An Empirical Study of Industrial Requirements Engineering Process Assessment and Improvement

IAN SOMMERVILLE and JANE RANSOM
Computing Department, Lancaster University

This article describes an empirical study in industry of requirements engineering process maturity assessment and improvement. Our aims were to evaluate a requirements engineering process maturity model and to assess if improvements in requirements engineering process maturity lead to business improvements. We first briefly describe the process maturity model that we used and modifications to this model to accommodate process improvement. We present initial maturity assessment results for nine companies, describe how process improvements were selected and present data on how RE process maturity changed after these improvements were introduced. We discuss how business benefits were assessed and the difficulties of relating process maturity improvements to these business benefits. All companies reported business benefits and satisfaction with their participation in the study. Our conclusions are that the RE process maturity model is useful in supporting maturity assessment and in identifying process improvements and there is some evidence to suggest that process improvement leads to business benefits. However, whether these business benefits were a consequence of the changes to the RE process or whether these benefits resulted from side-effects of the study such as greater self-awareness of business processes remains an open question.

Categories and Subject Descriptors: D.2.1 [**Software Engineering**]: Requirements/Specifications; D.2.8 [**Software Engineering**]: Metrics; D.2.9 [**Software Engineering**]: Management

General Terms: Experimentation, Measurement

Additional Key Words and Phrases: Software process improvement, process measurement, requirements engineering, empirical software engineering

1. INTRODUCTION

The notion of process improvement and consequent business benefits through increased product quality and productivity is rooted in the work of Deming [1982] in his work on statistical quality control. Deming's approach was geared

This research was partially funded by the European Commission's Framework 5 research program under contract number IST-2000-28760.
Authors' address: Computing Department, Infolab21, South Drive, Lancaster University, Lancaster LA1 4WA, UK; email: {is,bjr}@comp.lancs.ac.uk.

towards repeatability so that process performance was not dependent on individuals but was predictable, so that repeating the same work would produce, more or less, the same result. To achieve better results, the process therefore had to be improved and Deming successfully introduced a measurement-improvement cycle for industrial processes.

Deming's approach was adopted by Humphrey [1989] and others at the Software Engineering Institute in the 1980s, and they developed it for software process improvement. The seminal results of this work were the notion of software process maturity and an identification of a process maturity model with discrete levels of process maturity. This was an essential basis for the development of the Software Capability Maturity Model (CMM), described by Paulk [Paulk et al. 1993, 1995] which allows the maturity status of software development organizations to be assessed. Over time, the CMM has evolved from an assessment model to an essential component of an integrated approach to software process improvement [Ahern et al. 2001].

The general notion of a process maturity model has been extended at the SEI into other areas such as personnel assessment and systems engineering. The concept has also been taken up and incorporated into other process improvement models such as the European BOOTSTRAP model [Haase et al. 1994; Koch 1993; Kuvaja et al. 1994] and the Canadian TRILLIUM model [Coallier 1999]. The SPICE project [El Emam et al. 1997; Konrad and Paulk 1995, Paulk and Konrad 1994] is a continuing initiative to develop an international standard for software process improvement where a continuous improvement model is suggested that pulls together the different models that have been proposed or are in use. The CMMI model [Ahern et al. 2001] also integrates different models and has both staged and continuous versions.

Our previous work in RE process improvement resulted in the development of a maturity model for requirements engineering processes analogous to the CMM model for software processes [Sawyer et al. 1998, 1999a, 1999b; Sommerville and Sawyer 1997]. Like all other approaches to process maturity assessment, our work has been inspired by the initial concepts developed by Humphrey; but we have focused on requirements engineering (RE) processes—an area that has been mostly outside the scope of other process improvement models. Our rationale for concentrating on this early phase of the software process was that problems in this area have a profound effect on system development costs and functionality and that a very large number of organizations have poorly defined and informal RE processes.

System requirements engineering problems are universal and very significant indeed. In an early paper, Boehm [1983] suggested that errors in system requirements could be up to 100 times more expensive to fix than errors introduced during system implementation. Lutz [1993] showed that 60% of errors in critical systems were the results of requirements errors. In 1996, a survey of European companies found that more than 60% of them considered requirements engineering problems as very significant [Espiti 1996]. Recently, Hall et al. [2002] carried out a case study of 12 companies at different levels of capability as measured by the CMM. They discovered that, out of a total of 268 development problems cited, almost 50% (128) were requirements problems.

Company	Domain	Primary system type	Development mode
Company 1	Automation	Embedded	In-house
Company 2	Engineering supplies	Business support	Outsourced
Company 3	Electronic supplies	E-commerce	In-house/outsourced
Company 4	Wholesale/retail	E-commerce/e-service	Outsourced
Company 5	Aerospace	Embedded	In-house
Company 6	Media	E-service	In-house/outsourced
Company 7	Food processing	Business support	Outsourced
Company 8	Textiles	E-commerce	In house/Outsourced
Company 9	Building supplies	E-commerce (B to B)	In house/Outsourced

Fig. 1. The companies involved.

Clearly, significant benefits can accrue from improving the quality of requirements and, by implication, requirements engineering processes.

The study that we describe here was part of a European project called IMPRESSION, which had two related goals. First, it was intended to help organizations that were dependent on software to improve their requirements engineering processes. Second, and more broadly, the project wished to investigate whether or not improvements in business performance could be correlated to improvements in RE processes.

The IMPRESSION consortium consisted of two British universities, a Greek software development and consultancy company (ATC) and nine application users, each with systems at the requirements stage of development. These users (referred to here as Company 1 to Company 9) were also based in Greece and the systems they were developing covered embedded, e-commerce and e-service application areas. Some application users did all their development in-house, some used a mix of in-house and outsourced development and, for three companies, their software development was entirely outsourced. None had previous experience of formalized software process improvement. Application users received no direct funding from the project sponsors although they benefited from the consultancy provided by other partners. Figure 1 provides some summary information about the companies involved in our study.

Our role in the project was to provide advice on how the RE processes in the application users could be improved. Other partners defined and monitored business performance indicators, assessed the suitability of RE tools and techniques for the application users and facilitated the implementation of our recommendations into the application users' RE processes.

Our remit to propose RE process improvements gave us the opportunity to evaluate the RE process maturity model, the assessment process and the overall value of RE process improvement. We hoped to answer the following questions:

(1) What are the practical difficulties that arise in RE process maturity assessment?

(2) Is the RE process maturity model a useful basis for an RE process improvement initiative?
(3) Do improvements in RE process maturity lead to real business benefits?

The work described here is novel in that it represents the first systematic evaluation of the RE process maturity model across a range of companies. While the model has been used, both by ourselves and others, within individual companies the degree of formality with which it has been applied has been very variable. We believe that, while it has had a significant influence on some RE process improvement initiatives, it has not previously been used by us in a systematic way where "before and after" maturity assessments were carried out.

The remainder of the article focuses on our requirements maturity assessment work in the IMPRESSION project. We summarize our requirements engineering maturity assessment process, discuss the application of this model and how it was used to propose process improvements. We then present the results of a re-application of the assessment, reflect on the success of our process improvement proposals and discuss how we tried to assess business benefits from process improvement. Finally, we discuss the lessons learned from this study for the requirements engineering process maturity model and, more generally, for requirements engineering process improvement.

2. THE REQUIREMENTS PROCESS MATURITY MODEL

In a previous project (REAIMS), concerned with the improvement of processes for safety-critical systems development, we developed a maturity model for requirements engineering processes and an associated method for process assessment. This model and associated research has been documented in a book [Sommerville and Sawyer 1997] and in a number of papers [Sawyer et al. 1998, 1999a, 1999b]. This article is primarily concerned with the use and evaluation of the model rather than its development. Consequently, we present only a brief summary of the RE process maturity model to provide a context for the description of the process improvement study.

The requirements engineering maturity assessment method that we developed helps organizations to compare their RE processes to industrial good practice and to identify possible process improvements. At the heart of our method is a process maturity model for requirements engineering that complements the SEI's Capability Maturity Model, which does not cover requirements engineering processes. As many of the companies that we worked with were aware of the work of the SEI, we decided to use CMM-compatible terminology and therefore proposed three levels of RE process maturity corresponding to the first three levels in the CMM:

Initial. The requirements engineering processes are inadequate, and depend on the skills and experience of individual engineers. Timescales and budgets are often not met, requirements are not properly managed and requirements documents may be nonexistent or poorly structured and inconsistent.

Repeatable. Companies have explicit standards for requirements documents and descriptions and have introduced policies and repeatable procedures for requirements management. Basic cost and scheduling management procedures are in place and some advanced requirements engineering tools and techniques may be used.

Defined. The RE process for both management and engineering are documented, standardised and integrated into a standard process that enables active process improvement.

We chose three rather than five levels as in the CMM because our initial work in formulating the maturity model suggested that there were so few companies with defined RE processes that including the higher CMM levels (Managed and Optimizing) was not worthwhile. This data gathered in the study described in this article confirmed our view that RE process maturity was generally low and that higher levels of process maturity would not, at this stage, be useful.

