TUM

TECHNISCHE UNIVERSITÄT MÜNCHEN
INSTITUT FÜR INFORMATIK

Software Quality Models in Practice

Stefan Wagner, Klaus Lochmann, Sebastian Winter, Andreas Goeb, Michael Kläs, Sabine Nunnenmacher

TUM-I129

Quamoco – The Benchmark for Software Quality

Stefan Wagner
Klaus Lochmann
Sebastian Winter
Andreas Goeb
Michael Klaes
Sabine Nunnenmacher

Software Quality Models in Practice
Survey Results

Table of Contents

Introduction

The study focused on practitioners from German-speaking countries complemented with participants from all over the world to examine the practice of software quality models. It analyzes three main aspects of quality modelling: the usage of quality models in practice, the importance of quality attributes and techniques, as well as the potentials of certain improvements of quality models. After the introduction, key findings are summarized, the study approach is described and finally the results for each question are detailed.

Motivation

The quality of software systems is a complex concept comprising various aspects; hence, it is hard to define quality precisely. Quality models are a common way to cope with this problem in a structured manner. A quality model is "a model with the objective to describe, assess and/or predict quality."[3] It can be provided implicitly by certain tools or can be provided in such forms as international standards, general guidelines, or checklists. Since quality models have been subject to research for many years, a manifold and heterogeneous number of quality models emerged, which contains, for example, domain-specific models or test models. Some quality models describe software quality as a whole, others describe just certain quality aspects. Though detailed information about the actual usage and resulting problems is missing, it seems that the usage of quality models differs widely in practice. Furthermore, it is not clear to what extent quality models contribute to the quality assurance processes in which they are applied. Which of these models are actually used in practice and for which purposes?

Objective

We present an analysis of the use of quality models in the software industry that is based on a broad Web survey. The main objectives of the study are to identify the main classes of quality models that are used at present, to identify the quality assurance techniques that are applied with these quality models, and to identify problems that are related to these models. The study covers different company sizes and several domains to allow a broad view on software product quality.

Context

The study was performed in the context of the research project Quamoco that is founded by the German Federal Ministry of Education and Research (BMBF). Partners in Quamoco are Capgemini sd&m, Fraunhofer IESE, itestra, SAP, Siemens, and the Technische Universität München. All partners contributed to the development of the questionnaire.

Related Studies

In 1993 Davis et al.[2] published the results of a survey on the practice of software quality. Most of the questions were related to methodical topics rather than the use of models for product quality. Their results show that 41% of the respondents claimed not to have any formal quality assurance method at all; from the remaining respondents the majority (55 %) used their own method for quality assurance rather than an external one.

Hall[4] analyzed the quality-related activities and attitudes to quality of developers and managers in 1995. In his study, 99% of developers use programming standards and 74% employ code reviews or inspections. It was also found that 62% of the respondents collect software metrics data.

More recently, Jung, Kim, and Chung[5] published a survey on ISO 9126. They investigated whether the categorization of quality attributes is correct and reliable. This resulted in an improved categorization to provide guidance for revising the standard. This new categorization includes, for example, to separate security from functionality, which is the current proposal of the new ISO 25010.

A survey on software testing practices in Australia was conducted in 2004 by Ng et al.[7] among 65 different organizations. Their results show which testing methodologies and techniques as well as tools are used within these organizations. The survey also includes information about metrics, standards, and trainings related to software testing. The survey identifies current practices and therefore aims in a similar direction as our survey. Relevant results include that only 55.3% of the organizations in the survey agreed that metrics help to improve software quality. Furthermore, manual static analysis, that is, formal and informal reviews, is still more commonly used than automated, tool-based analyses.

Berntsson Svensson, Gorschek, and Regnell[1] interviewed 10 product managers and project leaders from 5 Swedish companies on their quality requirements practices. They found usability and performance being requirements considered most important. They also identified a major challenge with quality requirements to make them well specified and quantified so that they are testable.

Key Findings

This part of the report summarizes the key findings of the survey. It presents the interpretations that were drawn from the survey data. The key findings are highlighted by statements and contain insights on the type and usage areas of quality models, importance of quality attributes and quality assurance activities, as well as suggestions for improvements.

