A Review of Methods for Evaluation of Maturity Models for Process Improvement

Yeni Yuqin Li Helgesson, Martin Höst, Kim Weyns
Software Engineering Research Group
Department of Computer Science
Lund University, Sweden
(Yeni.Li_Helgesson, Martin.Host, Kim.Weyns)@cs.lth.se

Abstract

Maturity models are widely used in process improvement. The users of a maturity model should be confident that the weak points of the assessed processes can be found and that the most valuable changes are introduced. Therefore, the evaluation of maturity models is an important activity. In this paper, a mapping study of the literature on the evaluation of maturity models is presented. Two databases are searched resulting in a set of relevant papers. The identified papers can be classified according to six categories, namely the maturity model under evaluation, type of evaluation, relation of the evaluators/authors to the maturity model, level of objectivity, main purpose of the paper and size of study. Further, a framework of different evaluations of maturity models is developed, and the relevant papers are mapped to the framework. Finally, the relevant research on the evaluation of the maturity models in the CMM-family is discussed in more detail. The result of this mapping study is a clear overview of how the evaluation of maturity models has been done and some discussions are provided for further research on the evaluation of commonly used or newly-developed maturity models.

Keywords: mapping study, evaluation, maturity models, software process improvement

1. Introduction

In order to obtain improvements in the software development process and in the resulting product, the software process can be changed with the help of a software process improvement program. When deciding which improvements to implement, there are a number of different changes that are possible and it is in many cases not clear exactly how to change the process in the best possible way. This is one reason why maturity models are used during process improvement. This kind of model is often used as a guide in the improvement work in order to identify which improvements to introduce in the process and at which time. In practice, this is carried out in a number of phases. First the process that should be improved is assessed based on a maturity model. Then the result of the assessment is used to identify which improvements that should be introduced to the process to increase the maturity level of the process.

Some maturity models have been used to a large extent in process improvement in many software development organizations. Examples of this type of maturity models are CMM, CMMI, and the ISO standard based on the SPICE project (e.g. [1]). These models each have their own advantages and disadvantages, for example as described in [2]. To some extent, these models have been accepted as an international software process assessment state of practice.

In addition to the commonly used models, the basic methodology of using maturity models in process improvement is also useful in other domains, such as in human resources where People CMM has been used. Maturity models are also used in the area of safety culture assessments [3] and IT management [4]. Further on, maturity models are also defined for special purposes in cases where new models or adaptations are seen as more suitable than the traditional models. For example, in [5], a set of key practices from CMM is adapted to small organizations. Formulating maturity models is often seen as a useful way of communicating best practices, since it packages the knowledge in a way that makes it useful in improvement initiatives.

Since maturity models are commonly used in process improvement, it is important that they are effective in the sense that they identify the right improvement proposals. The users of the maturity models should be confident that the changes that are introduced really result in improvement and that there are no other changes that would result in significantly more value. This is the basic idea of maturity models, i.e. to outline a path to improvement based on well-proven and accepted improvement steps. If it is not clear that the models are effective and that they point at the right improvements, potential users will hesitate to use them.

An empirical approach to showing that a model or method is effective is to make an evaluation of the model and, based on that, draw conclusions on its effectiveness, and maybe also compare it to other models. Especially in cases when a new maturity model is developed, or a standard model is adapted, it is important to be able to show that it guides the user to the right improvements. However, it requires significant effort to evaluate this kind of models. If it is required that they are used in a large enough set of improvement initiatives in order to be able to draw significant results about effectiveness, this probably requires much time and effort. Instead, evaluations have also been conducted by other means. There are a number of ways that have been used in the literature, e.g. by investigating the differences between different assessors [6], or by analyzing the effect on one organization in a case study [7].

It is by no means easy for a researcher to decide how to evaluate a newly developed or existing model, or even to understand which evaluation approaches are available. This means that there is a need for a framework, describing the different kinds of evaluations that can be carried out for maturity models. In this paper such a framework is presented together with a mapping study [8] where the available literature on maturity model evaluation is mapped to the framework. That is, this paper summarizes what methods have been applied in the literature in order to evaluate maturity models. The result of this review is of course relevant to researchers who are planning to conduct this kind of evaluations. The result is also relevant to practitioners in the area of SPI since it provides an overview of existing evaluation methods, which is useful e.g. when interpreting the results of presented evaluations of existing models. For example, if a model is presented, and it is claimed that it is evaluated based on a set of studies, this study will make it easier to understand the completeness of the evaluations and to identify possible evaluation studies that also could have been conducted.

This means that the main contribution of this paper is not to say which model is the most effective but to summarize the evaluation methods that have been used. It would also be of interest to investigate the effectiveness of the different maturity models, but for most maturity models the data that is available in the selected papers is insufficient for this purpose. In terms of meta-analysis it was only possible to more qualitatively summarize the collection of evaluations of the most widely applied maturity model (CMM) as part of this research. This is discussed in Section 6.

2. Related work

Even if there are no standard evaluation frameworks for evaluating tools, models or frameworks in the field of software engineering, some researchers have done valuable work in this area already.

One such evaluation framework was developed by Kitchenham et al. [9,10]. They constructed a risk-based software bidding model and proposed an evaluation framework to evaluate their bidding model. The framework worked well as a validation framework, and first trials confirmed the benefit of the framework, and allowed for some adjustment to the framework according to the evaluation results. The evaluation process was confirmed to be important and the evaluation framework was suggested to be applicable to a broad range of models. This is related to the evaluation of maturity models, for example in that the evaluation task is complicated by the fact that it is very hard to judge whether decisions taken were right or other decisions should have been taken.

Another framework for evaluating tools and models in information system research has been proposed by Ågerfalk [11]. He makes a distinction between internal grounding, external theoretical grounding and empirical grounding. These three categories are closely related to the three types of evaluation proposed in this paper. The types of grounding described by Ågerfalk are not limited to maturity models or even to process improvement frameworks in general, but can be applied to a wide range of methods, while the types of evaluation presented in this paper are more specific for maturity models.

Pfleeger et al. [12] reported on the results of the Smartie project (Standards and Methods Assessment Using Rigorous Techniques in Industrial Environments), where they present a framework for evaluating software engineering standards. They discuss the difficulties in evaluating large process improvement frameworks and focus especially on the questions that should be asked during such an evaluation. However, no framework for maturity model evaluation was presented.

Finally, Vaishnavi et al. [13] have published a validation framework for maturity models, which they have used to validate their Formal Specification Strategies Maturity Model. Their framework focuses especially on the requirements for empirical validation of maturity models through application in case studies.

Systematic reviews and mapping studies have been conducted in different studies [14] in widely different areas such as cost estimation (e.g. [15]), open source software (e.g. [16]), and testing (e.g. [17]). However, to the best of our knowledge, none of them has looked specifically at the evaluation of maturity models in general. Nevertheless, for the most well-known maturity models, there have been meta-evaluations that summarize selections of previously published results about the evaluation of these specific maturity models. Often the scope of each of the individual evaluations is quite limited because of the large effort involved in evaluating a maturity model, therefore these meta-evaluations actually provide the most complete evaluation of the maturity models. Of course, this approach is only possible for well-established maturity models.