Essentially, our approach is based on the notion that the level of process maturity reflects the extent that good RE practices and used and standardized in a company. We identified 66 good practice guidelines, a method of assigning scores to them and an algorithm that determines the RE maturity level of an organization. These good RE practices fall into three categories, basic, intermediate and advanced:

Basic practices. There are 36 basic practices that are concerned with fundamental activities that are required to gain control of the RE process and, generally, are applicable across application domains. Most of the basic good practices are concerned with documenting, validating and managing requirements.

Intermediate practices. There are 21 intermediate practices are mostly concerned with the use of methodical approaches to requirements engineering and the use of tools. Again, they are mostly domain-independent.

Advanced practices. There are nine advanced practices that are concerned with methods such as formal specification and organizational change. The key distinction between advanced and intermediate practices are that advanced practices are mostly geared to critical systems development. This reflects the fact that the model was originally developed in collaboration with companies involved in safety-critical systems engineering.

Good practices are associated with important requirements engineering process areas as shown below (the area number reflects the chapter in our book [Sommerville and Sawyer 1997] describing these practices):

3 The Requirements Document
4 Eliciting Requirements
5 Requirements Analysis and Negotiation
6 Describing Requirements
7 System Modelling
8 Requirements Validation
9 Requirements Management
10 Critical Systems Requirements

Level	Criteria
Initial	Score of above 54 in Basic Guidelines.
Repeatable	Score of above 54 in Basic Guidelines and below 40 in (Intermediate + Advanced) Guidelines.
Defined	Score of above 54 in Basic Guidelines and above 39 in (Intermediate + Advanced) Guidelines.

Fig. 2. RE process maturity levels.

Some examples of the good requirements engineering practices that we proposed are:

3.1 Define a standard document structure (Basic)
4.3 Identify and consult system stakeholders (Basic)
6.2 Use language simply, consistently and concisely (Basic)
8.2 Organise formal requirements inspections (Basic)
4.10 Prototype poorly understood requirements (Intermediate)
9.6 Define change management policies (Intermediate)
10.6 Specify systems using formal specifications (Advanced)
10.8 Collect incident experience (Advanced)

We do not have space to discuss the details of these guidelines here. In our book [Sommerville and Sawyer 1997], each good practice guideline is described in detail, along with its expected benefits, the cost of introduction and the costs of using that guideline. Further examples of good practice guidelines are listed later in this article in Figures 9 and 10.

To assess process maturity, the assessor examines the RE processes used in a company and decides which practices that the company has adopted. This assessment cannot be carried out mechanically but requires judgment to recognize when particular ways of working correspond to the good practice recommendations. To reflect whether or not a practice has been institutionalized, the assessor allocates a *weighted score* to each practice according to its use within an organization as follows:

—Never used (Score = 0)
—Discretionary—used at the discretion of individuals (Score = 1)
—Normal—used by many teams but in different ways (Score = 2)
—Standardized—used throughout the company in a standardized way (Score = 3)

The maturity level is computed by summing these weighted scores for all Basic guidelines and for the Intermediate/Advanced guidelines. The score is therefore presented as a pair of numbers M | N where M is the weighted score for Basic guidelines and N the weighted score for Intermediate/Advanced guidelines.

The threshold levels in the model that we proposed are shown in Figure 2. These threshold levels and the scores associated with these levels were derived after extensive discussions with RE practitioners in industry and they

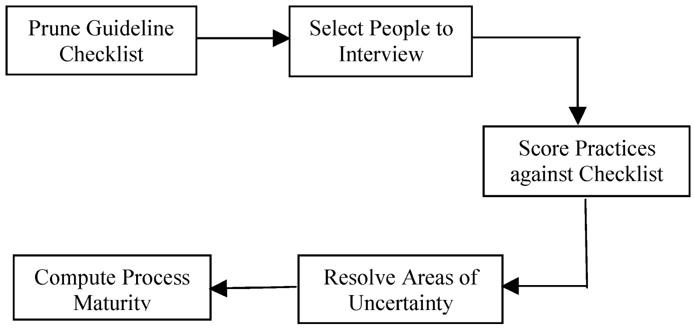

Fig. 3. Requirements engineering process maturity assessment.

reflect our judgment and experience in good requirements engineering practice. Essentially, to move from an Initial to a Repeatable level, a company must have a sound base of RE practices in place and to move from Repeatable to Defined, this must be supplemented by the use of methodical approaches.

As is now recognized in the CMMI approach [Ahern et al. 2001], the threshold levels in all discrete maturity models are arbitrary and this is certainly true of our model. Discrete models have the advantage that they establish goals for companies trying to increase their maturity but, as we discuss later, one of the results of this project was to convince us that a continuous rather than a discrete maturity model is more appropriate for most requirements engineering processes.

3. RE PROCESS MATURITY ASSESSMENT

The process maturity assessment method involves assessing the use of RE practices in an organization and using the results to determine a process maturity level. The process that we originally designed to assess the maturity level consists of five steps, as shown in Figure 3:

(1) *Prune Guideline Checklist*. Quickly identify practices that are never used in an organization (e.g., formal methods) to eliminate the corresponding guidelines from the assessment. Therefore, time is not wasted on unnecessary discussion.
(2) *Select People to Interview*. Select contacts within the organization who know the practices used and can advise on the level of usage.
(3) *Score Practices against Checklist*. Allocate each practice a score by interviewing the contacts.
(4) *Resolve Areas of Uncertainty*. Once an initial set of scores has been obtained, resolve all areas of uncertainty by further stakeholder discussion and ensure the scoring method is consistent.
(5) *Compute Process Maturity*. Compute the RE process maturity by calculating the total score and assigning a level as shown in Figure 2.

Maturity assessments are summarized in a maturity matrix as shown in Figure 4, which shows the number of good practices used and the overall maturity assessment. The maturity matrix includes columns for each of the three

	Basic	Intermediate	Advanced
Guideline Count	Basic Guideline Count	Int. Guideline Count	Adv. Guideline Count
Weighted Score	Basic Weighted Score	Int. Weighted. Score	Adv. Weighted. Score
Maximum Score	Max. score (Basic)	Max. score (Int)	Max. score (Adv)
Score %	Basic Score as a % of Maximum	Int. Score as % of Maximum	Adv. Score as a % of Maximum
Level	Assessed maturity level		

Fig. 4. The organization of a requirements maturity matrix.

categories of good practice guidelines, Basic, Intermediate and Advanced and information in each row about these guidelines:

- Row 1 shows the number of guidelines in each category that are actually used.
- Row 2 shows the weighted score, as explained above, reflecting the extent to which the guideline use is standardized in the organization.
- Row 3 shows the maximum possible score that can be attained if all guidelines in that category are used and standardized.
- Row 4 shows the percentage of this maximum actually achieved in this assessment.
- Row 5 shows the assessed maturity level (Initial, Repeatable or Defined).

The information in the maturity matrix is the basic starting point for discussions on process improvement. As we discuss in later sections, this is then supplemented by further process information to give a broader picture of the use of RE practices and possible process improvements.

4. THE IMPRESSION EXPERIMENT

The aim of our study was to collect information that would help us assess and improve the requirements engineering process maturity model. At the same time, the overall project had the goal of improving the requirements engineering processes in the application users and assessing the relationship between these improved processes and business performance. We were not directly involved in business performance measurement—other partners (City University, London) developed ways of monitoring business performance and applied this with the application users. This was done in parallel with the activities that we describe below.

The IMPRESSION project was an 18-month project with a 3-month startup and training phase followed by a 14-month assessment and improvement phase and then a final, 1-month, reporting phase. The stages of the work where we were involved were:

(1) *Method and Tool Development*. We adapted the RE process maturity assessment process to the circumstances of this project and trained consultants from ATC in the process maturity method. We also developed the IMPRESSION maturity assessment tool that supports the method (Months 1–3).
(2) *Initial Maturity Assessment of the Application Users*. This involved consultants from ATC visiting all of the application users, carrying out a process maturity assessment and collecting maturity information in the tool database. This was sent to us for detailed analysis (Month 4).
(3) *Process Improvement*. This stage involved us in producing a list of improvement recommendations and passing these to our consultant partners for implementation with the application users (Month 5). We provided support over the next 10 months to ATC during the improvement implementation phase of the project. (Months 6–16).
(4) *Process Maturity Reassessment*. ATC then reassessed the process maturity and, as before, transferred the results to us for analysis. (Month 17).
(5) *Business Performance and RE Process Improvement Analysis*. Working with our partners, we analysed the process improvement and business improvement data and tried to understand the relations between these (Month 18).

We discuss each of these stages in more detail in later sections of this article.

The IMPRESSION experiment involved real industrial companies with real problems. Unlike a university experiment carried out using students from the same background and pre-prepared materials, we had virtually no control over experimental variables such as the selection of subjects, their commitment and the projects used in the assessment. However, requirements engineering is primarily an industrial problem that is influenced by technical, business, economic and political factors. Working with industrial users reflects the reality of requirements engineering and we suspect that it is practically impossible to replicate this industrial reality in a university setting.