Usage of quality models

A wide variety of quality models is used. We found that almost all participants use some kind of quality model. A wide variety of these models is employed in practice. The spectrum ranges from laws and international standards to domain- and company-specific quality models. Examples are the general software quality standard ISO 9126, the medicine standard ISO 13485, and testing standards.

Company-specific quality models are widely used. Company-specific quality models are applied by more than 70% of the respondents. This is by far the most used type of quality model in our survey. On the contrary, only 28% of the respondents apply the standard ISO 9126. In addition to company-specific quality models, quality gates and defect classifications are among quality models that are employed often. Thus, both kinds should be incorporated in future developments of quality models as well.

Quality models typically are adapted. In total, almost 87% of the respondents adapt their used quality models to their needs. This can have two interpretations: (1) the existing quality models are not good enough to be used as is or (2) adaptation is an unavoidable, necessary task in using quality models. In detailed interviews conducted earlier by us, operationalizing quality models was one of the major problems encountered in practice. In the existing quality models, it was perceived that concrete measures that are needed for the operationalization are missing. The cause might be that the operationalization is too specific for a particular context and hence makes an adaptation indispensible.

ISO standards are not well accepted. The ISO 9126 standard is applied by 35 (28%) of the respondents. Of these, 25 employ a company-specific quality model in addition. Hence, the standard itself seems not to be sufficient for the needs of most companies. This result supports the common critique about the current ISO standard. Following from that and the above findings that company-specific quality models are predominant and the adaptation of quality models is common, we conclude that the existing standards are not sufficient to fullfil the needs of practice.

Quality model users are moderately satisfied with their models. In general, the respondents' satisfaction with the applied quality models is fair. On average, it lies slightly above a medium satisfaction. Therefore, we conclude that quality models are a good basis for quality management in practice but there is still room for improvements.

Aggregation is done often but not predominantly. With 47% of the valid responses, aggregation is a common activity in quality assessments. However, as more than half of the respondents do not aggregate, it is not predominant. Two interpretations are possible: (1) an aggregated quality statement is not required in practice or (2) common quality assessment methods and quality models do not sufficiently support aggregation. Hence, further analyses of the pros and cons of aggregation are necessary.

Quality Attributes

All quality attributes are important. Despite the overall high range and context-dependency of importance ranking of quality attributes, the average importance ranges only in the upper half. Most of the quality attributes belong to one of two importance ranks. Even the attribute considered least important was on average still ranked above medium importance. We therefore conclude that a widely applicable quality model needs to include the whole set of quality attribues.

The importance of quality attributes is context-dependent. There is, however, a high variance in the importance ratings of quality attributes. That leads to the conclusion that, in general, the importance of quality attributes differs strongly and is context dependent. This is expected because the environments and purposes of software systems differ strongly. For example, for a software system controlling the brakes of an aircraft, safety is essential whereas for a Web shop safety is not considered.

Functional suitability is most and portability least important. In terms of the content of the quality models, it appears that functional suitability is considered most important, whereas portability is considered least important. The ranking follows largely what a quality expert would expect as functionality in general is considered very important, followed by attributes as performance and reliablity. Portability and installability range at the lower end. By averaging the response, we can support these general tendencies.

Functional suitability is analyzed by the largest variety of techniques. Functional suitability was on average considered to be the most important quality attribute. Consequently, it also has the largest variety in techniques that are applied for it. Besides testing, customer feedback, formal reviews, and informal reviews are used for it by many respondents.

Quality assurance techniques

Quality models are highly used in most activities. Quality models are used for a variety of activities in the development and maintenance processes. The most usage, however, is in testing, architecture and design, the development of guidelines, and formal reviews. Additionally, testing and formal reviews together with customer feedback are considered most important for evaluating quality requirements. Especially testing was rated as very important for most quality attributes. Due to the fact that quality models are applied to support the important evaluation techniques, these techniques need to be particularly supported in future developments of quality models.