The most famous of these maturity models is probably the Capability Maturity Model Integration (CMMI) [18]. CMMI is a process improvement approach that provides organizations with the essential elements of effective processes. This framework, together with its predecessors and related frameworks, has to a large extent influenced the software industry and the research in software process improvement. McGarry et al. [19] have

published a large study with the result of CMM in over 90 projects. They also explicitly make a distinction between two kinds of evaluations: through direct measurements of an organization's results and a more subjective evaluation through surveys among organizations using the framework. Galin et al. [19] have summarized the results of 19 previous studies on the effectiveness of CMM programs. Both these papers explicitly state the need for more empirical, measurable evaluations of process improvement frameworks. Because of the importance of CMM and because of the large amount of publications available about CMM, Section 6 of this paper discusses the evaluations of CMM identified in this mapping study in more detail.

The SPICE (Software Process Improvement and Capability dEtermination) framework aims to deliver the ISO standard of software process assessment [21]. After the first and second versions of SPICE were published, large international SPICE trials were organized to evaluate different aspects of the framework. Jung et al. [22] and El Emam et al. [23] have published reports summarizing the goals and the results of these evaluations.

From these studies it is clear that there are many different ways to evaluate a maturity model and that there is little consistency in which evaluation methods are currently being used. This shows that there is a need for a broader summary of approaches for evaluating maturity models and classifying how models have been evaluated in the past.

3. Evaluation framework

It is important to distinguish between 'assessment' and 'evaluation' in the context of maturity models. Maturity models are used to assess and improve development processes and other processes. Also the maturity models themselves can be evaluated and improved and this is the focus of this mapping study. It can be noticed that both assessments and evaluations involve a set of questions that are asked by an investigator. In the assessment the objective is to understand and improve the process and in the evaluation the objective is to understand and improve the maturity model. An evaluation can be conducted independently of assessments or based on the results of assessments.

There are different possible ways to evaluate a maturity model. To classify the different ways of evaluating a maturity model, this paper proposes the following framework consisting of three types of evaluations, based on the experience of the authors in this field:

- Type 1: A type 1 evaluation is conducted "off-line", only by the authors of the evaluation without involving any outside experts. A type 1 evaluation can be done based on, for example, their knowledge of the processes it is intended to be used with or by comparing it with other similar frameworks. The evaluators could also be the authors of the maturity model itself.

- Type 2: A type 2 evaluation is conducted by involving practitioners, who are the experts on the type of process that is intended to be improved by the maturity model, but who have not been involved in the actual development of the maturity model. In a type 2 evaluation no real assessment is carried out, instead interviews, surveys or simulated assignments can be carried out.

- Type 3: A type 3 evaluation is conducted through real process improvement activities where the maturity model is used in a practical setting.

Type 1 evaluations requires the least effort of the three since they only require that people who have knowledge about it can spend time elaborating different aspects of the model. Type 2 evaluations are more complicated since they require the cooperation of practitioners. Type 3 evaluations can be regarded as the most costly since they require that an assessment is carried out and that the result is used not only to analyze the investigated process but also to analyze the maturity model. That is, if the assessment is carried out in order to evaluate the maturity model, the cost of this evaluation type can be regarded as very high. However, assessments can be carried out with the purpose of at the same time improving the process, and then the expected benefits can be used to justify this cost.

The three types of evaluations imply an order in which they can most logically be used to evaluate a maturity model under development. First the maturity model is evaluated by experts on the maturity model, then by experts on the process that is improved, and finally it is used in improvement programs. For example, a maturity model may be formulated for improvement of an IT management process. It could first be evaluated in a type 1 evaluation where experts on the model evaluate factors like understandability and internal consistency. After that, independent practitioners from both IT management and users of IT systems could evaluate the contents of the model in a type 2 evaluation to validate how well the model corresponds with the current state of practice. When these evaluations have been carried out it is reasonable to use the model in a series of assessments where extra effort is spent on evaluation of the developed maturity model, i.e. type 3.

It is of course also possible to iterate or to carry out the evaluations in any order that is seen as useful. In the example above, it can, for example, be necessary update and reevaluate the model based on the feedback from the earlier evaluations.

4. Research methodology

The research in this paper is conducted in two major steps. First a framework for the evaluation of maturity models is defined based on knowledge of the area and commonly cited literature. The defined framework is then used to classify papers in a mapping study of published literature in the area. The mapping study representing the major part of the presented research has two main objectives, both to identify literature and to investigate the usefulness of the defined framework for classification of research in the area.

The review focuses on the evaluation of maturity models. Mapping studies are to a large extent carried out in the same way as systematic reviews, as described by Kitchenham [24]. A difference is that while a systematic review is conducted in order to identify best practice in an area based on presented research the objective of a mapping study is more to classify conducted research [8]. However the systematic approach to identifying relevant research can be conducted in the same way. This research is carried out based on the guidelines for performing presented by Kitchenham [24]. The procedure of this mapping study includes the following steps: planning, defining research questions, searching the databases, discussion of validity, data extraction and synthesis of the results. These steps are described in the next subsections.

Three researchers were involved in this research, and they are the authors of this paper. From here on the term ‘authors’ is used to mean these researchers.

4.1 Planning

The goal of the mapping study is to find out how the evaluation of maturity models is conducted and to prove the usefulness of the framework introduced in Section 3. A review protocol was developed in the beginning of the mapping study. The protocol includes the research background, the research questions, search strategy, study selection criteria and procedures, quality assessment, data extraction and data synthesis strategies. The intention is that this review protocol should make sure that the study is undertaken as planned and not driven by researcher expectations. In this review protocol, the whole study timetable was not decided from the beginning but the actual timetable of the study and results produced were recorded as the study progressed. The research questions and article identification strategies are described in the following sub sections.

4.2 Research questions

In this research, the methods used for the evaluation of maturity models are classified and mapped to different types in the defined framework presented in Section 3. The result is an overview of which evaluation methods have been used. This can guide researchers in choosing suitable methods for future evaluations of maturity models.

This can be formulated in the following main research questions:

RQ1: What research has been conducted and reported in the area of the evaluation of maturity models?

RQ2: To what extent is the framework presented in Section 3 useful for classification of the approaches to the evaluation of maturity models?

RQ3: How can the framework be further extended in order to support researchers and practitioners developing evaluation approaches for maturity models?

Basically RQ1 is investigated through this mapping study, RQ2 in an extension of the mapping study, where the identified research papers are mapped to the structure of the framework. RQ3 is answered by reflections on the work with RQ1 and RQ2.

4.3 Search strategy and search process

4.3.1 Search resources

This study was planned to find relevant literature about the evaluation of maturity models. Based on the fact that most relevant papers for this study are in the software engineering area, two electronic databases were searched:

- INSPEC: This database is provided by Elsevier Engineering Information Inc. and the Institute of Electrical Engineers (IEE). It includes papers from 1969 to present.
- COMPENDEX: This database is provided by Elsevier Engineering Information Inc. It includes papers from 1970 to present.

Both databases intend to provide a complete coverage of the area, and include papers from all conferences, journals, and publishers (e.g. IEEE, ACM, Springer, and IEE). These two

databases are by many experts seen as the leading databases in e.g. Computer Engineering and Electrical engineering and Electronics. Both databases were accessed through Engineering Village (http://www.engineeringvillage2.org).

Searches in the databases were carried out with a search string that is presented below.

Besides these two databases, the Journal of Software Process Improvement and Practice was searched manually. This journal was chosen because it is known to include published software process improvement approaches. The papers of this journal from 1995 to 2009 were searched, i.e. all available issues at the time the database searches were carried out.