The fact that the work involved international partners offered us an important advantage from the point of view of experimentation. Because our application user partners normally worked in Greek rather than English, we were not directly involved in applying our maturity model. Hence, we avoided a problem of software engineering experimentation where the experiment is carried out by the inventor of some technique. Under those circumstances, it is very difficult for the experimentor to be objective and hence there is always some question mark over "successful" results. We are therefore confident that the results that we present in this paper are not strongly biased by our own desire for our model to "work".

5. METHOD AND TOOL DEVELOPMENT

It is a truism that any method has to be adapted for particular circumstances of use and the first stage in our work in the IMPRESSION project was to examine the maturity assessment process to assess what changes were required. The driving factor for change was the fact that we would be indirectly rather than directly involved in the assessment and improvement process and that our Greek consultant partners, who had experience of requirements engineering but not of RE process improvement, would be our interface with the application users.

The principal method change that we introduced was to make specific suggestions for improvements to be introduced rather than simply leave this entirely to the judgment of the assessor. To simplify the task of the assessor, we also developed a simple tool that supported the process of assessment, automated the maturity level computation and provided a series of process reports for the assessor and application users.

When originally developing the process maturity model, we assumed that a consultant would work with an organization and use their own knowledge of requirements engineering to propose process improvements that were appropriate to that context. In practice, after discussions with our technology partners, it became clear that some improvement guidance based on the good practice guidelines would be helpful. We therefore extended the maturity assessment method with additional activities geared to improvement rather than assessment. The improvement approach that we proposed was a cyclical process:

(1) Assess the process maturity of the organization.
(2) Using information on possible improvements suggested by the IMPRESSION tool, identify the most cost-effective improvements that can be introduced in the time available.
(3) Implement these improvements and then re-assess the maturity level. This later reassessment allows us to check if the improvements have been implemented and if implementing the improvements has adversely affected the use of existing good practice.

To support this process, we extended the maturity assessment method with two additional steps:

Compare Levels. This is a feedback facility to compare the results of the current assessment with the previous assessment, using the type of improvements made, prior to making further recommendations for improvements.

Recommend Improvements. This allows for the input of improvement recommendations so that we could, at the end of the assessment, analyse which guidelines have been implemented and to what extent.

The tool main control window display (Figure 5) shows these extra steps.

The IMPRESSION tool that we developed in the first phase of the project supports data collection, maturity assessment computation and report generation. It is a stand-alone tool, developed using Java and a Microsoft Access database, incorporating a model of the assessment process and providing user support at each stage of that process. It is calibrated with the 66 good practice guidelines and computation methods, interactively accepts scores and recommendations from a user, performs the maturity calculations and provides an extensive

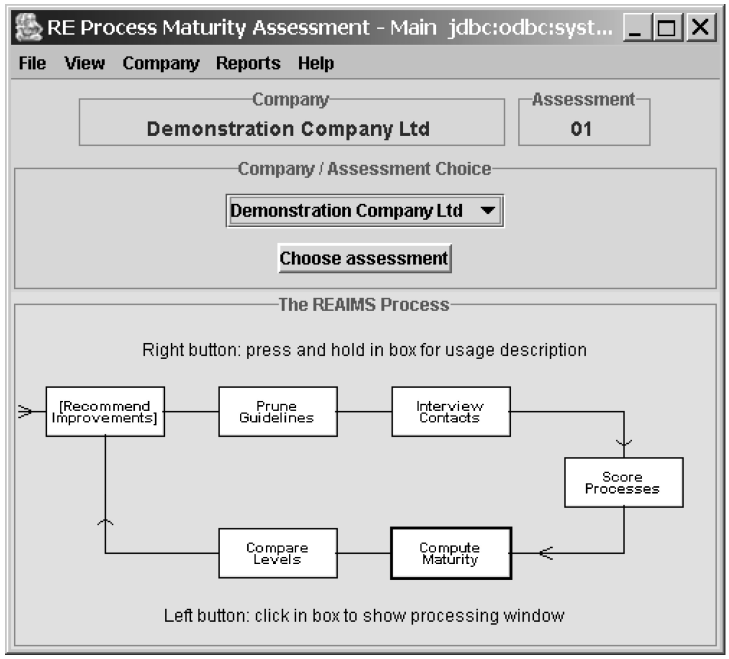

Fig. 5. The IMPRESSION maturity assessment tool.

reporting capability. We chose this approach to implementation rather than a spreadsheet approach because it made it easier to create and adapt reports about the processes in individual companies.

The primary motivation for developing this tool was the need for rapid maturity assessment. To assess process maturity, an assessor has to assess the level of standardisation for each of 66 Guidelines and allocate each a score. Then, to arrive at a maturity level and perform an analysis, these assessments must be separated, sorted and summed. Reports are then created for the client whose process maturity is being assessed.

The process maturity computations are not difficult but are tedious and prone to error. Report generation, however, can take a considerable time, as this involves identifying and collating the practices used in a company and presenting these in different ways. Manually, excluding the time required for information gathering, we estimated that computing the process maturity and generating the required reports for each company would take about 2 working days.

Our aim in developing the IMPRESSION tool was to reduce the time for information processing and report generation to 2–3 hours. The time required for information gathering at each company varied from 1 to 3 days initially, with slightly less time required for information gathering when maturity re-assessment was carried out. Therefore, with the tool, it was possible to carry out all initial maturity assessments within a month thus leaving us time to introduce process improvements and re-assess the process maturity. Developing the tool took about 3 weeks of effort so we estimated that the time savings were worthwhile. A further benefit of using the tool, rather than manual analysis, was that all data was expressed in a standard way so that it was simpler to make comparisons across the different companies involved.

6. INITIAL RE PROCESS MATURITY ASSESSMENT

The maturity matrices, described in Section 3.1, following the first assessments of the maturity of the application users are shown in Figure 6. These are ordered left to right depending on the number of basic guidelines that were used. The initial maturity assessment showed large variations in process maturity across the companies even although all of them were assessed at operating at an Initial level of maturity. For example, Company 6 made extensive use of good practice whereas Company 2 essentially did not do any requirements engineering in their systems' procurement processes.

As we were not involved directly in the assessment process, we had no intuition about the strengths and weaknesses of each company's processes so, in addition to the maturity matrices, we added a facility to the tool that presented an overview of each company's use of good RE practice according to process area.

These *Area/Strength matrices* for each company are shown in Figure 7, which identifies the different types of guideline (rows in the matrix) and classifies the company's use of guidelines of that type as weak (the leftmost column), average, good and strong (the rightmost column). For example, you can see from the area/strength matrices that Company 5 is strong in describing requirements and weakest in requirements elicitation, management and safety requirements (relevant for this company as it operated in the aerospace sector).

The value of the area/strength matrices was that they gave us an overall picture of the strengths and weaknesses of the application users and helped us decide where limited improvement effort should be focused. Of course, they are only a partial view of strengths and weaknesses—they do not show the extent to which practice is institutionalized, but we found them to be a valuable communication tool when discussing improvements with management in the application companies.

Observations from this initial maturity assessment are:

— Our suspicions that the application users were operating at low maturity levels were confirmed. All of the companies were operating at the Initial level of RE process maturity. However, we were pleased to see that Company 5 and Company 6 were fairly close to the weighted score of 55 in basic guidelines that represented the threshold between Initial and Repeatable levels of maturity.
— We were surprised that a number of companies successfully used intermediate and advanced good practices in spite of the fact that they did not really have a strong base of fundamental practice. This either meant that our notion of what was an intermediate or advanced practice was wrong or that companies could successfully use such practices without having their processes under control. The companies reported that they used these advanced

Company 6

	Bas	Int	Adv
G'lines	23	9	2
Wt. Sc.	45	17	4
Max	108	63	27
% Max	42%	27%	15%
Level	Initial+		

Company 5

	Bas	Int	Adv
G'lines	15	12	4
Wt. Sc.	37	29	8
Max	108	63	27
% Max	34%	46%	30%
Level	Initial+		

Company 1

	Bas	Int	Adv
G'lines	15	6	2
Wt. Sc.	23	7	2
Max	108	63	27
% Max	21%	11%	7%
Level	Initial		

Company 3

	Bas	Int	Adv
G'lines	14	2	0
Wt. Sc.	20	4	0
Max	108	63	27
% Max	19%	6%	0%
Level	Initial		

Company 8

	Bas	Int	Adv
G'lines	10	2	0
Wt. Sc.	18	5	0
Max	108	63	27
% Max	17%	8%	0%
Level	Initial		

Company 9

	Bas	Int	Adv
G'lines	12	4	1
Wt. Sc.	16	7	2
Max	108	63	27
% Max	15%	11%	7%
Level	Initial		

Company 4

	Bas	Int	Adv
G'lines	5	1	0
Wt. Sc.	8	1	0
Maxi	108	63	27
% Max	7%	2%	0%
Level	Initial		

Company 7

	Bas	Int	Adv
G'lines	5	1	0
Wt. Sc.	8	1	0
Max	108	63	27
% Max	7%	2%	0%
Level	Initial		

Company 2

	Bas	Int	Adv
G'lines	3	1	0
Wt. Sc	3	2	0
Max	108	63	27
% Max	3%	3%	0%
Level	Initial		

Fig. 6. The RE maturity matrices for all application companies after the initial maturity assessment.

