Available online at www.sciencedirect.com
SCIENCE DIRECT®
International Journal of Information Management 26 (2006) 30–43
International Journal of
Information
Management
www.elsevier.com/locate/ijinfomgt

Problems in measuring effectiveness in software process improvement: A longitudinal study of organizational change at Danske Data

Jakob Iversen[a,*], Ojelanki Ngwenyama[b,c]

[a]University of Wisconsin Oshkosh, 800 Algoma Blvd., Oshkosh, WI 54901-8675, USA
[b]Ryerson University, Toronto, Canada
[c]Virginia Commonwealth University, Richmond, VA 23284-4000, USA

Abstract

Software process improvement (SPI) is a widely recognized approach that software companies implement to improve quality, productivity, and time-to-market. Assessing and analyzing performance improvements are important SPI activities. However, many SPI managers have found it difficult to develop and implement effective performance measurement programs for SPI, in part because guidelines for conducting SPI measurements are scarce. We address this gap in the SPI literature by examining major problems that SPI change agents encounter when developing and implementing SPI measurement programs. We report on a longitudinal study of an SPI change initiative and the challenges that the SPI Team faced in dealing with the issues of measuring effectiveness of the initiative. We systematically analyze an SPI performance measurement program to understand its limitations and the problems that the SPI Team encountered when implementing it. We used an organizational change theory framework to derive theoretical and practical insights that can help managers and researchers develop and implement better SPI performance measurement programs.

Keywords: Software metrics; Software process improvement; Organizational change; Longitudinal process research

1. Introduction

The last decade has been one of continuous and unrelenting global competition in the software industry. A growing lower wage and a highly skilled software engineering labor force in India and other Asian countries have precipitated shifts in software production from Europe and North America to Asia. In response to these challenges, senior managers in European and North American software industries are adopting a variety of strategies to stay competitive. For example, some firms are outsourcing software development to Asia, and others are setting up Asian offices to take advantage of the lower cost of production. Still others are implementing various forms of organizational change, such as business process re-engineering; more flexible

*Corresponding author.
E-mail addresses: iversen@uwosh.edu (J. Iversen), ojelanki@ryerson.ca (O. Ngwenyama).

0268-4012/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijinfomgt.2005.10.006

approaches to organizing work, such as teams and autonomous work-groups; and software process improvement (SPI).

Many business firms view SPI as a strategic issue and are involved in organizational change initiatives to improve their software development practices. The fundamental objective of the SPI movement is to change the organizational practices of software development firms, in order to improve software quality and reliability, employee and customer satisfaction, and profitability (Hansen, Rose, & Tjørnehøj, 2004). One of the factors that affect the success of an SPI program is the presence of an effective measurement program to support decision-making and inform managers about the state of their software development practices. The measurement and analysis of organizational effectiveness is an important activity in any SPI initiative. Measuring aspects of organizational effectiveness in software development can provide management with critical information upon which to base decisions about allocating resources to effect change in specific areas.

Although there is broad agreement that a systematic and reliable methodology for measuring effectiveness is necessary for the success of SPI change initiatives (Goldenson, Emam, Herbsleb, & Deephouse, 1996; Goldenson & Herbsleb, 1995; Hayes & Zubrow, 1995), there is little agreement on what should be measured. Many researchers view the lack of a systematic and reliable measurement methodology as a major factor contributing to the high failure rate of SPI initiatives. For example, Bill Curtis, one of the authors of the Capability Maturity Model (CMM) (Paulk, Curtis, Chrissis, & Weber, 1993a), estimated that as many as 70% of the firms implementing SPI are unsuccessful (Curtis, 1996). More research is needed to develop systematic measurement programs for assessing effectiveness of SPI initiatives. The main objective in this paper is to develop an understanding of how to measure software development performance and how these measures can inform software development managers, SPI practitioners, and senior management.

The theoretical foundation of our analysis is organizational change theory, and our research methodology is based on longitudinal process analysis. The focus of our discussion is a longitudinal study of an SPI change initiative in the Danish software company, Danske Data. The rest of the paper is organized as follows: The next section describes the research methodology. Section 3 describes the organizational change theory framework we used in the study. Section 4 describes the case organization and the SPI program that is the subject of this study. Section 5 presents the empirical findings, and Section 6 outlines implications for future research on measuring effectiveness in SPI.

2. Research methodology

This research is part of a larger research project to improve the state-of-the-art of software development in four Danish software-developing companies (Mathiassen, Pries-Heje, & Ngwenyama, 2002). The research project was a longitudinal (from 1996 to 2001) process study of the SPI initiatives intended to advance the four companies up the CMM scale. Most of the researchers were engaged in the four companies and actively participated in the SPI projects, and were thus able to understand the SPI initiatives from the perspective of participant observers.

The major strength of longitudinal process research (LPR) is the opportunity to get close to the organization and find out what is really going on. However, close involvement can also be problematic if researchers do not systematically collect data to generalize findings (Foster, 1972; Mathiassen, 1998). We have attempted to address this potential problem by adhering to the five basic principles of LPR: (1) longitudinal engagement with the research site, (2) participant observation, (3) multiple sources of data; (4) systematic data gathering, and (5) reliable data recording or transcription (Lincoln & Guba, 1985; Monge, 1990; Pettigrew, 1985).