4.3.2 Search process

After some tried searches, the following search string was decided on for this study.

((Evaluating WN KY) OR (evaluation WN KY)) AND
(((maturity model*) WN KY) OR (CMM* WN KY) OR (SPICE WN KY))

The first line is intended to make sure that the papers have to do with the evaluation and the second line is intended to make sure that the paper is about maturity models. The term "WN" means existence, i.e. that the work before should exist in some part of the paper. "KY" denotes that certain words should appear in the title, abstract or key word list of the paper.

The search string was used on the two electronic databases on June 30th, 2009. After duplicate papers were removed, 1722 unique papers remained. After removing the papers that are obviously out of the area, 338 papers remained.

After this the Journal of Software Process Improvement and Practice was searched manually on July 23rd 2009. In this journal, 21 relevant papers were found. All of these papers were already included in the list of identified papers, which shows that the initial search was able to identify these important papers.

Afterwards, the references of a selection of the most relevant papers were manually checked and as a result, another 5 papers were added.

4.3.3 Inclusion and exclusion criteria

A paper is kept in this mapping study if it satisfies one of the following two criteria:

- The paper is the report of an evaluation of one or more maturity models to identify merits and/or weakness of the models.
- The paper is the report of experience of using one or more maturity models in organizations for process improvement and some feedback was provided in order to improve the models. This includes papers whose main purpose is not to evaluate the maturity model, but still present an empirical evaluation of the used maturity model.

The papers were first reviewed based on titles, abstracts and key words, and they were classified in three different types:

- Relevant papers: if the paper satisfies one of the two inclusion criteria.
- Process assessment papers: if the paper is related to process assessment with maturity models, but not related to the evaluation of maturity models.

- Excluded papers: other papers, which are not relevant to either evaluation of maturity models or process assessment.

One author reviewed all 343 papers and put them into these different types according to the previous criteria. Then the other two authors checked part of the results, and found that some disagreement existed. The conclusion was that the disagreement was large enough to warrant the extra work of rechecking the classification of the rejected papers. Therefore the other two authors reviewed those papers that were excluded or classified as process assessment related, and re-added some papers into the relevant papers group. When there was doubt about the classification of a paper, it was included in the relevant group, leaving the possibility to discard the paper during the next phase when the full papers were studied. The result of this stage was that 116 papers were classified as relevant to the evaluation and assessment of maturity models.

4.3.4 Classification

The three authors reviewed one third of the papers each, based on the full texts, and classified the papers into three types according to the framework in Section 3. After the first round of classifications, each author had some papers for which he/she wanted to discuss the classification. Then all authors reviewed those papers, and discussed their results together, and found they agreed on the result of the second round classification, so the classification result was accepted by the research team.

During the final step of the selection, 55 papers were excluded based on their full texts, 2 papers turned out to be duplicates that were not detected before because the names were slightly different, and 59 papers were selected as relevant papers for this mapping study.

The search process is summarized in Figure 1.

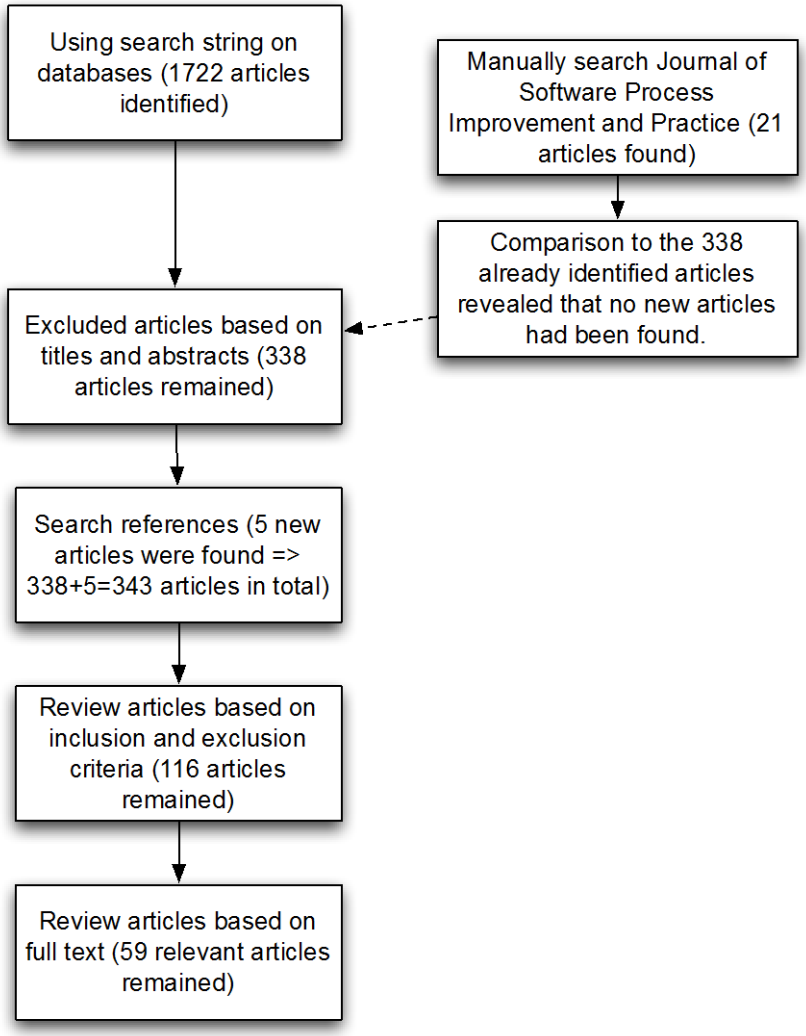

Figure 1. Paper identification process

4.4 Classification validity

The goal of this study is to cover as many as possible of the relevant research papers about the evaluation of maturity models. Nevertheless, it is likely that some relevant papers have been missed. This can be attributed to a number of different reasons. First of all, only papers published in English were considered. Because of practical reasons it was not possible to search through the scientific literature in other languages. Secondly, even in the English language there is some ambiguity and the search string described in Section 4.3.2 was chosen to include as many relevant papers as possible, while still returning a practical number of results. This means that some relevant papers that use a different terminology might not have been found. Thirdly, some lesser known journals and proceedings are not included in the electronic databases that were searched and any possible papers published in these collections were therefore not included in the results. A fourth reason is that the terms used were from a typical engineering perspective. It may be the case that the same type of model is used in other domains with other names. Finally, some papers can also have been rejected incorrectly during the selection process from the search results to the final list of relevant papers.

The scope of the mapping study is not limited to maturity models in the field of software engineering, although this is where maturity models have been the most popular and it is also the main focus of the mapping study. Because of the software engineering background from

the authors and the way the search string was constructed, the selection of relevant papers is more likely to have missed relevant papers from other fields where also a different terminology is more likely to be used.

The following measures have been taken to improve the validity of the research and to minimize the number of missed papers:

- The inclusion and exclusion criteria at every step were explicitly defined and agreed upon by all authors. This makes the results from different authors more consistent and objective.

- For the selection based on abstract, after one author classified all relevant papers, the other two authors rechecked the rejected papers and further added any papers that they considered as possibly relevant. At any time when one author was not certain about the classification, the other authors looked at that particular paper and the final classification was decided when all authors agreed on it.