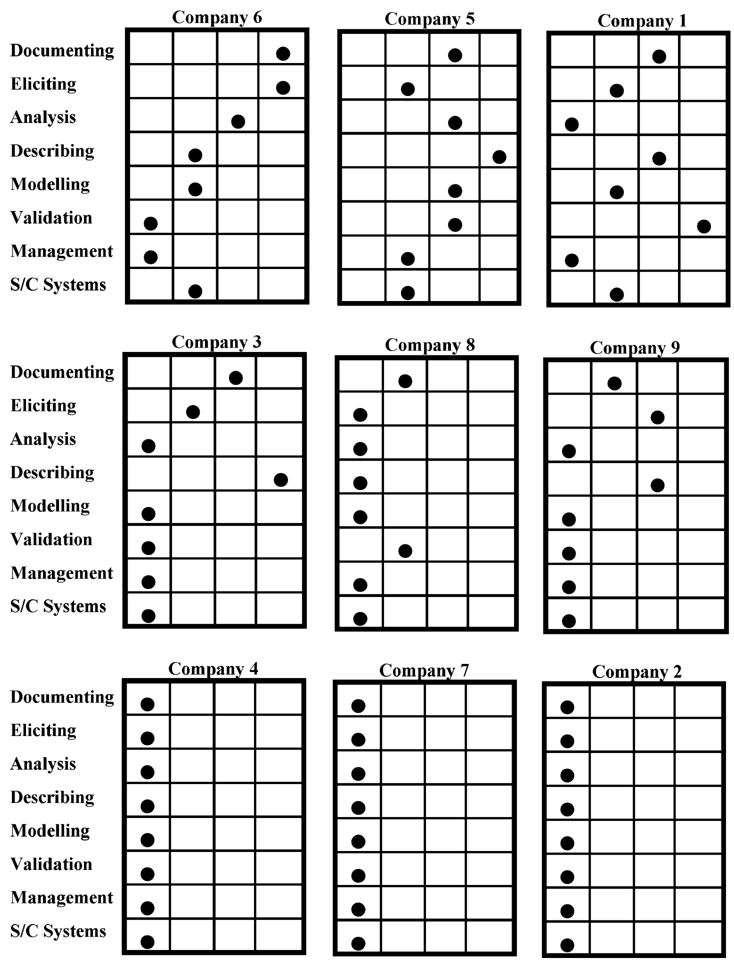

Fig. 7. Area/strength matrices after the initial maturity assessment.

practices, including the use of formal specification in one company, without difficulty. The maturity matrices show the rating of such companies as Initial+ rather than Initial to distinguish them from companies with few or no intermediate or advanced practices.

7. PROCESS IMPROVEMENT

While the benefits of process improvement are widely accepted, in practice, implementing real process improvement is a difficult task. When introducing new practices or improving existing ones, you have to consider the abilities, receptivity and experience of the personnel involved and the time and resources available. You have to ensure that business goals are not threatened by the process changes and that changes can be introduced without excessive costs.

In the RE Good Practice Guide [Sommerville and Sawyer 1997], we recommend that that new RE practices should be introduced incrementally and in such a way that the use of existing good practice is not compromised. An incremental improvement might be the introduction of a small number of new practices or increasing the score for a particular practice by one level. Based on the advice associated with the SEI's Capability Maturity Model for software, we recommend that companies should normally delay the introduction of intermediate and advanced practices until a good base of basic practices has been institutionalized.

The IMPRESSION project as a context for process improvement placed additional constraints on the choice of improvements. To deliver results to the project sponsor, our recommendations had to be implemented within the relatively short timescale of the project. It was a requirement of the application users that the process changes should not disrupt their normal business operations. Consequently, our goal was to generate the greatest increase in RE maturity for the lowest cost without compromising organizational profitability or operations. To do so, we made the following pragmatic proposals for maturity improvement:

(1) Focus improvements on areas of requirements engineering where the company is weak as suggested by the Area/Strength matrix.
(2) Consolidate and standardise practices that are already in use in the company.
(3) Only introduce new practices where the cost of introduction is low.

As most Intermediate and Advanced practices require additional training, their costs of introduction is fairly high. We therefore decided to recommend for each company that the focus of process improvement should be to bring in new Basic practices to strengthen the foundation for requirements engineering and to standardize Basic practices that were already used. Of course, the decision on what changes were actually made was the company's not ours. However, all companies accepted the wisdom of establishing a strong basis for RE and so followed our recommendations.

These proposals had to be interpreted for each of the application user companies involved. Companies with a relatively high maturity level are most likely to spend time consolidating existing practices. Low-maturity companies need to bring in new practices to establish a base for later standardization.

7.1 Identifying Weak Areas

The first stage in process improvement was to identify the areas where a company is weakest as these are where improvements are most likely to be cost-effective. We therefore identified the need for more detailed information on the use of good practice in different areas of the requirements engineering process than was available from the maturity and area/strength matrices. To provide this more detailed information for each company, we added a further report that showed how "strong" an organization is in each RE area as the number of practices actually used in that area expressed as a percentage of the total possible.

A sample of the report showing guideline use by process area for Company 5 is shown in Figure 8. The aim of this report was primarily to show guideline

Company 5 - Guideline use by process area				
	Requirements Document	**Elicitation**	**Analysis**	**Description**
Count	5	4	4	4
Maximum	8	13	8	5
% Usage	63	31	50	80
	Modelling	**Validation**	**Management**	**Critical systems**
Count	4	4	3	3
Maximum	6	8	9	9
%Usage	67	50	33	33

Fig. 8. Guideline use by process area for Company 5.

usage rather than standardization so we do not include the weighted score here. In the summary report, a column is displayed for each of the eight areas (on the left in Figure 8) showing:

Row 1: Guideline Counts. A count of the number of Basic, Intermediate, and Advanced guidelines that were used in this area

Row 2: Maximum. The maximum number of guidelines in this area.

Row 3: The percentage usage figure. The count expressed as a percentage of the maximum.

We decided to target an area for improvement if fewer than 50% of the Basic guidelines in an area were used. This was a pragmatic choice that was based on the assumption that introducing new guidelines probably required more support than standardizing guidelines that were already used. As our support resources were limited, we felt that this was the best way to help our application user partners.

You can see from Figure 8 that Company 5 is strongest in area 3 (the requirements document), area 6 (describing requirements) and area 7 (system modelling). For this company, therefore, we targeted improvements on the other assessment areas.

7.2 Recommending Improvements

Recall that our improvement strategy was to focus on introducing or standardizing the use of basic good RE practices in areas where the companies were weakest and to recommend guidelines that had relatively low costs of introduction. To help the consultant choose improvements to recommend to the company, we added two reports to the maturity assessment tool:

(1) A guideline usage report (Figure 9) that shows, for each guideline, the level of use of that guideline in a particular company. You can see from Figure 9 this that Company 5 has standardized the use of seven good practice guidelines and normally uses eight others across all their projects, although the

Company 5 – Basic Guideline Use		
ID	**Guideline**	**Score**
03.02	Explain how to use the document	Not Used
03.04	Make a business case for the system	Not Used
03.08	Make the document easy to change	Nor Used
04.01	Assess system feasibility	Not Used
04.02	Be sensitive to organisational and political considerations	Not Used
04.03	Identify and consult system stakeholders	Not Used
04.04	Record requirements sources	Not Used
04.05	Define the systems operating environment	Not Used
04.06	Use business concerns to drive requirements elicitation	Not Used
05.01	Define system boundaries	Not Used
05.03	Provide software to support negotiations	Not Used
05.04	Plan for conflicts and conflict resolution	Not Used
06.01	Define standard templates for describing requirements	Not Used
08.01	Check that the requirements document meets your standard	Not Used
08.02	Organise formal requirements inspections	Not Used
08.03	Use multi-disciplinary teams to review requirements	Not Used
09.02	Define policies for requirements management	Not Used
09.03	Define traceability policies	Not Used
09.04	Maintain traceability manual	Not Used
10.01	Create safety requirement checklists	Not Used
10.02	Involve external reviewers in the validation process	Not Used
Number of Not Used Guidelines = 21	**Not used score (21 * 0) = 0**	
05.02	Use checklists for requirements analysis	Normal
06.02	Use language simply, consistently and concisely	Normal
06.03	Use diagrams appropriately	Normal
06.04	Supplement natural language with other descriptions of requirements	Normal
07.01	Develop complementary system models	Normal
07.02	Model the systems environment	Normal
07.03	Model the system architecture	Normal
09.01	Uniquely identify each requirement	Normal
Number of Normal Guidelines = 8	**Normal score (8 * 2) = 16**	
03.01	Define a standards document structure	Standardised
03.03	Include a summary of the requirements	Standardised
03.05	Define specialised terms	Standardised
03.06	Lay out the document for readability	Standardised
03.07	Help readers find information	Standardised
05.05	Prioritise requirements	Standardised
08.04	Define validation checklists	Standardised
Number of Standardised Guidelines = 7	**Standardised score (7*3) = 21**	
Total number of Basic Guidelines = 36	**Score 37**	

Fig. 9. Use of Basic good practices in Company 5.

use may vary from project to project. For that company, there were no guidelines that were used in some projects but not in others.