The data collection for this study started in December 1996 and continued through December 2000. One of the authors communicated daily with the members of Danske Data, and made regular visits to the organization (at least one each month) to participate in regular project meetings. We also collected a broad range of organizational artifacts relevant to the SPI initiative at Danske Data. In addition to the participant observation, we recorded meeting minutes, and collected e-mail, memos, reports, project plans, and so on (Table 1). This multiple-source data collection strategy supports triangulation, systematic gathering, and reliable recording and transcription of data, which in turn helps to ensure the validity of our empirical findings.

Table 1
List of data sources

Data source	Explanation
Meeting notes	Researchers' notes from SPI Team meetings and performance measurement group meetings.
Meeting minutes	Minutes from SPI Team meetings and performance measurement group meetings.
Meeting transcriptions	Transcriptions of tapes recorded during SPI Team meetings and performance measurement group meetings.
Electronic messages	Includes e-mail, e-forms and other electronic exchanges among team members. Does not include telephone conversations.
SPI artifacts	SPI documents. Includes improvement plans, project plans, memos, and design documents.
Company artifacts	General documents describing the company, e.g., newsletters, web page, financial statements, etc.
Maturity assessments	CMM and BOOTSTRAP assessments uncovering several strengths and weaknesses in the firm's software development process.
Performance measurement results	Results from the performance measurement program, including estimates, completion dates, satisfaction surveys, etc. (see Table 3).

3. The organizational change perspective

Since SPI attempts to fundamentally alter the way the organization develops software, it is relevant to study SPI under the lens of an accepted organizational change framework. The theoretical framework for organizational change upon which our study is based was adapted from Applegate (1994), who suggests that to achieve effective and lasting organizational change, change agents must attend to more than just management and technical systems. To transform taken-for-granted work routines, behaviors and cultural habits, organizational actors must continuously learn and change. Therefore, change agents must attend to social and cultural issues, which are often the most difficult to address.

Applegate (1994) outlines an integrative longitudinal process model for organizational change that addresses the following three dimensions:

(a) *Aspects of organizational change*, which defines those aspects of the organization to which any organizational change initiative must attend. These are people, organizing structures, management and technical systems. In general, when the goal is fundamental change (competitiveness, innovation, performance, etc.), an episode of organizational change attempts to transform each aspect of the organization.
(b) *The change episode*, which concerns change mechanisms and the management of the change process. Here the change agent must attend to the change mechanisms (in our case, SPI and measurements) and important issues in managing change, such as vision, leadership, organizing, decision-making, and problem solving.
(c) *Organizational effectiveness measures*, which concerns what to measure to provide information upon which to base management decisions.

We explore these three dimensions further in the next sections.

3.1. Aspects of organizational change

Organizational change initiatives generally focus on changing various aspects of the organization from one state to another. For example, SPI seeks to transform software development practices, improving quality and

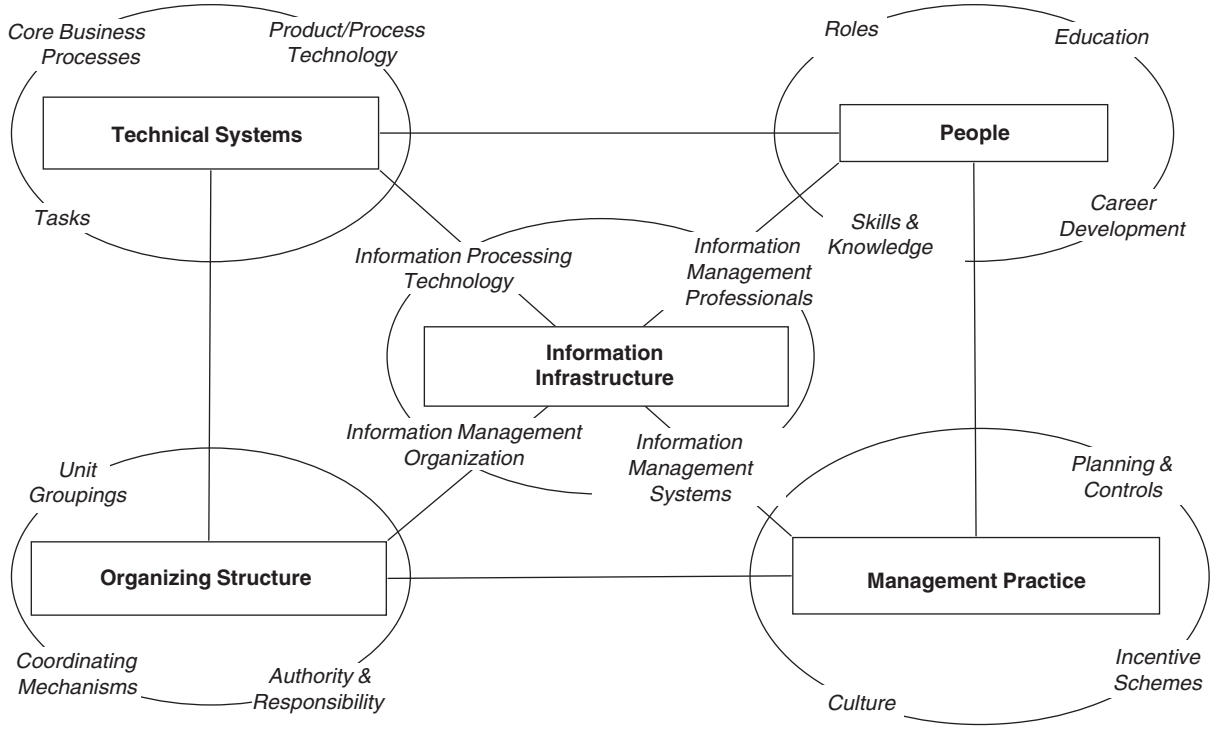

Fig. 1. Organizational framework (Applegate, 1994).

reliability, productivity, and customer and employee satisfaction. Viewed from an organizational change theory perspective, such a transformation of the work practices of a software development organization necessitates changes in all four aspects of the organization: people, technical systems, organizing structure, and management practice (Fig. 1).