- To check whether the search string returned all the most relevant papers, the most important journal in the field of interest, the Journal of Software Process Improvement and Practice, was searched manually based on the content of the article and no new relevant papers were found.

- To further complete the list of relevant papers, the most important papers in the list were selected and their reference lists were systematically searched for relevant papers. This resulted in the identification of an extra 5 relevant papers that had not been included in the list before. Three of these papers were from CrossTalk, a journal not included in the databases, the remaining two had not been returned in the automatic search results because they used a slightly different terminology.

These measures together give us a good degree of confidence that most of the relevant papers, at least in the software engineering field, have been identified, although there is a risk that some less influential papers have been missed. Therefore, this mapping study cannot guarantee completeness, but can still be trusted to give a good overview of the relevant literature on the evaluation of maturity models, especially in the field of software engineering.

4.5 Data extraction

The data extracted from each paper were maintained through the whole review process. After identification of the relevant papers, the following data were extracted:

- The source (journal or conference)
- Title
- Authors
- Publication year
- The evaluation or assessment type: type 1, 2 and/or 3
- Summary of the research, including which questions were solved

To be able to analyze the 59 papers there was a need to classify them in more ways than just according to the framework defined in Section 3. For this purpose further criteria for

classifying the papers were defined and discussed by the research team, based on what information was available in the papers. When needed the categories were updated or clarified during the classification process. The result is presented in Table 1.

Table 1 Information collected from papers

Evaluation type, select one of	Static evaluation only by the authors of the paper (type 1 according to the framework)
	Evaluation by external experts in the field (type 2 according to the framework)
	Evaluation through process assessment (type 3 according to the framework)
Size, report the following metrics	Number of case studies for type 3
	Number of independent people involved in type 2
Relation of the evaluators to the maturity model	Evaluation of their own MM
	Independent evaluation by others
	Part of an official evaluation of MM (e.g. SPICE trials)
Objectivity of evaluation, select one of	Objective
	Subjective
	Both subjective and objective
Purpose of the papers, select one of	Evaluation of MM is the only purpose of the paper
	Evaluation of MM is a main purpose of the paper
	Evaluation of MM is carried out in addition to another main purpose of the paper
Maturity model of interest, select one of	CMM*
	SPICE
	Special adaptation of an famous MM (e.g. CMM for "small organizations")
	New, self-developed MM
	Other maturity models, for example, BOOTSTRAP

Based on the criteria for classifying the papers, all relevant papers were reviewed and the corresponding data were extracted.

It is not easy to identify an evaluation as objective or subjective. In this study, an evaluation is identified as objective if it looks at the correlation between maturity level and objective data like faults/KLOC, productivity, etc. An evaluation is identified as subjective if interviews about the maturity model are used to evaluate the model, or if they look for correlation between maturity level and employee satisfaction and how people evaluate the maturity model. Some papers are identified as containing both a subjective and an objective evaluation.

The Carnegie Mellon Software Engineering Institute has published a series of frameworks since the original publication of CMM, for example CMMI [17]. The term CMM* is used in this study to refer to all related maturity models together.

The last categorisation partly overlaps with the third categorisation, especially concerning the self-developed maturity models. The last categorisation is mostly useful for being able to extract those papers that concern CMM* for further analysis in Section 6.

4.6 Data synthesis

The data synthesis was specified in the review protocol from the beginning of the mapping study. When there was any uncertainty about the classification of the primary studies, the issue was discussed by all authors until agreement was reached.

5. Results

5.1 Introduction

Appendix A lists the relevant articles from the mapping study. 59 articles are identified as relevant for this study. In this section, the classification of the articles according to Section 4.5 is discussed.

In Figure 2 the publication years for the identified articles are displayed. From this data it is not possible to identify any clear increasing or decreasing trend. Instead it can be concluded that there continuously have been published articles in the area.

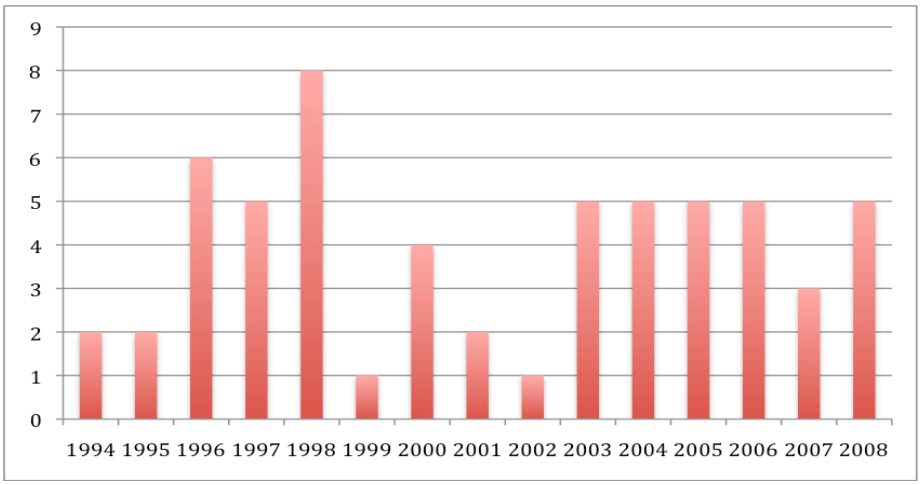

Figure 2. Histogram of publication year for the identified articles

5.2 Classification according to the framework

The evaluation type of each paper is shown in Table 2.

Table 2 Evaluation type of papers

Evaluation type	Identified articles	Number of articles
Type 1	6, 11, 28, 30, 35, 37, 41, 46, 50, 51, 52, 56	12
Type 2	1, 4, 54, 59	4
Type 3	2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 20, 21, 22, 23, 24, 25, 26, 27, 29, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 42, 43, 44, 45, 47, 48, 49, 51, 53, 55, 56, 57, 58, 59	50

Table 2 shows the classification with respect to evaluation type. It can be seen that 12 out of the 59 articles are classified as type 1, 4 articles are classified as type 2, and 50 articles are

classified as type 3. It should be noted that a few articles have been classified as more than one type, e.g. article 12 as both type 1 and type 3.

This result shows that most of the cases used to evaluate maturity models are type 3, in which maturity models are evaluated by using them in software process improvement. Only 4 out of 59 cases used type 2, and half of them were combined with a type 3 evaluation.

An example of how a type 3 evaluation can be carried out is presented in article 15, where data from already conducted assessments are used to evaluate CMM. Experience reports were collected from a database (PAIS). 707 of 948 reports were about CMM, which means that CMM could be evaluated with respect to internal consistency and correlation between related factors (dimensionality). This is a rather large evaluation with respect to number of included assessments.

An example of a type 1 evaluation is presented in paper 30 where two maturity models are compared by the researchers with respect to simplicity, validity, robustness, prescriptiveness, and analyzability.

5.3 Maturity models

The maturity models that were evaluated in the identified articles are of 5 types: CMM* (i.e. any type of model related to CM, such as CMM, CMMI), SPICE, adaptation of official Maturity Models (e.g. CMM for small organizations), self-developed maturity models, and other maturity models (common models but not so widely used as CMM or SPICE). The classification with respect to evaluated maturity models is presented in Table 3.