(2) An unused guidelines report for each company that identifies guidelines that could be introduced and that have a low or moderate cost of implementation.

Company 5 – Guidelines by Cost of Implementation		
Guidelines with a Very Low Cost of Implementation		
ID	**GUIDELINE**	**TYPE**
03.02	Explain how to use the document	Basic
03.04	Make a business case for the system	Basic
04.03	Identify and consult system stakeholders	Basic
Guidelines with a Low Cost of Implementation		
ID	**GUIDELINE**	**TYPE**
03.08	Make the document easy to change	Basic
04.01	Assess system feasibility	Basic
04.02	Be sensitive to organisational and political considerations	Basic
04.04	Record requirements sources	Basic
04.05	Define the systems operating environment	Basic
04.06	Use business concerns to drive requirements elicitation	Basic
05.01	Define system boundaries	Basic
05.04	Plan for conflicts and conflict resolution	Basic
08.01	Check that the requirements document meets your standards	Basic
08.03	Use multi-disciplinary teams to review requirements	Basic
09.04	Maintain traceability manual	Basic
10.02	Involve external reviewers in the validation process	Basic
04.07	Look for domain constraints	Intermediate
05.07	Use interaction matrices to find conflicts and overlaps	Intermediate
07.06	Document the links between stakeholder requirements and system models	Intermediate
10.05	Cross-check operational and functional requirements against safety requirements	Intermediate
09.09	Record rejected requirements	Advanced
Guidelines with a Low to Moderate Cost of Implementation		
ID	**Guideline**	**Type**
05.03	Provide software to support negotiations	Basic
10.01	Create safety requirement checklists	Basic

Fig. 10. Guidelines that could be introduced by Company 5.

Figure 10 shows the unused guidelines for Company 5. Our intention in this report was to identify guidelines with a low introductory cost so, for completeness, highlighted all such guidelines that a company did not use and not just Basic guidelines. It was then up to each company, to decide which of these to select for introduction and to set their own schedule for improvement. Our partners ATC acted as consultants during this period helping the application users prioritize their needs, select appropriate guidelines and introduce these in different projects within the company.

8. FINAL RE PROCESS MATURITY ASSESSMENT

The purpose of this assessment was to compare the results with the initial assessment to quantify the actual RE process improvements made. The maturity matrices following the second assessment are shown in Figure 11 and are presented in the same sequence as those in the initial assessment. To allow a simple comparison to be made with the previous assessment, the scores in the initial maturity assessment are shown in brackets. To illustrate how maturity levels changed, Figure 12 shows the rank order of the companies in the initial and final maturity assessment.

For each company, we show the weighted score of Basic practices (to the left of the vertical bar), the weighted score of Intermediate/Advanced practices (to the right of the bar) and the computed maturity level (I for Initial, R for Repeatable). From Figure 12, it is clear that all of the companies involved have increased their RE process maturity. Four companies moved from the Initial maturity level to the Repeatable level and the other four companies all introduced a significant number of good requirements engineering practices.

The matrices show that different companies have used different strategies to increase their maturity level. Company 6, for example, has introduced relatively few new guidelines (from 23 to 27 Basic guidelines) but has increased its maturity level by increasing and standardizing the use of good practice (shown by the increase in the weighted score from 45 to 62). By contrast, Company 5, which has also moved from Initial to Repeatable, has introduced 14 new Basic guidelines.

To give us a picture of the areas where each company has introduced improvements we modified the area/strength matrices to show before and after strengths and weaknesses. Figure 13 shows the area/strength matrices as estimated in the final maturity assessment. The large dots represent the strength in the first assessment and the small black dots represent the scores in the re-assessment.

The area/strength matrices that were generated after process improvements were introduced reveal information that is hidden by the quantitative presentation in the maturity matrices. For example, although Company 7 has introduced five new basic guidelines, this has made no practical difference to its strength in requirements engineering. It is still weak in almost all areas. By contrast, Company 4 whose initial estimate was the same as Company 7's has introduced 16 new guidelines and has increased its strength in a number of RE process areas.

The data collected during the process maturity assessment allowed us to see the changes that individual companies made to their process. To provide management information, we created summary reports for each company which described the changes that the company had made and predicted how these might affect business goals (discussed later). The improvement summary report for Company 5 is shown in Figure 14.

9. RE PROCESS IMPROVEMENTS AND BUSINESS PERFORMANCE

One of the overall aims of the IMPRESSION project was to investigate the relationship between RE process improvement and business benefits. Our hypothesis was that there would be a correlation between the extent of RE process improvement and the business benefits gained.

Company 6

	Bas	Int	Adv
G'lines	27(23)	9 (9)	2 (2)
Wt. Sc.	62 (45)	17 (17)	4 (4)
Max	108	63	27
% Max	57%	27%	15%
Level	Repeatable		

Company 5

	Bas	Int	Adv
G'lines	29 (15)	12 (12)	4 (4)
Wt. Sc.	76 (37)	29 (29)	8 (8)
Max	108	63	27
% Max	70%	46%	30%
Level	Repeatable		

Company 1

	Bas	Int	Adv
G'lines	25 (15)	6 (6)	2 (2)
Wt, Sc.	44 (23)	7 (7)	2 (2)
Max	108	63	27
% Max	41%	11%	7%
Level	Initial		

Company 3

	Bas	Int	Adv
G'lines	28 (14)	2 (2)	0 (0)
Wt. Sc.	51 (20)	4 (4)	0 (0)
Max	108	63	27
% Max	47%	6%	0%
Level	Initial		

Company 8

	Bas	Int	Adv
G'lines	29 (10)	2 (2)	0 (0)
Wt. Sc.	68 (18)	5 (5)	0 (0)
Max	108	63	27
% Max	63%	8%	0%
Level	Repeatable		

Company 9

	Bas	Int	Adv
G'lines	29 (12)	4 (4)	1 (1)
Wt. Sc.	60 (16)	7 (7)	2 (2)
Max	108	63	27
% Max	56%	11%	7%
Level	Repeatable		

Company 4

	Bas	Int	Adv
G'lines	21 (5)	1 (1)	0 (0)
Wt. Sc.	38 (8)	1 (1)	0 (0)
Max	108	63	27
% Max	35%	2%	0%
Level	Initial		

Company 7

	Bas	Int	Adv
G'lines	10 (5)	1 (1)	0 (0)
Wt. Sc.	19 (8)	1 (1)	0 (0)
Max	108	63	27
% Max	18%	2%	0%
Level	Initial		

Company 2

	Bas	Int	Adv
G'lines	15 (3)	1 (1)	0 (0)
Wt. Sc.	29 (3)	2 (2)	0 (0)
Max	108	63	27
% Max	27%	3%	0%
Level	Initial		

Fig. 11. Final maturity matrices.

Initial Maturity Level	**Final Maturity Level**
Company 6 45 | 21(I)	Company 5 76 | 37 (R)
Company 5 37 | 37 (I)	Company 8 68 | 5 (R)
Company 1 23 | 9 (I)	Company 6 62 | 21 (R)
Company 3 20 | 4 (I)	Company 9 60 | 9 (R)
Company 8 18 | 5 (I)	Company 3 51 | 4 (I)
Company 9 16 | 9 (I)	Company 1 44 | 9 (I)
Company 4 8 | 1 (I)	Company 4 38 | 1 (I)
Company 7 8 | 1 (I)	Company 2 29 | 2 (I)
Company 2 3 | 2 (I)	Company 7 19 | 1 (I)

Fig. 12. Comparison of initial and final maturity levels.