Transforming the people aspect requires the change agent to address changes in role responsibilities, developing new skills and knowledge, career development, and education. New role responsibilities must be clearly defined because stability of the work processes depends on them. Everyone in the organization must clearly understand the new role they are to play. To assist in the transition, change agents must also define the new skills and knowledge necessary to carry out new roles, and devise education programs to support the development of these.

Technical systems comprise core business processes, product/process technologies and tasks. Core business processes and tasks in software development include development processes such as design, programming, and testing. Product/process technologies include development methodologies, configuration management tools, and so on. Changes in these need to be carefully planned and executed.

Organizing structure refers to how the workgroup conducting software development is organized, how work activities are coordinated and managed, and how authority and responsibility are shared in the work group. Fundamental organizational change often necessitates redesign of the structures upon which work is organized and carried out. Finally, management practice concerns shaping and maintaining organizational culture, how planning and control is carried out in the organization, and incentive schemes for inducing changes in behavior and high level of performance.

3.2. The change episode

Successful change management requires careful attention to selecting change mechanisms and managing the change process (Kanter, 1983; Quinn, 1996). In SPI, the change mechanisms are well defined in such methodologies as CMM (Paulk, Curtis, Chrissis, & Weber, 1993b) and IDEAL (McFeeley, 1996). However,

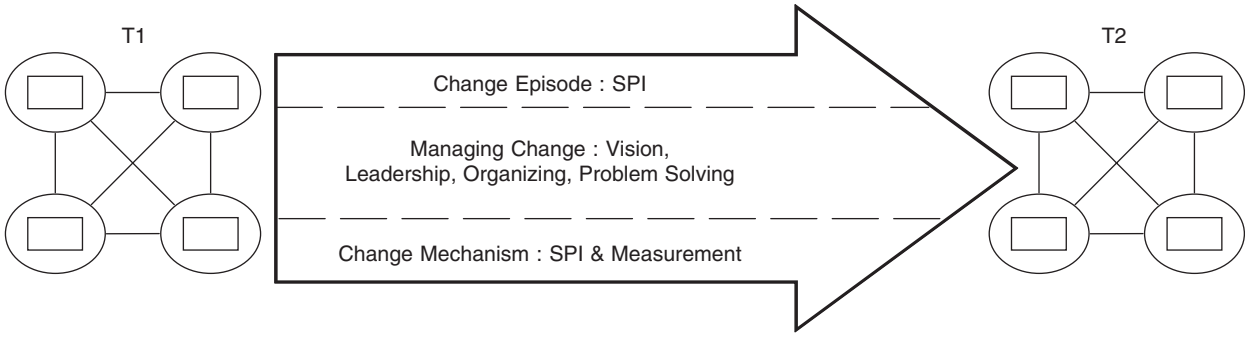

Fig. 2. Longitudinal perspective of organizational change (Applegate, 1994).

the main concerns of managing the change process are strategic vision, leadership, organizing for change and problem solving, as illustrated in Fig. 2, which shows an organization at the time T1, which is transformed through the change episode into the organization at the time T2.

Senior management is responsible for defining and articulating the vision of the new organization, shaping its culture, defining parameters of success, and defining how it will interact with key stakeholders. Another important issue of change is leadership, which deals with developing and shaping management style, communication, and social interaction within the organization. These social aspects of the organizational environment are often essential to ensuring the viability of change initiatives.

Two other important management issues of the change episode are organizing and problem solving. How the change initiative is organized and the perceived power, status and responsibility of the change agents will affect its success. The change team needs to be visible and credible. This requires that individuals with recognized expertise and company-wide recognition be members of the change team. Failure to appoint such persons to the team could send a signal to the rest of the organization that senior management is not committed to fundamental change. Problem-solving strategies also need to be well thought out. Does the change team hand down the decisions and solutions, or do they involve employees in the problem-solving and decision-making process? These important issues need to be addressed before the change process starts. Senior management needs to determine how the change initiative will be organized before it is announced.

3.3. Organizational effectiveness measures

Applegate (1994) suggests that to measure progress in organizational change initiatives, it is necessary to benchmark initial performance and to conduct interval evaluations of process performance, stakeholder satisfaction and results. Benchmarking is a systematic approach to continuously compare and measure a firm's performance against leading firms in the industry. It also can be done internally, whereby one development process can be compared with another highly successful development process. Benchmarking is intended to help the firm identify, implement, and perfect the best practices in the industry. In some cases, benchmarking may be very difficult or very expensive. In these cases, we can use baseline measurements, where the measurements are evaluated relative to a fixed norm or to previous measurement results that have been given the status of a baseline. An example of such a baseline could be CMM, which is used for assessing the maturity of a software development process.