Table 3 Evaluated Maturity Models

Maturity model type	**Identified articles**	**Number of articles**
CMM*	3, 4, 5, 6, 9, 13, 16, 19, 20, 22, 30, 31, 34, 35, 36, 38, 41, 43, 44, 45, 47, 48, 50, 52, 57, 59	26
SPICE	8, 11, 15, 18, 19, 23, 24, 25, 26, 27, 28, 36, 52	13
Adaptation of official MM (e.g. CMM for "small organizations")	7, 10, 14, 21, 29, 46, 53	7
Self-developed MM	1, 2, 12, 17, 32, 33, 37, 39, 40, 42, 44, 51, 54, 56, 58	15
Other maturity models, for example, BOOTSTRAP	49, 55	2

Table 5 shows that 26 articles were about evaluation of CMM*, 13 were about SPICE, 3 of which were about the evaluation of both CMM and SPICE. 7 were about adaptations of official maturity models. 15 evaluated their self-developed maturity models. One of them compared with CMM also. Only two papers were about evaluation of other common maturity models.

More than half of the studies were about CMM* or SPICE. About one fourth of the evaluations were about self-developed maturity models. Only a few were about adaptation of official maturity models, and even fewer cases were about evaluation of other common

maturity models. In Section 6 the articles presenting research on CMM* are discussed in more detail.

5.4 Relationship of the authors to the maturity model

Table 4 shows that 16 of 59 articles were written by researchers that had developed their own maturity model and evaluated it, of course these are nearly exactly the same articles as in the category of self-developed maturity models in the previous sections. Most of these maturity models are from the software engineering field, though this can partly be explained because we focused our search on this area. The evaluation of the maturity models is an important step in the development of it, and it is normal that the first evaluations are performed by the developers of the model.

Table 4 Relationship of the authors to the evaluated maturity model for each of the papers

Evaluation group	**Identified papers**	**Number of papers**
Evaluation of their own MM	1, 2, 7, 12, 15, 17, 32, 33, 37, 39, 40, 42, 51, 54, 56, 58	16
Independent evaluation by others	3, 4, 5, 6, 8, 9, 10, 13, 14, 16, 20, 21, 22, 23, 25, 28, 29, 30, 31, 34, 35, 36, 38, 41, 43, 44, 45, 46, 47, 48, 49, 50, 52, 53, 55, 57, 59	37
Part of the official evaluation of the MM	11, 18, 19, 24, 26, 27	6

6 of the 59 articles are part of an official maturity models evaluation, and all of those are from the SPICE trials, since this is the only maturity model that has been systematically evaluated in an official set of organized trials.

Most of the articles, 37 out of 59, discuss the evaluation of internationally accepted models by researchers or professionals that were not directly involved in the development process of the maturity model. These articles are most likely to be unbiased and can evaluate how easy the evaluated maturity model is to use as a finished tool for an organization without the involvement of the authors of the maturity model.

This classification is also important because the context in which the evaluation was performed has a large influence on the way the evaluation can be performed. In the evaluations of self-developed maturity models, most studies are quite small because of the huge effort required in this kind of evaluation. The advantage of these evaluations is that the developers of the maturity model are directly involved in the application of the model and can check that all participants in the study apply the model consistently. In the evaluation of internationally renowned maturity models such as CMM, there is often a lot of data available because many organizations are using these models. However, the difficulty here lies in that the researchers have very little control over how these organizations apply the maturity model, making it hard to compare data from different organizations.

5.5 Objectivity

The evaluation objectivity of relevant articles is shown in Table 5. Articles in the subjective category are those using questionnaires and interviews to investigate the applicability of the maturity models and whether the effects of the maturity model are experienced as positive.

The papers in the objective category investigate the direct effect of the maturity model on metrics such as productivity and fault rate.

McGarry et al. [19] note that many evaluations are based on mostly subjective judgments of the effectiveness of maturity models and that there is a need for more objective evaluations. Table 5 shows that the mapping study identified more subjective than objective papers, but the difference is relatively small. Because both a subjective and objective evaluation are important in the complete evaluation of a maturity model, it is positive that there are many papers in each of the categories and that there are many articles combining both approaches.

Table 5 Evaluation objectivity

Evaluation objectivity	**Identified articles**	**Number of articles**
Objective	5, 6, 12, 13, 15, 18, 19, 20, 23, 24, 25, 26, 31, 32, 34, 38, 43, 45, 55, 59	20
Subjective	2, 4, 7, 9, 14, 16, 17, 21, 22, 29, 30, 33, 37, 39, 40, 41, 46, 48, 49, 50, 51, 52, 54, 57	24
both subjective and objective	1, 3, 8, 10, 11, 27, 28, 35, 36, 42, 44, 47, 53, 56, 58	15

It should be noted that this classification is not always easy to make. For the articles of type 3 of the framework, all maturity model assessments automatically contain a subjective factor. This is not what is meant with the subjective category in this study. The subjective/objective classification only takes into account the other variables studied, which are in each study being compared to the measured maturity levels. For some papers it can be hard to extract how some of the study data was collected making the classification of objective or subjective more uncertain. For the articles from type 1 and 2 of the framework the classification of objective or subjective is based on the used evaluation methods.

5.6 Size

The size of each study represents how many persons were involved in the study, how many companies or how many divisions of a company were included in the study, etc.

The extracted size data about the papers shows that it is generally very hard to generalize the size data or even to compare the size of a study presented in one article with another since different measures are reported in the different articles. Therefore the size data cannot be summarized statistically, nevertheless it can be seen that the list of relevant articles contains everything from very small to very large-scale evaluations.

For example, 545 survey participants were involved in a survey about CMM KPA in article 4, and this covers about half of the companies in CMM-based SPI programs at the time. 6 persons in one organization were involved in the evaluation of a requirements quality model in paper 54. In article 42, 18 companies were involved in a case study. The experience of 30 platforms in around 10 divisions was reported in article 33. However, in article 35, only one organization was involved in the evaluation, and in article 29, one case study was done in a virtual software organization.

6. Analysis of research on CMM*

In this work many articles on software process improvement based on different maturity models have been identified and analyzed. Based on this it is natural to say something about the general effectiveness of the models in process improvement. It would, for example, be interesting and valuable to determine the mean value of the effects of improvements based on different models. However, based on the identified articles it is not possible to do exactly this investigation for a number of reasons.

One reason that it is not possible is that it is very different how many studies there are for the each model. For CMM* and SPICE there are rather many articles identified, while there are only a few studies for each of the other models. Another reason is that the focus of the identified studies differs very much. Concerning SPICE, the focus of the studies is, as described above, not really on effectiveness, but more on e.g. consistency of different reviewers when interpreting factors of the model.

Concerning the articles on CMM* they investigate rather different factors, which makes the analysis of effectiveness from the articles impossible. However, it possible and relevant to further analyze the evaluations of the maturity models in the CMM family by summarizing the research that has been conducted. With CMM* we denote all models in the CMM family, i.e. both CMM and CMMI.

26 articles evaluating CMM* have been identified in this study and together they present many different types of evaluations. The largest group of these evaluations (articles 3, 5, 13, 34, 38, 45 and 47) presents metrics collected from many projects across different maturity levels, at one or more organizations. These articles make it possible to statistically analyze the benefits and/or costs associated with an organization reaching a higher maturity level. This is the strongest proof for measurable improvements in an organization from using software process improvement. Article 20 even presents a meta-analysis of a number of these metrics from different independent usages of CMM*. Another similar overview of accumulated evidence of software process improvement in general, including a number of cases using CMM* can be found in [25].