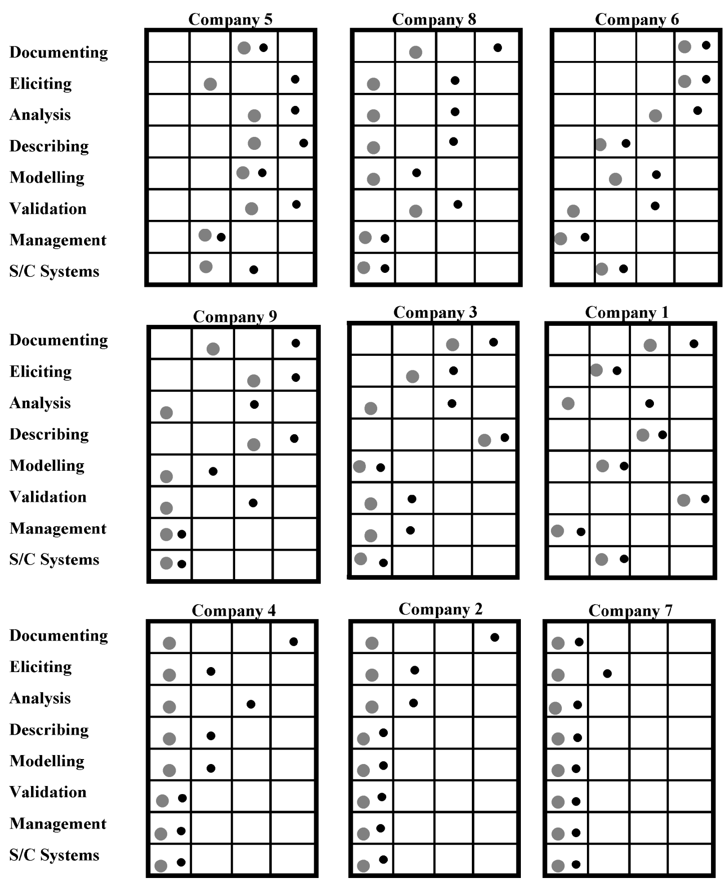

Fig. 13. Area/strength matrices following process re-assessment.

The company improved 7 out of 8 recommended practices, introduced 14 out of 14 recommended practices and improved its process maturity by 105%. It now uses 22 practices at a Normal level and 23 at a Standard level. All new practices were introduced at the Normal or Standard level of usage and, as all previously used practices were at the Normal level, all improvements have been to the Standard level.

There is a significant increase in strength for Eliciting Requirements and a modest increase in Requirements Analysis and Negotiation, Requirements Validation and Critical Systems Considerations. We would expect to see a corresponding improvement in Stakeholder Business Goals.

Company 5 was previously strong in areas affecting System Business Goals so we would not expect these to be able to improve to any great extent.

Fig. 14. Summary improvement report for Company 5.

This hypothesis raises a number of questions that, as we discuss below, turned out to be very difficult to answer and which seem to us to be quite fundamental to any studies of the relationships between software engineering process improvement and business performance. These questions were:

(1) How do we measure relative process improvement across the application users?
(2) How can we, in a relatively short-term project, measure business benefits that *may* result from process improvement and can we compare these benefits across the application users?
(3) Can we be confident that any observed business benefits resulted from process improvements rather than some other factors that may or may not be related to the improvement exercise?

In the IMPRESSION project, other partners, working the application users, devised the approach used to assess business performance and to measure performance improvement. We therefore do not discuss the approach used to the measurement of business benefits in any detail. However, the companies involved were satisfied that it presented a reasonably accurate picture of the actual changes in the business over the lifetime of the project.

9.1 Comparing RE Process Improvement

To assess if business benefits correlates with process improvement, we must first try make a comparative assessment of the improvements of the different companies. This assessment can be carried out against using either a relative or an absolute measure. A relative measure looks at how much each individual company has improved relative to its own starting point; an absolute measure looks at improvements in some value irrespective of its baseline. Initially, we proposed to use a relative measure based on the percentage improvement in process maturity for each company. This was calculated as:

$$Improvement = \frac{current\ assessment\ score - previous\ assessment\ score * 100}{previous\ assessment\ score}.$$

% Improvement		# Practices Improved		Basic Maturity Increase	
Company 2	867%	Company 9	26	Company 8	50
Company 4	375%	Company 8	24	Company 9	44
Company 8	278%	Company 5	21	Company 5	39
Company 9	275%	Company 3	18	Company 3	31
Company 3	155%	Company 4	18	Company 4	30
Company 7	100%	Company 1	16	Company 2	26
Company 5	105%	Company 2	15	Company 1	21
Company 1	91%	Company 6	12	Company 6	17
Company 6	38%	Company 7	5	Company 7	8

Fig. 15. Comparative maturity improvements.

The results of this assessment are shown in the left-hand column in Figure 15. It is clear that using the percentage improvement as a measure of improvement is biased towards companies with a very low initial maturity.

We also calculated the relative improvements of each company based on two absolute measures:

(1) The number of good RE practices improved or introduced.
(2) The increase in Basic maturity—recall that we focused improvements at this level.

As you would expect, the rank ordering of improvement based on these measures, as shown in columns 2 and 3 of Figure 15, is fairly similar. However, it is radically different from the ordering that results when a relative improvement assessment is calculated.

To illustrate the difficulties of making sense of these results, consider Company 2 and Company 6. According to the relative improvement table, these had the highest and the lowest percentage improvements to their process maturity. Yet, using the absolute value assessment, we see that the changes in each company were very similar. What differed, radically, of course were the starting points of each company. Company 6 had the highest maturity in the initial assessment and Company 2 the lowest. Company 6 decided that their improvements should focus on standardizing existing practice and only introduced a small number of new practices; this required little effort on their part. Company 2, by contrast, were very committed to the project but, because of their low baseline and knowledge level had to expend a great deal of effort to achieve a comparable absolute improvement.

This analysis highlighted to us the difficulties in comparing process improvement across companies. Is it meaningful to say that Company A has improved more than Company B? Unless all companies that are being assessed start around the same level of process maturity then it is difficult to interpret any relative assessment of process improvements. This seems to us to be a general problem with any experiments involving the introduction of new software techniques or technologies into different organizations. We cannot control the baseline so cannot therefore determine the relative improvements in a reliable way.

9.2 Assessing and Comparing Business Benefits

The "holy grail" of the software engineering research community is to demonstrate that a proposed innovation (be it in methods or tools) leads to business benefits from companies adopting that innovation. To make an assessment of the business benefits of RE process improvement, we had to choose appropriate metrics that reflected these benefits and then take measurements before and after the process improvements were introduced. Our project partners proposed the following approach to make this assessment:

(1) Through discussion with the user partners, discover business goals that were likely to be supported by RE process improvement.
(2) Identify how we could assess if the company was making progress towards its business goal. To do this, we identified a number of critical success factors (CSFs) that were related to the business goals—we discuss this further below.
(3) Identify specific measurements that we could make that were related to the CSFs. We measured the values of key performance indicators (KPIs) where a KPI is a *measurable* process or product attribute. By examining changes in the KPIs, we could assess whether or not each company was making progress towards its process improvement goals and hence its business goals.

The high-level, strategic business goals for the application users reflected the business improvements that they wished to make. Different organizations had different goals but we finally settled on six business goals shared by all companies. These are shown in the left column in Figure 16. Improving RE process maturity was identified by all of our partners as a business goal in itself. The reason for this was that they had all experienced problems with their RE processes and felt that, even if no other benefits ensued, better RE processes would be of value to the business.

To help us decide whether or not progress was being made towards a business goal, we identified 11 associated critical success factors that are shown in the right column of Figure 16. Each CSF can be thought of as a software process improvement goal—they were not just concerned with RE processes. For example, to reach the business goal of improving software productivity, associated CSFs were to reduce the amount of code that had to be reworked and to reduce the amount of effort devoted to rework. Improving RE processes should contribute to these goals, for example, reducing requirements rework means that less effort overall will be spent on rework.

To enable us to assess progress towards each CSF, we associated one or more Key Performance Indicators with each CSF. A KPI is a process or product measure that can be made at different stages in the software process. Examples

Business goal	Critical Success Factors
Improve RE process maturity	Increase the use of RE standards.
Improve managerial satisfaction	Improve the traceability of requirements throughout the system lifecycle. Reduce uncertainty throughout the system lifecycle.
Improve user satisfaction	Reduce the number of complaints from internal and external users.
Improve the efficiency of the system implementation process	Get the system right first time Reduce time to deployment of software Reduce the deviation from project schedules Reduce the deviation from project budgets
Improve the maintainability of deployed systems	Improve software extendability Reduce maintenance costs
Improve the productivity of software development	Reduce code production related to rework Reduce the effort devoted to rework

Fig. 16. Business goals and critical success factors.

of KPIs that we identified are:

—The number of requirements engineering standards used;
—The percentage of system requirements that had to be reworked;
—The elapsed time between system conception and deployment;
—The time overrun on projects;
—The effort devoted to rework;
—The number of system modifications resulting from requirements engineering errors.

We identified 17 KPIs in total with each CSF associated with one or more KPIs. Some KPIs were associated with more than one CSF. For space reasons, we do not show the complete mapping here. Rather, Figure 17 shows the KPIs associated with the "Get the system right first time" CSF.

Each application user set up a team, selected relevant KPIs and recorded values regularly over the course of the project. The KPIs were initially measured for an existing "reference" project, to establish a baseline and subsequently for the project developed during IMPRESSION, termed the "demonstrator" project. The difference in the KPIs between the reference and demonstrator projects points to an improvement, or degradation, in business performance. User involvement in this process was critical—they chose KPIs that they believed were relevant to their business and that could indicate progress towards their business goals. This meant that the application users felt that the study

CSF	Associated KPIs
Get the system right first time	Percentage of requirements satisfied in initial design
	Number of requirements change requests relating to requirements errors per month
	Effort devoted to rework
	Number of system modifications resulting from requirements engineering errors

Fig. 17. A CSF and associated KPIs.