The first activity of any measurement program is to perform an assessment of the current performance before the change initiative starts. After the change initiative has been implemented periodic assessments can tell managers how they are progressing toward the stated goals. The three areas of interest in measuring organizational effectiveness are process performance, stakeholder satisfaction, and results. For example, senior managers might be interested in how the software quality assurance process is performing. We can measure this both in terms of impact on key stakeholders such as customers, and in terms of results such as cost, number of software defects, and time-to-market. Table 2 illustrates some aspects of organizational effectiveness that senior managers might want to measure.

Table 2
Organizational effectiveness measures (Applegate, 1994)

Process performance	Stakeholder satisfaction	Results
• Time • Quality • Cost • Innovation	• Customer/supplier satisfaction • Employee satisfaction	• Market measures • Financial measures

Competitive benchmarks.

4. The case organization, Danske Data

Danske Data was a subsidiary of Den Danske Bank Group, and their primary business function was the development of business management software for Den Danske Bank Group. Danske Data had expertise in the development of banking, insurance, mortgage, and financing applications. They had approximately 850 employees located at four geographically dispersed development centers. Projects had varied sizes; most projects were small and short-term, with 3–5 people on the project team for 6–12 months, but some were major projects with strategic implication for the entire corporation. This included the Y2K project and a project to enable the central systems to run in foreign branches. Danske Data mainly developed systems for the central mainframe platform. Security and reliability were the main requirements of the systems, as data were mirrored in real time between the two operation centers in Århus and Copenhagen. The remainder of this section discusses various aspects of the organization's SPI program. We use the framework presented in Fig. 2 to structure the discussion.

4.1. Strategic vision for SPI

The primary goal for the SPI effort at Danske Data was to improve productivity by 10%. At a press conference that initiated the SPI research project, the Senior Executive Vice President of Danske Data stated,

> … in Danske Data we count on improving efficiency by at least 10% through this project […]. That is the equivalent of 30 million DKK.

Every employee of Danske Data was made aware of this goal, and of the Senior Executive Vice President's concern for the SPI initiative. A second goal of the SPI initiative was to move Danske Data from level 1 to level 2 on the CMM scale.

4.2. Leadership and management

Danske Data had a management hierarchy consisting of four levels: project manager, Vice President, Senior Vice President, and Senior Executive Vice President. Project managers were in charge of regular projects, whereas Vice Presidents managed high-profile projects. It was recognized from the beginning of the SPI effort that leadership would be very important to the success of the effort. An experienced Vice President was appointed project manager, and other Vice Presidents served as team members. The idea was that each division would have one representative on the project team. Although several of the appointed representatives were not active on the project, those who did participate were committed to the project and well respected in the rest of the organization. The three Senior Vice Presidents and the Senior Executive Vice President made up the steering committee, further adding to the image of an extremely high-profile, organization-wide project.

4.3. Organization of the SPI initiative

Fig. 3 shows how the organization of the SPI project was structured around an improvement group consisting of middle managers (Vice Presidents). The SPI project manager had been very skillful in the politics

of the organization and had managed to get the project off the ground and gain commitment from key stakeholders. The researchers and consultants were considered to be active participants in the improvement group, although they were not quite as active as the internal members of the group. The group of SPI practitioners, researchers, and consultants was named SPI Project Team.

Since Danske Data was at CMM level 1 (Iversen, Johansen, Nielsen, & Pries-Heje, 1998), the improvement group did not function as an actual software engineering process group (Fowler & Rifkin, 1990), but acted more as a catalyst for the actual improvement effort. Improvement projects, each of which dealt with one improvement area, made the detailed decisions on how and what to improve. The improvement projects were staffed with well-respected people who were knowledgeable in the area to be improved. Although the SPI project had a very high profile in the organization, management support was not always as strong as desired by SPI practitioners. For instance, in staffing the improvement projects, it was difficult to get the desired number of people, and the right people at the right time, causing delays in the improvement plans.

4.4. Problem solving

As in most highly visible organizational programs, decisions concerning the SPI initiative involved a good deal of social influence and negotiation. To make decisions and have them implemented, the SPI project team obtained support from key actors in the organization. This resulted in a lengthy deliberation process, but when agreement was achieved, there was usually solid organizational commitment to the decision. The process of acquiring team members and other resources for the improvement projects clearly illustrates this. The conventional way of ensuring team members for a project was to request the relevant Senior Vice Presidents to release a particular individual for some period of time. The Vice President then decided if that individual best fit the project, or alternatively freed up another individual. However, the improvement group realized early on that the success of the improvement projects would depend on having experienced and qualified people involved. So they started out by talking to the people that they knew would be able to do the job and be interested in doing it, in order for them to put pressure on their Vice President. A formal request to the VP was made only after the support of the desired people was ensured. This strategy was successful in some cases, but in others the Vice Presidents felt unable to free the desired resource due to other obligations.

4.5. Measurements

A central theme of the SPI effort was to increase the role and significance that measurements play in the organization. It was realized early that one of the main distinguishing features of this effort, compared to previous improvement efforts, was the reliance on valid and reliable data collection methods. A number of measurement and assessment activities were developed and conducted.

