement initiative.
(ii) 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).
(iii) 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 was in correct Japanese.
- content was correct in software-engineering terminology.
- content was presented in a Japanese style.

This translation strategy implied that the content of the Japanese version was not a word-for-word exact translation of the English original. This was deliberately avoided 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 subleader 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 hyperlinked 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 Fig. 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 e-mail 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, eight-subjects were later removed owing 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 [8], which is good.

4.6 Analysis

The analysis of the rating data is divided into two steps as described in the method section. First all the data were analysed to obtain an overall view of the trends in the data. Then checks were made 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 Fig. 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

Table 2: SPI factors identified in the case study

Factor	SPI goals
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, e.g. automatic reporting tools and development tools
4 Training	Ensure quality and acceptance of CMM training and integrate into all levels of the FX training programme
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

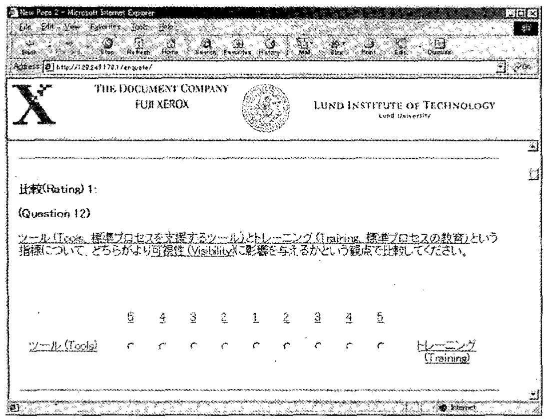

Fig. 6 *Screen-shot from the web-based rating system*

normally distributed. It was found that the data for four out of the seven factors were normally distributed. In the case of the other three, the test could not reject that the data were normally distributed at a 0.05 significance level.

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 with the ordinal nature of the ratings one can apply the test. To investigate the relationships between the rated factors, a null-hypothesis was formulated such that there are no significant differences between the seven factors:

$$H_0: R_1 = R_2 = \cdots = R_7$$

The alternative hypothesis was formulated such that there is a significant difference between at least one pair of the seven factors:

$$H_1: R_m <> R_n$$

for at least one pair of factors m, n.

The test showed that there is a significant difference in the data set at the 0.05 significance level. As the normal distribution of all the factors could not be proved, a Kruskal–Wallis test was also performed. This test gave the same result as the ANOVA test. The final test that it would be desirable to perform is the Fisher PLSD test to

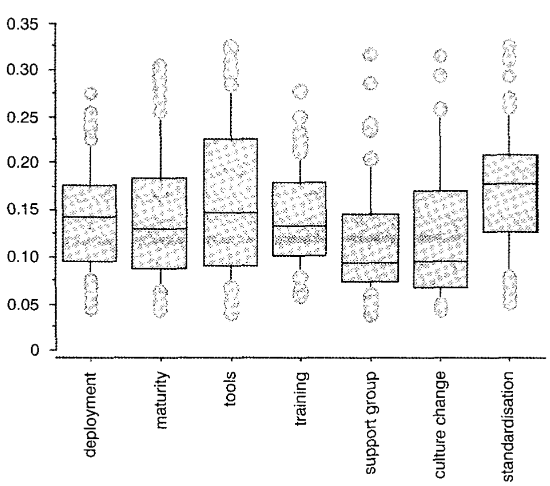

Fig. 7 *Box plot of all ratings*

Table 3: Aggregated results from the AHP calculations

Overall/Rating	Mean	Standard deviation	Standard error
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

investigate the individual relationships in the data, but this requires that the preconditions for the ANOVA be 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 [17]. The whole analysis procedure is illustrated in Fig. 8.

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 Fig. 9. In Fig. 9, if hypothesis H_0 is rejected for a pair $[m, n]$ at the 0.05 level, this is denoted by a S at the point of intersection between two factors. The factors are ordered according to their arithmetic mean rating.

The results indicate that the factors standardisation and tools seem to be the most firmly anchored in the organisation. To investigate whether 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 were 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, against 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.