The collection of this kind of metrics is very valuable but is often quite difficult in practice. Therefore, another common way to evaluate the effects of CMM* is with a large survey among organizations that are using CMM*, often with a specific focus. This approach is used in five of the identified articles (articles 4, 16, 22, 36 and 53), where the first and the last article focus especially on the problems faced by small and medium-sized businesses.

Three other articles (articles 9, 35 and 57) present experience reports of using CMM* for process improvement. These articles do not focus on presenting measured benefits or on collecting large amounts of data, but instead on describing in detail the positive effects and practical problems experienced by one organization starting out with software process improvement.

Three articles (articles 19, 31 and 43) present a statistical analysis of the data from a large number of CMM* assessments to investigate the internal consistency of the assessment methods.

All the articles discussed above are clearly type 3 evaluations. A number of type 1 evaluations of CMM* have also been performed. Three articles focus on one specific element of software engineering such as testing (article 6), security (article 41) or software architecture (article 50)

and study how the best practices in this area are supported by CMM*. This allows them to identify some very specific possible shortcomings in CMM* and to propose an own framework that complements CMM* and specifically targets this area.

Another type 1 evaluation is to compare the assessment method of CMM* to some other common process improvement methods such as BOOTSTRAP (article 30), GQM (article 44) or SPICE (article 52).

Finally, there is also one article (article 48) that offers a detailed discussion of the advantages and disadvantages of using CMM* levels as a requirement when selecting contractors from the point of view of the US Department of Defense.

7. Discussion

For a more detailed discussion of the distribution of the relevant articles it is also important to look at the combination of different classifications. More than half of the articles present evaluations of CMM or SPICE, and these evaluations are done mainly as type 3. About one fourth of the articles are evaluations of self-developed maturity models, and these evaluations are also done mainly as type 3.

All reported type 1 evaluations are done independently by others than the model developers. Possibly this type of evaluation has also been done by the developers themselves but then these evaluations have not been reported separately, and only those done by others have been reported. Only two articles used only type 2 evaluation. Both of them are evaluations of self-developed maturity models. Most of the type 3 evaluations are independent evaluations by others, and only few of them are part of the official evaluation. This is reasonable, because with a type 3 evaluation, the model should be used by others and the model's effectiveness is assessed according to that.

Most of the self-developed maturity models were evaluated by the developers first, because the related articles are classified as evaluation of their own maturity models. This conforms to what we would expect to find from our own experience. Most of the independent evaluations are done regarding CMM and related models. This shows that CMM has been used for quite some time, and quite a few results have been published about it.

Most of the evaluations of self-developed maturity models are side effects of the articles, which shows that the developers of the maturity models usually publish the result of the model evaluation together with their maturity models, and that the evaluation results are only a small part of their articles. When the official evaluations are published, the evaluations of maturity models are usually the only purpose of the articles, and the models themselves are not explained, because the models were published already.

The evaluations of self-developed maturity models are more often subjective than the independent evaluations by others. As expected, all official evaluations are either classified as objective or as both objective and subjective. None of these evaluations are classified as subjective.

Regarding the relation between the objectivity of evaluations and the purpose of the articles, most of the objective evaluations are published as the only purpose of the articles. These objective evaluations are done by others than the model developers, so they usually only report the results of evaluations, but many articles by model developers explain the model and then some result from evaluation. This result is shown by the fact that most of the subjective

evaluations are published as a side effect of the articles. Both subjective and objective evaluations are usually published as main purpose of the articles, and at the same time some other evaluations are published as only purpose or side effect, because the evaluation may be published separately or together with the model.

For the papers whose side product is maturity model evaluation, most of them (two thirds) are evaluation of self-developed maturity models. The rest of them are about evaluations of adaptations of official models. This shows that evaluation results are often published together with a model only when the model is new or has been changed for a special purpose. Otherwise there is no need to explain a model that is already published, or even has been commonly used.

The presented mapping study also shows that there are large differences in the size of the evaluations in the relevant articles. Some articles combine data from hundreds of assessments at many different companies, while others are limited to a few projects at one organization. Generally, it is hard to compare this aspect of the studied papers because not all studies report all details of the data collected. An evaluation with many companies involved is not necessary a larger study than another evaluation with only one company involved, because of the differences in the size of the companies and the number of people involved in the studies.

8. Conclusions

In this paper, a mapping study of the evaluation of maturity models is presented. 59 relevant articles are selected from two electronic databases. The selected articles are classified according to the evaluation types, the relation of the evaluators to the maturity model, the objectivity, the purpose of the articles and the maturity model of interest.

A framework of different evaluations of maturity models is developed, and the relevant articles are mapped to the framework according to their classified evaluation types. The result of the mapping study shows that the type 3 evaluation, i.e. evaluation is conducted by actively using the maturity model in a process improvement effort, is the most commonly used evaluation method. Type 2, i.e. the evaluation involves some external experts from the field but without conducting an assessment, is the least used method. Only those type 1 evaluations done independently by others than the model developers were found in the review. Even though most of the published type 1 evaluations are not done by the developers, the evaluations are still classified as subjective, because the evaluations are done off-line and are not the result from using the model in process improvement.

A lot of results are reported about the evaluations of CMM (including other models in the same family) and SPICE. Especially CMM has been evaluated in many different ways and the advantages and disadvantages of this framework have been documented in many research papers. The evaluations of some self-developed maturity models were also found. It is easy to see that quite a few maturity models have been developed, but the evaluation results for self-developed maturity model are not many. It shows that it is quite difficult to put a new maturity model into use when not much evaluation result has been reported for it. To have more users is very important for the development of a maturity model because it means more feedback, more evaluation results, and the model itself can be improved faster.

This mapping study has looked into relevant articles and extracted data of published evaluation methods of maturity models. This can give a good overview for the researcher and

practitioners in this area of how the evaluations have been done, and it can give them guidance in what kind of methods can be chosen for future evaluations of maturity models.

Future work building on the research presented in this paper could, for example, focus on how this framework can be used to identify possible shortcomings in how some maturity models have been evaluated. This way, further interesting opportunities for the evaluation of some important maturity models could be identified. Currently there is not enough empirical data available in the literature to determine how effective developed models are in different situations in a meta-analysis. This research points to the need for providing this kind of empirical data, and an interesting area for future research would be to do this kind of meta-analysis when such data is available.

Acknowledgment

The work is funded by the Swedish Emergency Management Agency under a grant for FRIVA, Framework Program for Risk and Vulnerability Analysis of Technological and Social Systems.

References (in citation order)

1. Jung, H.-W., Hunter, R., The Relationship Between ISO/IEC 15504 Process Capability Levels, ISO 9001 Certification and Organization Size: an Empirical Study, *Journal of Systems and Software* **2001; 59**, pp. 43-55.

2. Wang, Y., Court, I., Ross, M., Staples, G., King, G., Dorling, A. Quantitative evaluation of the SPICE, CMM, ISO 9000 and BOOTSTRAP, *In proceedings of Software Engineering Standards Symposium and Forum*, **1997**, pp. 57-68.

3. Fleming, M. Safety Culture Maturity Model. *Offshore Technology Report*, **2000; 049**, 2001.

4. Luftman, J.N. Managing the Information Technology Resource: *Leadership in the Information Age*. Prentice-Hall, **2003**.