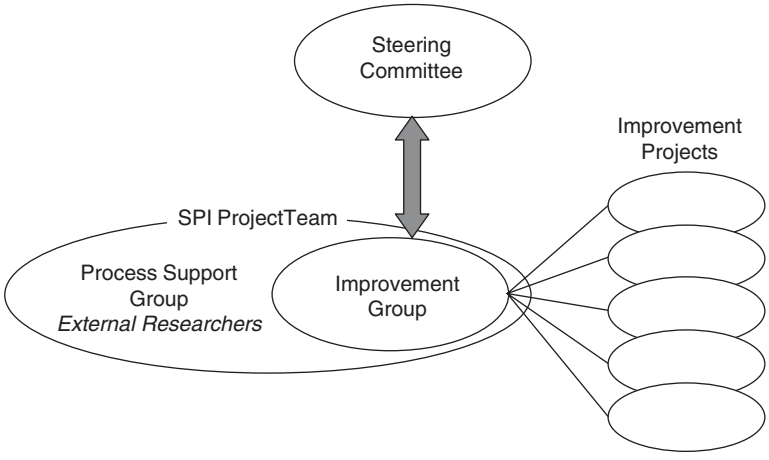

Fig. 3. Organization of the SPI project in Danske Data.

Table 3
Indicators of the measurement program in Danske Data

Indicator	Definition
Productivity	Resources used to develop the system relative to volume of project (hours/FP)
Quality	Number of error reports, both absolute and relative to volume of project (errors/FP)
Adherence to schedule	Variation from agreed time of delivery, both absolute and relative to volume of project (hours)
Adherence to budget	Variation from estimated use of resources (DKK)
Customer satisfaction	Satisfaction with the development process and the implemented solution (questionnaire)
Employee satisfaction	Satisfaction with the development process (questionnaire)

A large portion of the efforts was focused on how to demonstrate reliably that the overall 10% improvement goal was achieved. The SPI project team decided to put a measurement program into place that would measure productivity, along with a number of other factors related to the software development process (e.g., quality, adherence to schedule, customer satisfaction). There were two reasons for introducing these other measures: (1) productivity improvements should not be achieved at the expense of other vital areas, and (2) if the desired improvements in productivity were not achieved, the improvement group would have other measures to show potential successes. Introducing a measurement program was also perceived as introducing a best practice into the organization, and it thus became an improvement area in itself.

Table 3 lists the specific data that were collected in the measurement activity. The analysis of these data was expected to show if, and to what extent, the improvement activities were having the intended effects on the organizational performance. The SPI Team decided to collect data after development projects were completed in order not to disturb the projects unduly. Results were reported quarterly. It was thus not possible to use the measurements to improve an ongoing project, but the project managers were able to use the results to improve how they would run their next project.

The first measurement report covered projects completed in 3rd quarter 1997, and was seen as an experiment, with information only on productivity, adherence to schedule, and adherence to budget. The layout of the report left a lot to be desired. The second report, covering projects completed in 4th quarter 1997, improved considerably over the first, adding customer and employee satisfaction. However, it became clear that the automatic function point counts were flawed, so productivity had to be excluded from the report. These problems, as well as concerns over the data quality, caused senior management to withhold the first two reports from public availability to the rest of the organization. This caused the SPI Team to suspend the measurements while initiating a formal improvement project for the measurement program. The third report, covering projects completed in 2nd quarter 1999, had sufficient data quality to warrant a wider distribution. Since then, all reports have been available to all developers and managers in Danske Data. However, the problem of a reliable size measurement still was not solved, and the measurements were thus unable to provide information that would allow the SPI Team or senior management to declare victory on the 10% productivity improvement goal.

Table 4 lists all the detailed measurements in the measurement framework provided by Applegate (1994). Since Danske Data has only a single customer, who in turn primarily uses Danske Data for its IT services, market measures were replaced by the performance measurements that are most relevant to the customer. The table shows that the measurement program did not cover process performance. This aspect was, however, covered by several other initiatives in the organization (Andersen, Arent, Bang, & Iversen, 2002; Iversen et al., 1998).

5. Research findings

A careful analysis of the empirical data reveals the key problems that managers, practitioners, and the SPI group encountered with the measurement program. Although the team members considered the measurement program to have a great value for the organization, they encountered several relevant problems. In the following, use the Applegate framework (Fig. 1) to ensure broad coverage, as we categorize and discuss the problems that the SPI Team encountered during the change initiative.

Table 4
Measures in Danske Data

Strategic focus of measures		
Process performance	Stakeholder satisfaction	Results
• Specific process areas of 1st and 2nd maturity assessments • Project manager self-assessments • SPI risk assessments	• Customer satisfaction • Employee satisfaction	• Adherence to schedule • Adherence to budget • Productivity • Quality

Baseline: 1st maturity assessment.
Benchmark: 1st and 2nd maturity assessments.

5.1. Technical systems: size measurement

Technical systems are concerned with the concrete tasks involved in conducting change efforts. A number of the indicators rely on an accurate measurement of the output of a software development project (e.g., price, quality, and adherence to schedule). This typically is measured as the size or complexity of the software that the project produced. Early in the project, DD management selected function point analysis (IFPUG, 1994) as the measurement method. However, to minimize the disturbance of the development projects, it was decided to automatically calculate the function points. This was possible only because it also had been decided to measure the project size after the project had been completed. The results of initial experiments looked promising enough to be included in the first performance measurement report (3Q97).

There were problems, though. One was related to enhancement projects that continued work on an existing system. Such projects were accredited the entire function point count of each module they modified, even if the modifications were miniscule, and thus would appear to have an unrealistically high productivity. Another problem was that the function points were counted across disparate development and application environments. The IFPUG definition allows for normalizing such differences between projects by using the value adjustment factor (VAF), which can normalize the function point count by up to $\pm 35\%$. Although IFPUG recommends counting and comparing only within homogenous environments when VAF is not used, Danske Data decided not to use VAF due to the complexities involved in implementing it in the automated solution.