Testing is most important. Dynamic testing of a software system is still considered the most important technique for evaluating quality requirements. Customer feedback and formal reviews follow closely. Similiarly to the importance of quality attribues, the variance is large. This suggests that there are important differences between companies.

Testing, code analysis, and measurement are often used daily. The quality assurance techniques applied on the most frequent basis in practice are testing, code analysis, and measurement. Each of these techniques is used daily by more than 20% of the respondents, testing even by more than 40%. A possible explanation is that they have the highest potentials for automations. This finding implies that the support by a quality model needs to allow a daily usage.

Customer feedback and formal reviews are employed mostly at milestones. Contrary to the daily usage of testing, code analysis, and measurement, the execution of customer feedback and formal reviews as quality assurance techniques is mostly restricted to specific milestones. Similarly, an explanation could be that these two techniques require a high degree of manual effort. Furthermore, it is probably not easily possible to get customers for such feedback. Hence, the quality model support for these activities is sufficient at milestones.

Testing is used the most for all quality attributes except maintainability and portability. For evaluating quality attributes, testing was considered most important. It is also used for the most quality attributes. No other technique was mentioned that often and for such a variety of quality attributes. However, it ranks lower for maintainability, for which formal and informal reviews are mentioned more often. Also for portability, the dominance is not clear as formal reviews are mentioned almost as often as testing.

Improvements

All improvements have high potential. On the whole, quality models are regarded as having a high potential for improvements. The improvements that obtain the largest numbers of high ratings are the following fields: defining evaluation criteria, quantifying quality, and transparent definition of quality. The support of these tasks should have a high priority in future improvements of quality models.

Functional suitability should be improved. On average, functional suitability was responded most often as a quality attribute that should be improved first. Over 80% of the respondents mention this quality attribute and it is by far the quality attribute most often placed on the first rank. The quality attributes that also exhibit a high potential for improvements are reliability, performance, maintainability, operability, security, and safety. Consequently, future quality models should consider these quality attributes. On the other end of the spectrum, installability, compatibility, and portability seem not to be in need of an urgent improvement.

Methodology

This part describes the methodology that was applied in the study. The objectives of the study are defined, the design and the population of the study are presented, and an overview of the implementation and the execution of the study is provided.

Study definition

The aim is to identify the current state of software quality models in practice. This study examines the practitioners' understanding of software product quality, the methods they apply, and their opinions regarding strengths and weaknesses of existing quality models.

The objective of the study is to analyze **product quality models** for the purpose of **characterization** with respect to **their usage (where and for what) and possible improvement potentials** from the viewpoint of **quality managers and quality model users** in the context of **the software developing industry**.

The main research questions are:

1. What quality models (including standards and laws) are used in which contexts?

2. What quality attributes are important?

3. How is quality assurance influenced by quality models?

4. What are current problems with quality models and which potentials for improvement exist?

Study design

The study is divided into two phases. Phase 1 consisted of interviews in a face-to-face mode. The design of the questionnaire was identified and refined in a series of workshops. The final version of the questionnaire contained 15 open questions and 12 closed questions. Phase 1, as a whole, serves as a pre-study for Phase 2. Detailed information about Phase 1 was published separately.[8]

Phase 2 comprises a research survey using an Internet-based questionnaire. We used the data of the interviews conducted in Phase 1 to refine the design of the Internet-based questionnaire. Additionally, problems that occurred during the interviews originating from ambiguous wording were resolved. The questionnaire was initially tested internally with employees of the Quamoco industrial partners to detect its potential weaknesses and to estimate the time required for completing it. Afterwards the questionnaire was further refined to eliminate the problems detected in the internal tests.

The final version of the Internet-based questionnaire consists of 23 questions that are divided into the four groups:

1. The role of quality models in the company

2. Quality requirements and evaluation of quality

3. Improvement of quality models

4. General information about the company

The last question of each group was an optional open question, which allowed the respondents to state additional comments of any kind. Based on the experience of the pre-test, we estimated at most 25 minutes for completing the questionnaire. For the final survey, the questionnaire was implemented in LimeSurvey, a free Web survey tool.