5. Hosny, H.M. The CMM Software Quality Assurance Process Scaled Down for Small Organizations, *In proceedings International Conference on Electrical, Electronic and Computer Engineering, ICEEC*, **2004**, pp. 291-294.

6. Fusaro, P. and El Emam, K. and Smith, B. Evaluating the Interrater Agreement of Process Capability Ratings, *proceedings Software Metrics Symposium*, **1997**.

7. Diaz, M., Sligo, J. How Software Process Improvement Helped Motorola, *IEEE Software*, **1997**; **11(4),** pp. 75-81.

8. Kitchenham, B. Budgen, D., Brereton, P. O. Using mapping studies as the basis for further research – A participant-observer case study, *Information and Software Technology*, **2011; 53**, pp. 638-651.

9. Kitchenham, B., Pickard, L., Linkman, S. and Jones, P., A framework for evaluating a software bidding model. *Information and Software Technology* **2005; 47(11),** pp.747-760.

10. Kitchenham, B. and Linkman, S., Experiences of using an evaluation framework. *Information and Software Technology* **2005; 47(11),** pp.761-774.

11. Ågerfalk, P.J. Grounding Through Operationalization - Constructing Tangible Theory in IS Research. *ECIS 2004 Proceedings*. **2004**.

12. Pfleeger, S.L., Fenton, N., and Page, N. Evaluating software engineering standards. *Computer* **1994; 27(9),** pp.71-79, 1994.

13. Vaishnavi, V.K., and Fraser, M. D., Validation framework for a maturity measurement model for safety-critical software systems. *Proceedings of the Annual Southeast Conference*, **1998**; pp. 314-322. Marietta, GA, USA.

14. Kitchenham, B., Brereton, O.P., Budgen D., Turnera, M., Bailey, J. and Linkmana, S. Systematic literature reviews in software engineering - A systematic literature review. *Information and Software Technology* **2009; 51(1),** pp.7.

15. Jørgensen, M. A review of studies on expert estimation of software development effort, *Journal of Systems and Software*, **2004**; 70(1-2), pp. 37-60.

16. Stol, K-J., Ali Babar, M. Reporting empirical research in open source software: the state of practice, *Proceedings International Conference on Open Source Systems*, **2009**; Skövde, Sweden, pp. 156-169.

17. Engström, E., Runeson, P. Software product line testing – A systematic mapping study, *Information and Software Technology*, **2011; 53**, pp. 2-13.

18. SEI, Capability Maturity Model Integration, Version 1.2, CMU/SEI-2006-TR-008(2008): *Carnegie Mellon Software Engineering Institute*, **2008**.

19. McGarry, F., Burke, S. and Decker, B. Measuring the impacts individual process maturity attributes have on software products. *International Software Metrics Symposium, Proceedings*, **1998**; pp. 52-60. Bethesda, MD, USA.

20. Galin, D., and Avrahami, M. Are CMM program investments beneficial? Analyzing past studies. *IEEE Software* **2006; 23(6),** pp.81-87.

21. El Emam, K., Drouin, J.N., and Melo, W., SPICE - the theory and practice of software process improvement and capability determination, Los Alamitos, California: IEEE Computer Society, **1998**.

22. Jung, H.W., Hunter, R., Goldenson, D.R., and El Emam, K. Findings from phase 2 of the SPICE trials. *Software Process Improvement and Practice* **2001; 6(4),** pp.205-242.

23. El Emam, K. and Goldenson, D.R. SPICE: an empiricist's perspective. *Proceedings of the Second IEEE International Software Engineering Standards Symposium (ISESS'95) Experience and Practice* (Cat.No.95TB8092), **1995**; pp. 84-97. Los Alamitos, CA, USA.

24. Kitchenham, B. Guidelines for Performing Systematic Literature Review in Software Engineering. *Technical report*, Keele University and University of Durham, **2007**.

25. Krasner, H., Accumulating the Body of Evidence for The Payoff of Software Process Improvement, *Software Process Improvement*, IEEE Computer Society Press, **2001**, ISBN 0-7695-0999-1, pp 519-539

Appendix A: Relevant papers from the mapping study

1. Alvaro, A., de Almeida, E.S. and Meira, S.L. A software component maturity model (SCMM), *33rd Euromicro Conference on Software Engineering and Advanced Applications, EUROMICRO '07,* **2007**.

2. April, A., Hayes, J.H., Abran, A. and Dumke, R. Software maintenance maturity model (SMmm): the software maintenance process model. *Journal of Software Maintenance* **2005; 5,** vol. Res. Prat. (UK).

3. Bellini, E. and Lo Storto, C. CMM implementation and organizational learning: Findings from a case study analysis. *Portland International Conference on Management of Engineering and Technology*, **2006**.

4. Brodman, J. G. and Johnson, D. L. What small businesses and small organizations say about the CMM. *Proceedings - International Conference on Software Engineering*, **1994**.

5. Brodman, J.G. and Johnson, D. Return on Investment from Software Process Improvement as Measured by U.S. Industry. *Crosstalk*, **1996**.

6. Bueno, P.M.S., Crespo, A.N. and Jino, M. Analysis of an artifact oriented test process model and of testing aspects of CMMI. *Product-Focused Software Process Improvement. 7th International Conference, PROFES 2006. Proceedings* **2006; 4034**.

7. Cass, A., Volcker, C., Ouared, R., Dorling, A., Winzer, L. and Carranza, J.M. SPICE for SPACE trials, risk analysis, and process improvement. *Software Process Improvement and Practice (UK)* **2004; 9(1).**

8. Cater-Steel, A., Toleman, M. and Rout, T. Process improvement for small firms: An evaluation of the RAPID assessment-based method. *Information and Software Technology*, **2006**.

9. Cattaneo, F., Fuggetta, A. and Lavazza, L. Experience in process assessment. *Proceedings of International Conference on Software Engineering*, **1995**.

10. Cepeda, S., Garcia, S. and Langhout, J. Is CMMI useful and usable in small settings? One example. *CrossTalk,* **2008**.

11. Coletta, A. SPICE trials feedback and tuning of the measurement framework. *Conference on Software Measurement and Management*, **1996**.

12. Cook, C.R. and Visconti, M. New and improved documentation process model. *Fourteenth Annual Pacific Northwest Software Quality Conference*, **1996**.

13. Diaz, M., Sligo, J. How Software Process Improvement Helped Motorola. *IEEE Software,* **2007**.

14. Egan, I., Ritchie, J.M. and Gardiner, P.D. The role of CMM as a tool for detecting, measuring and evaluating performance change in the mechanical design process. *5th International Conference on Integrated Design and Manufacturing in Mechanical Engineering (IDMME 2004),* **2004**.

15. El Emam, K. Internal consistency of the ISO/IEC 15504 software process capability scale. *Proceedings of International Software Metrics Symposium*, **1998**.

16. El-Emam, K., Goldenson, D., Mccurley, J. and Herbsleb, J. Modeling the likelihood of software process improvement: An exploratory study. *Empirical Software Engineering*, **2001**.

17. Ericson, T. and Subotic, A. and Ursing, S. TIM - a test improvement model. *Software Testing Verification and Reliability*, **1997**.

18. Fusaro, P., El Emam, K. and Smith, B. Evaluating the interrater agreement of process capability ratings. *Proceedings. Fourth International Software Metrics Symposium* (Cat. No.97TB100196), **1997**.