On recognizing the difficulties with attaining accurate and acceptable size measurements, the team compared a number of approaches to deriving size measures, but all were rejected, because none were viewed as providing accurate measurements. Consequently, the SPI project team eliminated the size measurement from the overall measurement program and decided to concentrate the effort in the improvement program on the indicators that did not require the use of a size measure. Excluding a size measure seriously impeded reaching the original objective of measuring efficiency and productivity, as there was no longer a measure of the output of the software projects.

5.2. People: customer and employee satisfaction

To measure stakeholder satisfaction, two questionnaires were developed to elicit data from the project team members and customers after project completion.

> So far 8 groups have answered the questionnaire [on customer and employee satisfaction], and we have received some criticism on the questionnaire and the way in which feedback has been given. We need to think a little more about that and make a concept to give standardized feedback. (Measurement Project Manager, SPI Team meeting, January 22, 1998)

The problems with the questionnaires were related primarily to the customer questionnaire. A small analysis was conducted to determine the nature of the problems:

> We send questionnaires to project managers and ask them to administer them to those people who were involved in the acceptance test of the system. And what we then ask are managerial, contractual, overall process-related questions relating to how the project was conducted. Then some random tester has to answer if commitments have been met. He hasn't seen the contract or anything. That's bound to go wrong, and that's why management can't recognize reality in these numbers. (Vice President, SPI Team meeting, April 22, 1998)

The questionnaire was subsequently revised to include perspectives of both the competent decision-maker (the customer) and those users who participate in the actual development process.

5.3. Management practice: data discipline

After the disappointment that the second measurement report could not be published, the SPI group decided to establish an improvement project to improve the quality of the measurement report so much that it would be impossible for management to deny making it public. In June 1998, the performance measurement program was established as a separate project under the SPI project. It was given its own resources and staffing. The main success criterion for the project was that a measurement report should be completed in April 1999, containing data on all six indicators and from all projects completed in 1Q99. Compared to the second report, this report should have improved the measurement process for all the indicators, and the layout of the report also should be enhanced. Among other things, the data quality of each indicator should be displayed, for instance as a percentage of complete data sets (as in Fig. 4).

Only 10 projects had reported complete data in 1Q99, and the data they had reported could not necessarily be trusted. It was therefore decided not to make a measurement report, but instead produce a memo to senior management describing the problems with data discipline and recommending a strategy for alleviating the problems. On September 21, 1999, the measurement report for 2Q99 was presented to management. The report contained data from 51 of the 80 projects that completed in the period. The report stated that: "*[This] is the best data foundation ever. The report's result should therefore provide a credible picture of the actual state of affairs in FSS*". The report also contained a word of caution: "*The data discipline should not only be maintained but should be clearly improved to ensure the data foundation for the future measurements [...] but commitment from the management hierarchy is required to reach a better result*".

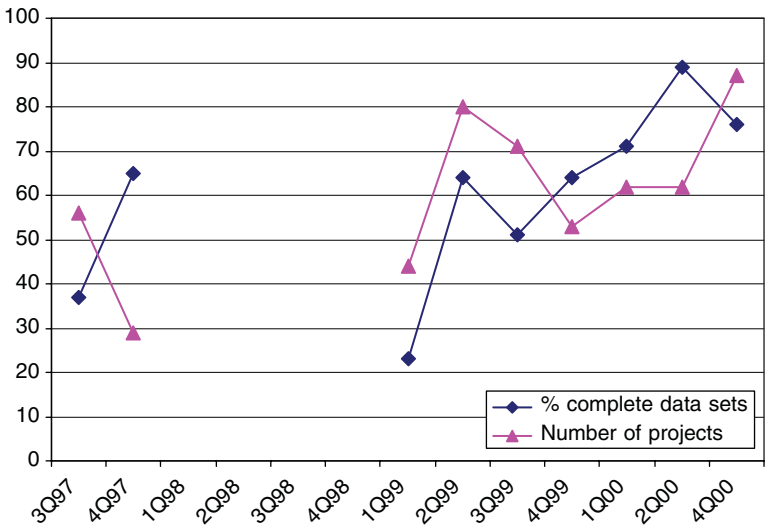

Fig. 4. Data discipline in the performance measurement program. No measurements were taken in 1998.

Management decided to distribute the next report (3Q99) to project managers, and every report from 4Q99 on to all employees of FSS. The metrics team saw this decision as a step in the right direction, and it expressed the kind of recognition that the group had hoped for since management withheld the first measurement report in September 1997.

Fig. 4 shows the number of projects completed in each quarter, along with the percentage of those projects that submitted complete data to the measurement program. As the figure shows, there was a definite improvement in data discipline over the period.