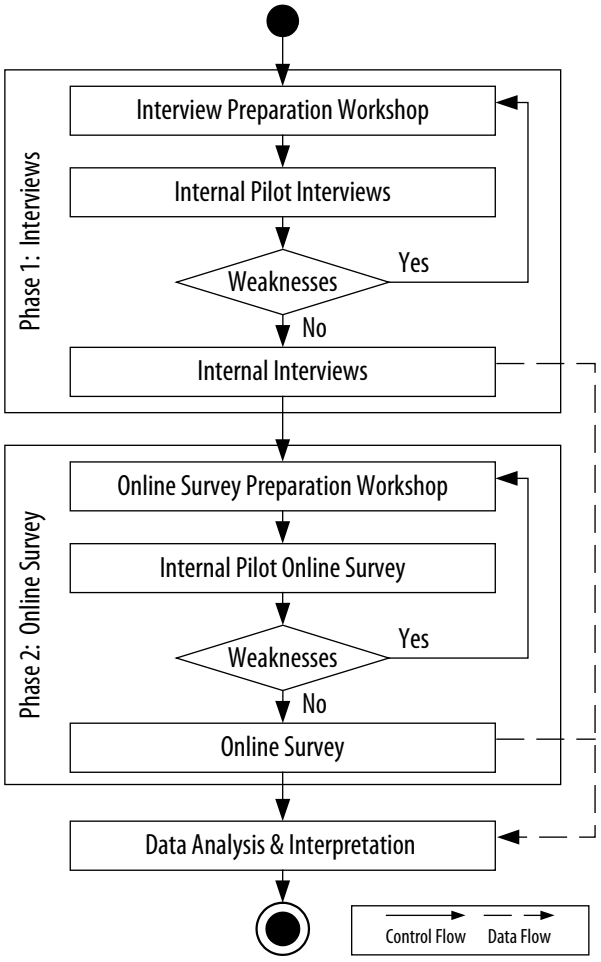

Figure 1 Sequence diagram of the study design.

Study population

The population of this study consists of quality managers and quality model users employed at software development organizations that are distributed over the whole world. The recipients of the Internet-based questionnaire were selected using convenience sampling. Based on data from the project partners, a list of recipients was compiled. This selection resulted in a sample of 515 persons. We contacted the persons from the sample, out of which 125 completed the questionnaire. This corresponds to a response rate of 24.3%. Although we used personalized contacts only and sent one follow-up e-mail to those who had not answered, we did not establish pre-contacts.