	# KPIs used in analysis	# KPIs 'improving'	# KPIs 'regressing'	# KPIs 'stable'
Company 1	11	9	2	0
Company 2	5	4	0	1
Company 3	11	11	0	0
Company 4	3	3	0	0
Company 5	12	11	1	0
Company 6	11	8	1	2
Company 7	5	4	0	1
Company 8	8	8	0	0
Company 9	9	7	2	1

Fig. 18. KPI assessment per company.

had real business value and none of them dropped out in the course of the experiment.

Figure 18 shows the number of KPIs used by each company and the number of KPIs that indicated improvements. The changes in the KPIs across all companies varied dramatically from −100% to +600%. Figure 19 is a histogram that shows, for each company (1–9), the changes in each KPI. Across all companies, we measured about 75 KPIs in all.

You can see that, in the vast majority of cases, there was up to a 100% improvement in the measured KPI. A small number of KPIs (6) showed a degradation and a small number (4) showed no change. In some cases, there was no change in a KPI because there was no real scope for improvement in that KPI from the baseline as the company was already performing well. Some of the decreases in KPIs resulted from differences beween the baseline and the demonstrator project.

Equally, a small number of KPIs indicated an improvement of more than 100% with, in one case, a massive improvement of almost 600%. This latter case was a change in the KPI "Number of unique requirements captured".

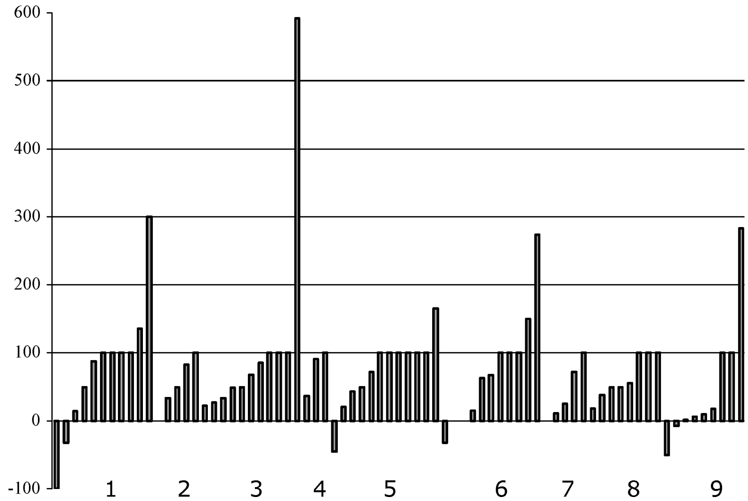

Fig. 19. KPI change for each company.

Essentially, the company's practice in stating requirements was to bundle several distinct requirements under a single requirements statement. After introducing the practice of uniquely identifying each requirement, the number of distinct requirements increased dramatically.

Because of the differing businesses of the companies involved, each company chose KPIs that were relevant for its business. As is clear from Figure 18, companies chose different KPIs and the number of KPIs used in different companies varied from 5 to 12. Some of the KPIs identified were not practically measurable in the limited time available (e.g., those relating to long-term system maintenance) and others were not relevant for specific companies. For example, companies that outsourced software used fewer KPIs than those that used in-house development. Therefore, our third question "How do we compare business improvement across several companies" was not readily answerable because of the lack of a common base set of KPIs used by all companies.

This presented us with a problem that we had not anticipated at the start of the study. We had hoped that all companies would chose the same way to assess business improvements so that we could try to correlate these with RE process improvements. Ideally, we would have liked to have seen the greatest business benefits emerging in the companies who had made most progress in RE process improvement. However, because of the significant variation in the measures used by each company, it was impossible to make sensible comparisons of their relative business improvement.

We are well aware of the methodological weaknesses of this approach to assessing business improvement. Observations were taken using only a single project as a baseline and using a single demonstrator project in each company. The fact that measurements were taken over a relatively short timescale means that the actual values of the KPIs are only an indicator of change. The actual values measured are clearly subject to error. However, we believe that the direction of movement of the indicators rather than their absolute value is significant—movement in a positive direction suggests that the changes introduced contributed to business improvement.

In all companies we saw positive improvements for most KPIs so concluded that our study led to real business benefits for the participants. This conclusion was reinforced by discussions with application user company management who were happy with the changes to the RE process that were made and who intuitively felt that there had been progress made towards their business goals.

There are general issues that arise from our work here that we believe are applicable to other technology transfer projects. The fact that we identified business goals rather than simply technical process goals was a critical factor in convincing company management that participation in the study was worthwhile. The CSFs and KPIs were developed through consultation with the application users and this consultation process further reinforced their commitment to the work. As researchers, it is important that we bear in mind that our research goals are rarely the goals of our business partner and accepting an imperfect method is perhaps worthwhile if it opens up the possibility of active industrial commitment to new methods and techniques.

9.3 RE Process Improvement and Business Benefits

The results of our study indicated that when the RE Process Maturity of an organization improved, an improvement was also observed in business performance indicators. However, we cannot draw any definite conclusions about whether or not the magnitude of the business performance improvement is correlated with the magnitude of the RE process improvement.

As we have discussed in Sections 9.1 and 9.2, making relative comparisons of improvements in process maturity and business goals was impossible. Company 2 had a relative improvement of process maturity of over 800% and Company 6 had an improvement of 38%. Even in the absolute tables, Company 2 did better than Company 6 yet when we look at the improvements in the KPIs shown in Figure 19, these suggest that, if anything, Company 6 improved more than Company 2. Our suspicion here is that improvements become more effective once some threshold level of maturity has been reached—Company 6 had reached this level so gained value from its process improvements; Company 2 started from such a low point that the threshold level was never reached.

The data collected suggests that involvement in the process improvement experiment resulted in business benefits for all of the companies involved. This was certainly reflected in the qualitative impressions of company representatives who, universally, felt that participation had been worthwhile and that they had gained some benefit from being involved. Of course, these benefits could have come from some other external factors but, given that nine companies from quite different sectors were involved, it is unlikely that there were comparable external reasons for the business benefits across all companies.

The data also supports the hypothesis that improving the maturity of RE processes leads to business benefits but we do not have sufficient data to propose any formula that would allow us to predict the business improvements resulting from some given level of process improvement. It is also the case that participation in the study had significant side-effects for the companies that may have influenced the results.

It is possible that it is the participation in the process of process improvement rather than the improvements themselves that is most significant. The Hawthorne effect is a well-known experimental effect where it is participation in an experiment that leads to changes in outcome rather than the changes proposed in the experiment. Furthermore, our partner, ATC, noted that, in some instances, the improvement in business performance could be attributed to a new awareness of business issues brought about as a side-effect the IMPRESSION exercise rather than by RE process improvements themselves. This increased awareness may have led to business process changes apart from changes to RE practice. It could be the case that this increased self-awareness is the key factor that led to business benefits and not the improvements in RE processes.

We believe that this increased self-awareness is a factor in improvement but we suspect that the truth is that the improvements resulted from a mixture of increased awareness and better practices. To judge the actual contribution of better practice would require the improvement exercise to be continued in the same organizations where (we hope) that the self-awareness factor would be less significant.

10. CONCLUSIONS AND LESSONS LEARNED

This industrial study in process improvement provided valuable information about requirements engineering process improvement. It demonstrated that the RE process maturity model was usable in practice across a range of companies and that it was a useful basis for supporting RE process improvement. Our study provided some evidence to confirm the hypothesis that improving RE process maturity contributes to improved business performance. However, because of differences between companies, it was impossible for us to judge whether or not there was a correlation between the degree of process improvement introduced and the business benefits that resulted from this.

When some method or tool is used in a real industrial setting, we learn useful lessons about how it can be improved. The particular value of the study was the lessons that we learned about the RE process maturity model, the good practice guidelines that we proposed as part of that model and the maturity assessment process.

10.1 Lessons Learned

Our initial development of the RE process maturity model was carried out under the assumption that maturity assessment would either be done by companies themselves or by consultants who worked closely with them. We assumed that they would be sufficiently familiar with the processes and activities used that they could identify improvements that could be introduced.

Perhaps the most important lesson that we learned as far as the process maturity model was that this was an optimistic assumption. Knowing the process maturity was not enough—companies also wanted to know what improvements could be introduced. Therefore, the amendment to the maturity model where we introduced a process improvement recommendation stage as shown in Figure 5 is an important result of this study. This change from a passive assessment model to an active improvement model and the development of tool support means that the startup costs for process improvement are reduced and hence increases the practical usefulness of the approach.