Next the hypotheses was formulated as follows:

$$H_0: R(\text{manager})n = R(\text{engineer})n, n = 1, \ldots, 7$$
$$H_1: R(\text{manager})n <> R(\text{engineer})n, n = 1, \ldots, 7$$

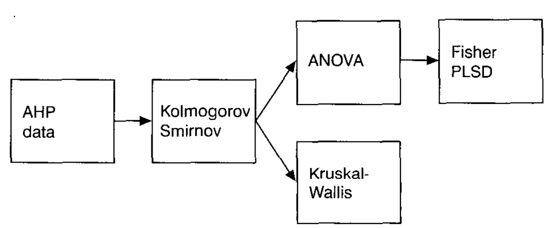

Fig. 8 *Data-analysis procedure for all subjects*

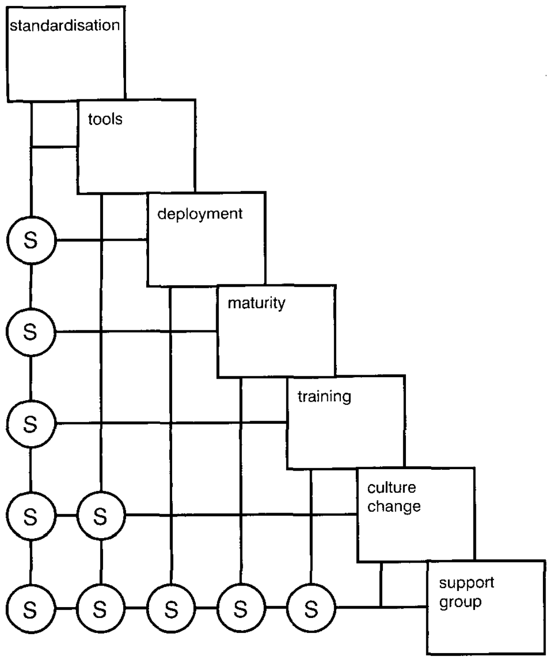

Fig. 9 *Significant relationships between factors according to Fisher's PLSD*

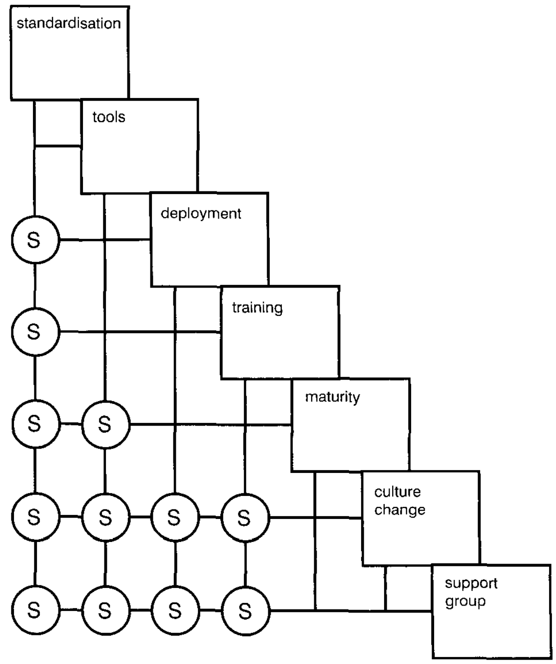

Fig. 11 *Significant relationships between factors in the 'engineering' group according to Fisher's PLSD*

The null hypothesis states that there is no statistical difference between each pair of factors from the respective group and the alternative hypothesis states that there is a statistical difference.

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 Fig. 10.

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 Figs. 11 and 12 for engineers and management, respectively. It can be seen that the order of factors is different in the two groups.

The results indicate that there is a difference between the groups regarding maturity and standardisation. Note, though, that the groups agree on the importance of tools.

4.7 Validity

The threats to the validity of the survey have been evaluated according to the lists presented by Wohlin *et al.* [18]. Only significant threats to validity are discussed.

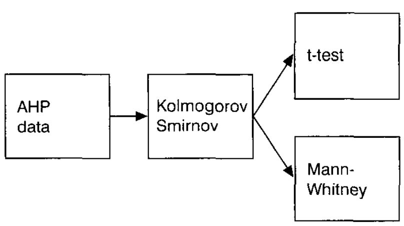

Fig. 10 *Data-analysis procedure for comparison of groups*

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

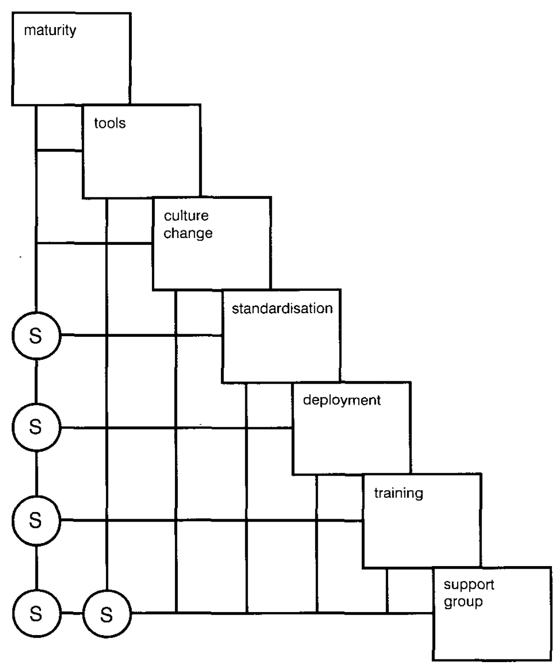

Fig. 12 *Significant relationships between factors in the 'management' group according to Fisher's PLSD*