19. Fusaro, P., El Emam, K. and Smith, B. The internal consistencies of the 1987 SEI Maturity Questionnaire and the SPICE capability dimension. *Empirical Software Engineering (Netherlands)*, **1998**.

20. Galin, D. and Avrahami, M. Are CMM program investments beneficial? Analyzing past studies. *IEEE Software*, **2006**.

21. Grunbacher, P. Software assessment process for small software enterprises. *Proceedings of the EUROMICRO Conference*, **1997**.

22. Herbsleb, J.D. and Goldenson, D.R. A Systematic Survey of CMM Experience and Results. *Proceedings of ICSE 18*, **1996**.

23. Ho-Won J. Evaluating the internal consistency of ISO/IEC TR 15504 process capability measures. *Software Process Improvement and Practice (UK),* **2003**.

24. Ho-Won J. Evaluating the ordering of the SPICE capability levels: an empirical study. *Information and Software Technology (Netherlands),* **2005**.

25. Ho-Won J. Evaluating interrater agreement in SPICE-based assessments. *Comput. Stand. Interfaces (Netherlands),* **2003**.

26. Ho-Won J. and Hunter, R. Evaluating the SPICE rating scale with regard to the internal consistency of capability measures. *Software Process Improvement and Practice (UK),* **2003**.

27. Ho-Won J., Hunter, R., Goldenson, D.R. and El-Emam, K. Findings from phase 2 of the SPICE trials. *Software Process Improvement and Practice (UK),* **2001**.

28. Hong G., King, G., Ross, M. and Staples, G. SPICE: a suitable model for managing the software process in virtual software environments. *Proceedings 3rd Conference on Quality Engineering in Software Technology and VDE-ITG Workshop on `Testing Non-Functional Software-Requirements'*, **1999**.

29. Ibrahim, L. and Pyster, A. A single model for process improvement. *IT Professional (USA)*, **2004**.

30. Jarvinen, J. On comparing process assessment results: BOOTSTRAP and CMM. *Software Quality Management II. Managing Quality Systems* **1994;1**.

31. Jung, H.-W. and Goldenson, D. R. The internal consistency and precedence of key process areas in the capability maturity model for software. *Empirical Software Engineering*, **2008**.

32. Kamata, I., Higasi, S. and Moriyama, F. Introduction and evaluation of SPA method for telecommunication software development. *Sixteenth Annual Pacific Northwest Software Quality Conference Joint ASQ Software Division's Eighth International Conference on Software Quality*, **1998**.

33. Kim, K.T. A case study on architectural maturity evaluation: experience in the consumer electronics domain. *On the Move to Meaningful Internet Systems: OTM 2008 Workshops*.**2008**.

34. King, J. and Diaz, M. How CMM Impacts Quality, Productivity, Rework, and the Bottom Line. *Crosstalk*, **2002**.

35. Krasner, H. Using the SEI's systems engineering capability maturity model to evaluate an organization's capabilities. *Information Systems Analysis and Synthesis. Proceedings of the International Conference on Information Systems Analysis and Synthesis, ISAS'96*, **1996**.

36. Kuilboer, J.P. and Ashrafi, N. Software process and product improvement: An empirical assessment. *Information and Software Technology*, **2000**.

37. Kwak, Y. H. and Ibbs, C.W. Berkeley Project Management Process Maturity Model: Measuring the value of project management. *IEEE International Engineering Management Conference*, **2000**.

38. Lawlis, P., Flowe, R. and Thordahl, J. A Correlation Study of the CMM and Software Development Performance. *Crosstalk*, **1995**.

39. Leem, C. S., Kim, S. and Lee, H. J. Assessment methodology on maturity level of ISMS. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, **2005**.

40. Leem, C. S., Kim, B. W., Yu, E. J. and Paek, M. H. Information technology maturity stages and enterprise benchmarking: An empirical study. *Industrial Management and Data Systems*, **2008**.

41. Lee, Jo., Lee, Ji., Lee, S. and Choi, B. A CC-based security engineering process evaluation model. *Proceedings 27th Annual International Computer Software and Applications Conference. COMPAC 2003*, **2003**.

42. Leem, C. S. and Yoon, Y. K. A maturity model and an evaluation system of software customer satisfaction: the case of software companies in Korea. *Ind. Manage. + Data Syst. (UK)*, **2004**.

43. Lester, N.G., Wilkie, F.G., McFall, D. and Ware, M.P. Evaluating the internal consistency of the base questions in the express process appraisal. *33rd Euromicro Conference on Software Engineering and Advanced Applications*, **2007**.

44. Makarainen, M., Vierimaa, M. and Kinnula, A. Evaluating the process-experiences from three methods on the improvement of the DSP SW process. *SPI 98 Improvement in Practice: Reviewing Experience, Previewing Future Trends, The European Conference on Software Improvement*, **1998**.

45. McGarry, F. and Burke, S. and Decker, B. Measuring the impacts individual process maturity attributes have on software products. *International Software Metrics Symposium, Proceedings*, **1998**.

46. Mishra, A. and Mishra, D. Software quality assurance models in small and medium organisations: A comparison. *International Journal of Information Technology and Management*, **2006**.

47. Nelson, K.M., Armstrong, D., Buche, M. and Ghods, M. Evaluating the CMM level 3 KPA of Intergroup Coordination: a theory-based approach. *Inf. Technol. Manag. (Netherlands)*, **2000**.

48. O'Connell, E. and Saidian, H. Can you trust software capability evaluations? *Computer (USA)*, **2000**.

49. Pederiva, A. The COBIT Maturity Model in a vendor evaluation case. *Information Systems Control Journal*, **2003**.

50. Poort, E. R., Postema, H., Key, A. and De With, P. H. N. The influence of CMMI on establishing an architecting process. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*), **2007**.

51. Rye, K.-S. A study on Data Quality Management Maturity Model. *The 7th International Conference on Advanced Communication Technology, ICACT 2005*, **2005**.

52. Salviano, C.F. and Figueiredo, A.M.C.M. Unified basic concepts for process capability. *SEKE 2008, The 20th International Conference Proceedings on Software Engineering & Knowledge Engineering*, **2008**.

53. Shackleton, P., Fisher, J. and Dawson, L. Evolution of local government e-services: the applicability of e-business maturity models. *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*, **2004**.

54. Smith, R.C. Using a quality model framework to strengthen the requirements bridge. *Proceedings Third International Conference on Requirement Engineering*, **1998**.

55. Stienen, H., Engelmann, F. and Lebsanft, E. BOOTSTRAP: five years of assessment experience. *Eighth IEEE International Workshop on Software Technology and Engineering Practice incorporating Computer Aided Software Engineering* (Cat. No.97TB100145), **1997**.

56. Vaishnavi, V. K. and Fraser, M. D. Validation framework for a maturity measurement model for safety-critical software systems. *Proceedings of the Annual Southeast Conference*, **1998**.

57. Van der Velden, M.J., Vreke, J., van der Wal, B. and Symons, A. Experiences with the capability maturity model in a research environment. *Software Quality Journal (UK)*, **1996**.

58. Visconti, M. and Cook, C.R. Evolution of a maturity model-critical evaluation and lessons learned. *Software Quality Journal (USA)*, **1998**.

59. Wilkie, F.G., McFall, D. and McCaffery, R. An evaluation of CMMI process areas for small-to-medium-sized software development organizations. *Software Process Improvement and Practice*, **2005**.