We were always aware that the maturity categorisation into Initial, Repeatable and Advanced was a somewhat arbitrary one but felt that identifying these different levels gave companies a target to aim at in the maturity improvement. However, these levels are too coarse to provide effective feedback on maturity improvement and the study showed that we need a continuous rather than a discrete model where companies can see progression according to guidelines used, standardized, etc. They also require more than a simple assessment but an across the board analysis of their strengths and weaknesses. This reflects experience with the CMM and movement towards a continuous model as reflected in CMMI [Ahern et al. 2001]. The guidelines are already organized into process areas such as requirements elicitation, requirements analysis, etc. and we are now considering a redesign of the maturity assessment based around these areas.

We also found that our classification of the good practice guidelines was far too complex. First, there are too many categories of costs of introduction of guidelines—we had initially defined eight possible cost categories. These could be simplified to Low, Moderate and High. Second, our classification of each guideline as Basic/Intermediate/Advanced needs to be adapted for different application domains. A guideline that might be basic guideline, such as "define traceability policies" for a technologically advanced company is perhaps an advanced guideline for a company operating at a very low level of maturity. We now think that it is probably the case that using area and cost of introduction as the classification attributes for the guidelines is all that we need and that the basic/intermediate/advanced classification is unnecessary.

Our process maturity model was initially developed for companies working in critical systems development and two of the application users in the IMPRESSION project fell into this category. However, the other companies involved were either involved in business systems development or companies who outsourced all of their systems development from external contractors. Realistically, these companies will never achieve the high levels of process maturity that can be expected from companies developing critical systems. Similarly, there are some guidelines that are specific to some domains and, indeed, outsourcing may require new guidelines to be introduced.

This project has shown that the same model can be applied across different types of company but that there is probably a need for sector-specific variants of the model, especially if the different levels are maintained. Therefore, if we maintain the staged model, a company that outsources all of its software development might be rated as operating at an Intermediate level of maturity with a lower overall score in good practice use than a company that develops critical systems. We are now considering how the model should evolve to make it more configurable.

10.2 Tool Support

We believe that a fundamental success factor in this study was the development of the software tool to support the process maturity model. This made it possible for nine analyses and reanalyses to be carried out in a relatively short period and for the method to be applied by an organization that was not involved in its development. Without this tool, the costs of data collection and analyses would have *appeared* prohibitive and we do not believe that the commitment of our industrial partners could have been maintained.

Our general conclusion here is that if you are involved in any study that involves extensive data collection then you should develop the support tool before starting to collect data. This significantly lowers the initial commitment costs from industry and so lowers the barriers to participation in software engineering experiments.

10.3 Conclusion

Our involvement in this process assessment and improvement study was a very valuable exercise that provided encouraging feedback about the usefulness of the process maturity model. It also revealed some fundamental difficulties of assessing and comparing process maturity and process improvement.

If we return to the questions identified in the introduction to this article, how well have these been answered?

First, we wanted to discover the practical difficulties that arise in RE process maturity assessment. We have not mentioned these in the article simply because we had no real problems in this respect. We anticipated problems where our industrial partners were unable to engage in the process because of other commitments but these problems simply did not arise. We believe that this was largely due to the fact that we presented the work to them from a business rather than a technical perspective.

Second, the IMPRESSION project confirmed the usefulness of the RE process maturity model and improvement framework across a range of different types of company. Without the structured approach provided by the model, assessing RE process maturity and, critically, selling the notion of RE process improvement to application users would have been much more difficult. The fact that an external consultant could apply the model without previous experience in maturity assessment was also an encouraging result.

Third, we have not unequivocally demonstrated that RE process improvement leads directly to business benefits. This is extremely difficult in a short-term project. However, the evidence that was produced strongly suggested that participation in a process improvement process resulted in benefit and it is likely that at least some of this benefit resulted from better requirements engineering practice.

The study highlighted a number of key points that researchers interested in assessing the value of new tools and methods should bear in mind:

(1) Industry is primarily interested in business benefits rather than technical measures that compare existing methods with new approaches. Whatever

the weaknesses of the method used to assess business benefits, the fact that we talked to companies in these terms meant that they engaged in the study and were committed to it. To engage industry in technology transfer, we cannot just present technical solutions but must also relate these to the real needs of the business.

(2) Relative comparisons of the efficacy of tools and methods will inevitably be very difficult. There is so much variation in the practices used in different industries that creating a common baseline will always be very difficult. We will probably never discover the "holy grail" where we can relate business improvements directly to the introduction of new technologies. As researchers and reviewers, we must understand the fundamental difficulties here and not expect too much from empirical software engineering. We should see measurements made in industry as contributing to a body of evidence about some tool or technique and should not expect them to stand on their own.

There is no question that experimentation in software engineering is very difficult and fraught with uncertainties. Controlled experiments (often using students) are so far removed from industrial reality that their conclusions will always be open to question. Industrial experiments suffer from all of the messy problems of reality that mean generalizations are difficult to sustain. In an ideal world, we would, of course, have a structured approach where we move from the laboratory to industry. In reality, this is often impossible and our experience in the IMPRESSION project suggests that, in spite of the uncertainties in the results, industrial experiments are an effective means of highlighting practical problems with software engineering techniques and demonstrating the value of such techniques to industry.

ACKNOWLEDGMENTS

We would like to thank our partners in the IMPRESSION project—Athens Technology Centre, Greece and City University, London—who made this work possible. Particular thanks are due to Andreas Kalligeris, Christina Sorovou and Despina Politou from ATC who worked with our industrial users to oversee the requirements engineering improvement exercise and collect assessment data and to Bill Karakostas from City who devised the business benefit assessment scheme. Pete Sawyer from Lancaster provided helpful comments on earlier drafts of the article.

REFERENCES

AHERN, D. M., CLOUSE, A., AND TURNER, R. 2001. *CMMI Distilled*. Addison-Wesley, Reading, Mass.

BOEHM, B. W. 1983. The economics of software maintenance. In *Proceedings of Software Maintenance Workshop* (Washington, D.C.), 9–37.

COALLIER, F. 1999. TRILLIUM: A model for the assessment of telecom product development and support capability. In *Software Process Improvement*, R. B. Hunter and R. H. Thayer, Eds. IEEE Computer Society Press, Los Alamitos, Calif.

DEMING, W. E. 1982. *Out of the Crisis*. MIT Press International, Cambridge, Mass.

El Emam, K., Drouin, J., and Welo, M. 1997. *SPICE: The Theory and Practice of Software Process Improvement and Capability Determination*. IEEE Computer Society Press, Los Alamitos, Calif.

Espiti. 1996. Software process improvement on the right road with ESPITI—The ESPITI European Survey Results. *ESPITI Newsletter Issue 2*. Available at: http://www.cse.dcu.ie/cse/international/trispin/News2.html#espiti.

Haase, V., Messnarz, R., Koch, G., Kugler, H. J., and Decrinis, P. 1994. Bootstrap: Fine tuning process assessment. *IEEE Software 11*, 4, 25–35.

Hall, T., Beecham, S., and Rainer, A. 2002. Requirements problems in twelve software companies: An empirical analysis. *IEE Proceedings: Softw. 149*, 5, 153–60.

Humphrey, W. 1989. *Managing the Software Process*. Addison-Wesley, Reading, Mass.

Koch, G. 1993. Process assessment: The 'BOOTSTRAP' approach. *Inf. Softw. Tech. 35*, 6/7, 387–403.

Konrad, M. and Paulk, M. 1995. An overview of SPICE's model for process management. In *Proceedings of the 5th International Conference on Software Quality* (Austin, Tex.).

Kuvaja, P., Similä, J., Krzanik, L., Bicego, A., Saukkonen, S., and Koch, G. 1994. *Software Process Assessment and Improvement: The BOOTSTRAP Approach*. Blackwell Publishers, Oxford.

Lutz, R. R. 1993. Analysing software requirements errors in safety-critical embedded systems. In *Proceedings of RE'93* (San Diego Calif.).

Paulk, M. C., Curtis, B., Chrissis, M. B., and Weber, C. V. 1993. Capability maturity model, Version 1.1. *IEEE Softw. 10*, 4, 18–27.

Paulk, M. C. and Konrad, M. 1994. An overview of ISO's SPICE project. *IEEE Comput. 27*, 4, 68–70.

Paulk, M. C., Weber, C. V., Curtis, B., and Chrissis, M. B. 1995. *The Capability Maturity Model: Guidelines for Improving the Software Process*. Addison-Wesley, Reading, Mass.

Sawyer, P., Viller, S., and Sommerville, I. 1998. Requirements process improvement through the phased introduction of good practice. *Softw. Proc. J. 3*, 1, 19–34.

Sawyer, P., Sommerville, I., and Kotonya, G. 1999a. Improving market-driven RE processes. In *Proceedings of International Conference on Product-Focused Software Process Improvement (Profes '99)* (Oulu, Finland), 222–236.

Sawyer, P., Sommerville, I., and Viller, S. 1999b. Capturing the benefits of requirements engineering. *IEEE Softw. 16*, 2, 78–85.

Sommerville, I. and Sawyer, P. 1997. *Requirements Engineering: A Good Practice Guide*. Wiley, Chichester.

Received August 2003; revised April 2004 and September 2004; accepted September 2004