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 so that 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 at 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 were tested for normal properties, and the tests applied (ANOVA and t-test) are standard tests for the analysis. In the cases where normality could not be established, nonparametric tests were 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 progressed. 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 preoperational 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 was simply no more time to work on these and the study had to be performed.

Evaluation apprehension: Subjects may purposely not answer the survey 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 [16]. 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 corporations developing software. Making changes in the way the company functions takes an extremely long time and standardisation is difficult owing 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. In particular, 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. In particular, 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 of removing inconsistent subjects.

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 [19] after the results of the investigation were presented.

5 Summary and conclusions

In a software-improvement programme, 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-programme 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 programme by providing decision-support information for SPI work.

Based on a predetermined 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 pairwise 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 Acknowledgments

The authors thank Fuji Xerox for extensive co-operation and support for the project, especially, Dr. Jun Miyazaki and Atsushi Nakamura. Mark Amaya of Software Process People and Technology provided many illuminating insights into the process-improvement work at Fuji Xerox. Finally we 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, Boston, MA, 1989)
2 PAULK, M.C., CURTIS, B., CHRISSIS, M.B., and WEBER, C.V.: 'Capability maturity model for software, version 1.1'. CMU/SEI-93-TR-24, Software Engineering Institute, Feb. 1993
3 CATTANEO, F., FUGGETTA, A., and SCIUTO, D.: 'Pursuing coherence in software process assessment and improvement', *Softw. Process Improv. Pract.*, 2001, **7**, pp. 3–22
4 HERBSLEB, J.D., and GOLDESON, D.R.: 'A systematic survey of CMM experience and results'. Presented at 18th International Conference on Software engineering, ICSE'96, IEEE Computer Society Press, 1996
5 BANDINELLI, S., FUGGETTA, A., LAVAZZA, L., LOI, M., and PICCO, G.P.: 'Modelling and improving an industrial software process', *IEEE Trans. Softw. Eng.*, 1995, **21**, (5), pp. 440–454
6 SOMMERVILLE, I., SAWYER, P., and VILLER, S.: 'Managing process inconsistencies using viewpoints', *IEEE Trans. Softw. Eng.*, 1999, **25**, (6), pp. 784–799
7 VAN SOLINGEN, R., and BERGHOUT, E.: 'The goal/question/metric method' (McGraw–Hill International (UK) Ltd., 1999)
8 KARLSSON, J., and RYAN, K.: 'A cost-value approach for prioritising requirements', *IEEE Softw.*, 1997, **13**, pp. 67–74
9 SAATY, T.L.: 'The analytic hierarchy process' (McGraw–Hill, New York, 1980)
10 http://www.focalpoint.se/index_eng.html, home page of Focal Point AB. Accessed 2nd March 2002
11 FORMAN, E., and PENIWATI, K.: 'Aggregating individual judgements and priorities with the analytic hierarchy process', *Euro. J. Oper. Res.*, 1998, **22**, pp. 165–169
12 MONTGOMERY, D.C.: 'Design and analysis of experiments' (John Wiley & Sons, Singapore, 1997)
13 SIEGEL, S., and CASTELLAN, N.J.JR.: 'Nonparametric statistics' (McGraw–Hill International, USA, 1998)
14 THAYER, R.H.: 'Software engineering project management', in Thayer, R.H. (Ed.): 'Software engineering project management'(IEEE Computer Society, 2000), pp. 72–104
15 MCCONNELL, S.: 'Software project survival guide' (Microsoft Press, USA, 1998)
16 YIN, R.K.: 'Case study research design and methods' (Sage Publications, Beverly Hills, CA, 1994)
17 BRATTHALL, L.WOHLIN, C.: 'Understanding some software quality aspects from architecture and design descriptions'. Proceedings of the International Workshop on Program comprehension, 2000, pp. 27–36
18 WOHLIN, C., RUNESON, P., HÖST, M., OHLSSON, M., REGNELL, B., and WESSLÉN, A.: 'Introduction to experimentation in software engineering' (Kluwer Academic Publishers, USA, 2000)
19 HUMPHREY, W.S.: 'A discipline for software engineering' (Addison–Wesley, USA, 1995)