A recurring problem with the performance measurement program was to get project managers and developers to report accurate and complete information in the time registration database. Two of the initial problems were that this information was not previously used for anything, and the user interface for the application used to enter the data was highly confusing, giving rise to many wrong entries. To alleviate this problem, project managers were educated on the use of the system and how the data were processed and used. Moreover, as part of the project to improve the measurements, the team planned to improve the user interface of the time registration system to avoid misunderstandings. The following quote illustrates the problem:

> Some of the [improvement in] data discipline is due to the information meetings. People are aware that this is actually used for something. Of course, it can still get better. Ann-Dorte actually e-mails everybody prior to each measurement: ''Do remember to add data to POKA [project registration tool]'', and it is actually only 2 fields in relation to this that are interesting. I know that when you create a project, there are at least 50 fields to fill out. (SPI Team Project Manager, SPI Team meeting, April 22, 1998)

5.4. Management practice: using the data

A major problem with the measurement program was that the data were not used for anything. No one trusted the data quality sufficiently to rely on the measurement reports for making hard decisions. For these purposes, the qualitative data that came out of the quality review meetings for every project were considered far better.

> We … first had to make people used to registering the necessary data, and only when this was 'back bone' could we start improving the validity. …We actually did not manage to get valid data, as we spent too much time just getting people to register data…. If you don't want to be fighting against windmills you have to make registering the necessary data a natural and unwavering part of the project.
>
> We experienced that the measurement program lost the attention of management, partly because the novelty disappeared, and because they (and we) couldn't use the data for anything, as we continued to doubt the causes of what we saw in the results, and the lacking data discipline was still an issue, that made us uncertain what we could conclude based on the quantitative data. (SPI Team and Measurement Team member, Email, May 28, 2001)

5.5. Organizing structure: project organization

One of the problems in establishing a high-quality measurement program was how to organize the effort. In the initial phase (approximately 3–5 months) when the program was being established, it was run as a very loosely structured project, with participants from several geographical locations. Very little formal project management was applied to the project. After the regular measurements started in June 1997, one team member was responsible for collecting and analyzing the data, and producing the first two reports. After the second report was completed, but neither report was approved for publication in the organization, the SPI Team recognized at a meeting in April 1998, that it would be necessary to increase staffing and impose more formal project management practices on the performance measurement program. This would include making a project contract, project plans, etc. This turned out to be a very positive development for the project, as it allowed the team members to spend more time on working on the details of the measurement program. Ultimately, this increased attention led to the successful publication of all quarterly measurement reports, beginning in 3Q99.

6. Discussion

The Danske Data case illustrates many of the problems in implementing an SPI measurement program. SPI measurement programs not only encompass technical issues of measurement and accuracy; they also involve such issues as organizational culture and policy, management practices, resource allocation, and stakeholder influence processes (Iversen & Mathiassen, 2003). For example, while the problems of measuring project size and the accuracy of the function point measurements are related primarily to the technical systems issues, problems of data discipline are organizational and cultural issues, which, in turn, are related to the roles and the distribution of responsibilities among the development team members. The problem of publishing the reports is another example of an organizational culture issue that also relates to management practice and stakeholder influence.

Even in a highly successful company dedicated to improvement and quality, such as Danske Data, it is difficult to develop and implement a systematic measurement program. A comprehensive and authoritative framework can be useful in providing a structure and level of legitimacy to measurement programs. This is important because measurement programs are highly contentious organizational issues affecting a variety of stakeholders. Evidence from the case study suggests that conflicts about measurement goals can often influence perceptions of success or failure on SPI initiatives.

Another important issue is how to determine the effectiveness of an SPI measurement program. Our research suggests that it is necessary to measure the effectiveness of the measurement program. After some time, the SPI Team of Danske realized that this was necessary and attempted to address the issue by introducing a metric for the effectiveness of data discipline (percentage of projects with complete data sets—c.f. Fig. 4). Such meta-indicators are crucial to understanding the source of the problems of the measurement program.

Conflicts on measurement goals are another possible source of problems. At Danske Data, the measurements were never intended to provide significant benefits to the project managers who were burdened with providing the data. With the first measurement reports, some efforts were made to provide feedback that would allow project managers to compare their projects to the Danske Data average. While the SPI Team viewed such feedback as crucial to the success of the program, several factors made direct and timely feedback difficult. First, data were collected for completed projects only. Then the results were only made public in quarterly reports, which meant that several months elapsed from the completion of a project until a project manager could see and compare his/her project's performance to others in the organization. Further, only summary information (such as averages) were published in the reports. Consequently, it was difficult for the project managers to conduct a detailed comparative analysis. A second contributing factor here is the organizational culture. While senior managers at Danske Data wanted to use the measurement program to improve performance, they were uncomfortable with the potential for competitive behavior among the project managers. Such behavior was seen as dysfunctional to organizational cooperation.

Lessons:

- understand the tradeoff between disturbing the projects and providing value to them,
- recognize the synergy between making data public and the accuracy and completeness of data sets,
- need to consider how a measurement program affects many parts of an organization (i.e., five categories in Applegate),
- understand the cultural implications of measurement frameworks,
- need to understand how to manage change programs (four parts in Fig. 2: vision, leadership, organizing, problem solving).

Measuring the effectiveness of SPI change initiatives is a complex undertaking, as these initiatives target various dimensions of the organization. Evidence from this longitudinal study suggests that the development, implementation and success of such a measurement program can benefit from a comprehensive authoritative framework. Such a framework could help guide SPI managers in their work, saving much time and cost. It can also enable early convergence on the central issues of the various stakeholders. Currently, no comprehensive

measurement framework exists in the SPI literature. Our research suggests that although Applegate's framework for organizational change (Applegate, 1994) is general, it does address the critical dimensions SPI change. As such, it can be adopted and specialized for software organizations. The primary strength of the Applegate framework is that it has been tested in field studies in 110 companies of all sizes and industry segments that implemented a range of organizational improvement programs (Applegate, 1994). Early assessments from our studies of SPI suggest that it can be useful for guiding both field studies and research and development of SPI measurement programs.