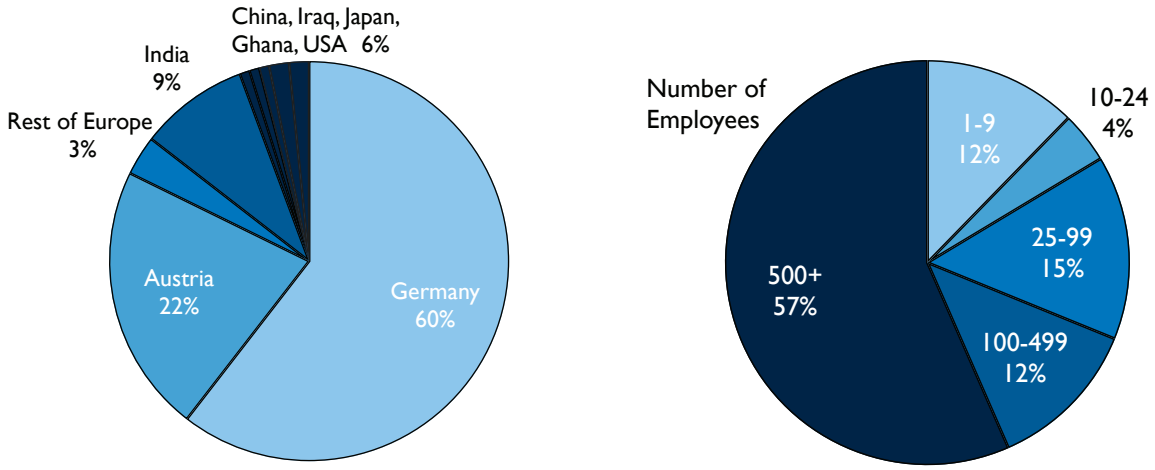

Figure 2 Countries and employees

The respondents come from 12 countries. The majority of the respondents work in Europe, mainly Germany and Austria. Hence, we have an emphasis on German-speaking countries. This is complemented by 18% participants from other European countries, Asia, Africa, and North America. Forty-four percent of the respondents work as project managers. Twenty-eight percent of the respondents are normal employees or line managers. The experience of the respondents is important to ensure that the respondents are able to answer the questionnaire meaningfully. Twenty-nine percent of the respondents have between 11 and 15 years of professional experience in the area of software engineering; only 15% of the respondents have less than five years of professional experience. The majorirty of the respondents see development as the primary task of their department. More than a third consider their department mainly in quality assurance. The participating organizations cover all sizes in terms of the number of employees with an emphasis on larger companies.

The study covers organizations that develop business information systems, software for embedded systems, development tools, and platforms. This software is used in all major domains ranging from telecommunication to finance.

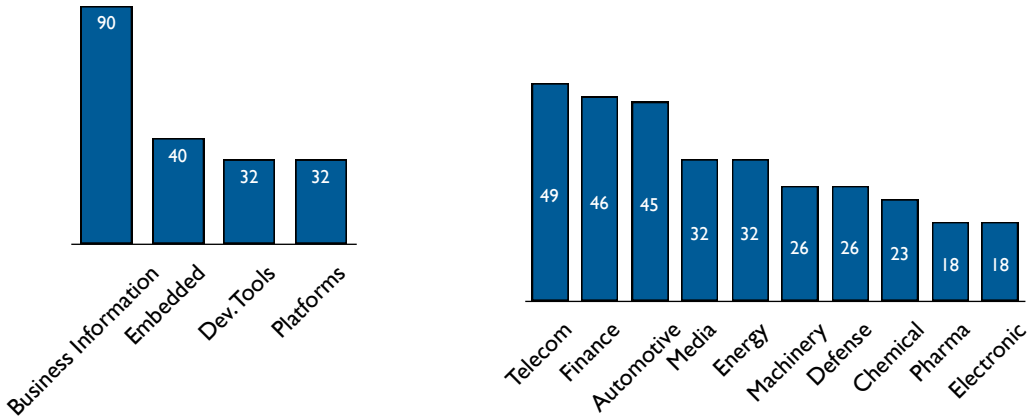

Figure 3 a) Types of software b) Software usage domains

Disclaimer

Limitations are a common threat in empirical field research. To reach validity of the constructs used in the study, we only interpreted what the question directly asked for and set them into relation to the information that we have learned from conducting the interviews of Phase 1. Moreover, we did a pre-test of the questionnaire, which gave additional insights for improving the questions' design. All completed questionnaires were carfully inspected and checked for consistency where appropriate, because a web survey is not administered and hence no additional help can be given to the participants. Respondents may have tried to manipulate the results because they have some interest in the outcome. For example, experts for a specific quality assurance technique may have a stake in having this technique rated as very important in this study.

We assured the recipients that the responses are kept separately from the authentication tokens, i.e., tokens and survey responses cannot be matched. This supports that respondents report all quality-related problems located in their companies. The survey was conducted as a broad study, i.e., many different companies and individuals were involved. Hence, the results should be generalizable to other companies and countries because in total the 125 participants came from 12 countries. Most participants have a long experience in software development, but experts with less experience were also among the respondents. Likewise, we cover different company sizes, different domains, and different types of systems.

Detailed Results

This part details the survey results. They are structured in the same way as the questionnaire is organized. Each question is first described generally and then descriptive statistics are given. This includes the use and adaptation of quality models and the level to which the respondents' needs are satisfied, the importance of different quality attributes, and the usage of quality models for specific activities. Furthermore, improvement potentials are discussed.