Acknowledgments

This research was supported by The Danish National Center for IT Research. Without the support and assistance of Danske Data, this research would not have been possible. Also thanks to all the employees of Danske Data who directly and indirectly provided assistance in the making of this paper.

References

Andersen, C. V., Arent, J., Bang, S., & Iversen, J. H. (2002). Project assessments. In L. Mathiassen, J. Pries-Heje, & O. Ngwenyama (Eds.), *Improving software organizations: From principles to practice* (pp. 167–184). Upper Saddle River, NJ: Addison-Wesley.

Applegate, L. M. (1994). Managing in an information age: Transforming the organization for the 1990s. In R. Baskerville, S. Smithson, O. Ngwenyama, & J. I. DeGross (Eds.), *Transforming organizations with information technology* (pp. 15–94). Amsterdam: Elsevier Science B.V.

Curtis, B. (1996). *Improving software organizations with the cmm [slide set]*. Copenhagen: Metodica.

Foster, M. (1972). An introduction to the theory and practice of action research in work organizations. *Human Relations*, *25*(6), 529–556.

Fowler, P., & Rifkin, S. (1990). *Software engineering process group guide* Technical report no. CMU/SEI-90-TR-24. Pittsburgh, PA: Software Engineering Institute.

Goldenson, D. R., Emam, K. E., Herbsleb, J., & Deephouse, C. (1996). *Empirical studies of software process assessment methods (No. ISERN-97-09)*. Kaiserslautern: Fraunhofer—Institute for Experimental Software Engineering.

Goldenson, D. R., & Herbsleb, J. D. (1995). *After the appraisal: A systematic survey of process improvement, its benefits, and factors that influence success* Technical report no. SEI-95-TR-009. Pittsburgh, PA: Software Engineering Institute.

Hansen, B., Rose, J., & Tjørnehøj, G. (2004). Prescription, description, reflection: The shape of the software process improvement field. *International Journal of Information Management*, *24*, 457–472.

Hayes, W., & Zubrow, D. (1995). *Moving on up: Data and experience doing cmm-based process improvement* Technical report no. CMU/SEI-95-TR-008. Pittsburgh, PA: Software Engineering Institute.

IFPUG. (1994). *Function point counting practices manual release 4.0*. Princeton Junction, NJ: The International Function Point Users Group (IFPUG).

Iversen, J., Johansen, J., Nielsen, P. A., & Pries-Heje, J. (1998). Combining quantitative and qualitative assessment methods in software process improvement. *Paper presented at the European Conference on Information Systems (ECIS 98)*, Aix-en-Provence, France.

Iversen, J. H., & Mathiassen, L. (2003). Cultivation and engineering of a software metrics program. *Information Systems Journal*, *13*(1), 3–20.

Kanter, R. M. (1983). *The change masters: Innovation for productivity in the American corporation*. New York: Erwin.

Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. Thousand Oaks, CA: Sage Publications.

Mathiassen, L. (1998). Reflective systems development. *Scandinavian Journal of Information Systems*, *10*(1–2), 67–118.

Mathiassen, L., Pries-Heje, J., & Ngwenyama, O. (Eds.). (2002). *Improving software organizations: From principles to practice*. Upper Saddle River, NJ: Addison-Wesley.

McFeeley, B. (1996). *Ideal: a user's guide for software process improvement* Technical report no. CMU/SEI-96-HB-001. Pittsburgh, PA: Software Engineering Institute.

Monge, P. R. (1990). Theoretical and analytical issues in studying organizational processes. *Organization Science*, *1*(4), 406–430.

Paulk, M. C., Curtis, B., Chrissis, M. B., & Weber, C. V. (1993a). *Capability maturity model for software, version 1.1* Technical report no. 93-TR-024. Pittsburgh, PA: Software Engineering Institute.

Paulk, M. C., Curtis, B., Chrissis, M. B., & Weber, C. V. (1993b). Capability maturity model, version 1.1. *IEEE Software*, *10*(4), 18–27.

Pettigrew, A. M. (1985). Contextualist research and the study of organizational change processes. In E. Mumford (Ed.), *Research methods in information systems*. Amsterdam: North Holland.

Quinn, R. (1996). *Deep change*. San Francisco: Jossey Bass.

Jakob H. Iversen holds an M.Sc. in software engineering and a Ph.D. in computer science from Aalborg University, Denmark. He is currently an Assistant Professor of MIS at the University of Wisconsin Oshkosh. His research focuses on agile software development, software process improvement and measurement. He is a co-author of *Improving Software Organizations: From Principles to Practice*.

Ojelanki K. Ngwenyama is Professor of Information Technology Management at Ryerson University, Toronto, Canada. He earned his Ph.D. from the Thomas J. Watson School of Engineering, SUNY-Binghamton. He is on leave from his position as Professor of Information Systems and Director of the International Scholars Program at Virginia Commonwealth University, Richmond, VI, USA. He is also a Visiting Research Professor at Aalborg University, Denmark and Docent at University of Jyväskylä, Finland. Ojelanki has also served on the faculties of University of Michigan Business School, USA; The Rotterdam School of Management, Erasmus University, The Netherlands; University of Pretoria, South Africa; York University, Toronto. His current research focuses on organizational issues of information technology. His papers have appeared in a range of international journals.