Company-specific quality models are widely used

As the term "quality model" is used heterogeneously in the research community as well as in practice[3,6,8], it can mean a variety of models and standards to practitioners. The aim of this question is to capture all these models and to determine which of them are often used in software development. Therefore, we employ the following question:

Question 1 Which quality models/standards do you use for assuring the quality of your products?

The participants were asked to answer one or more of the options "ISO 9126", "ISO 25000", "domain-specific", "company-specific", "laws", "quality gates", "defect classification", "reliability growth models", "none", or "other". In case of domain-specific or other models, the participants could give the name of the model as free text. The chart shows the results as the number of answers per option.

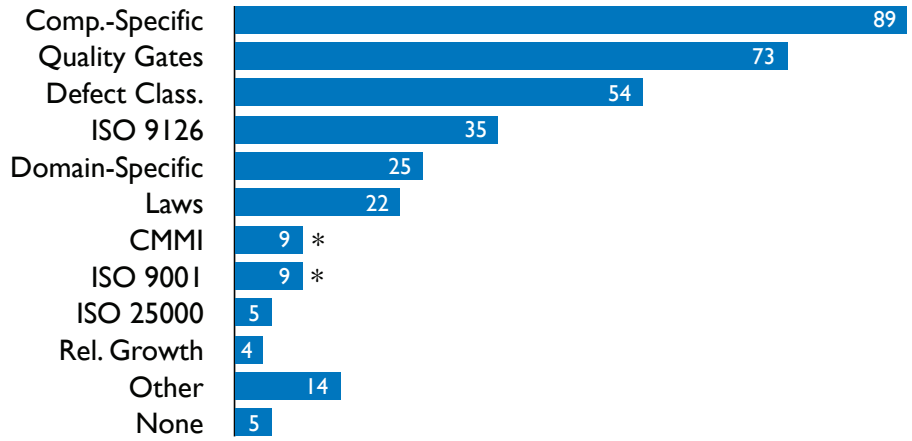

Figure 4 Quality models used in practice

The predominant kind of quality models that are used in practice are company-specific models. Almost three quarters of the respondents use this type of quality model., i.e., company-specific models are used widely in practice. Well-established practice are also quality gates and defect classifications, with about half of the participants employing them. As these terms are rather vague, the actually used models can, however, differ to a high degree. In the first study phase, we found that defect classifications are often only a prioritization of defects.

The data also suggests that the standard ISO models are not highly accepted. ISO 9126 is adopted by less then a third of the participants and ISO 25000 only at 4%. For the latter, a reason might be the recent and yet incomplete publication of the standard. If ISO 9126 is used, 25 out of 35 respondents additionally apply a company-specific model. This suggests that the standard model itself cannot fulfill the needs of the software quality experts.

Domain-specific quality models, and laws were also mentioned frequently. The mentioned domain-specific models originated from the area of pharmaceutics and medicine (ISO 13485, IEC 60601, IEC 62304, MDD, CFR) with 11 occurrences in total, testing (4), safety (4), security (3), public or military (3), and accessibility (2). Five respondents answered that they use the new ISO 25000 and only four employ reliability growth models. Five respondents (4%) answered that they do not use quality models.

For this question a sanity check was performed, i.e., "none" was not supposed to appear together with any other field. The check was successful for all 125 responses. In addition, the answers for "other" led us to corrections in four cases where it was obvious that the mentioned models belong to one of the predefined categories. The remaining models in "other" were mainly "CMMI" and "ISO 9001" with nine occurrences each. This is why we introduced them as separate categories and removed them from the "other" category.

Quality models typically are adapted

The second question relates to the finding for the first question that standard quality models are commonly used in conjunction with company-specific models. We explicitly ask about the adaptation of the quality models that are used:

Question 2 Do you adapt the quality models that you use for your products?

The participants had three options for answering this question. They do not adapt the used quality models ("no"), they adapt them "for classes of products", or "for each product" separately. Furthermore, a "don't know" answer was possible. Fig. 5 shows the relative distribution of the answers including the "don't know" answers.

Quality models are commonly adapted. Ignoring the "don't know" answers, 13.2% do not adapt their quality models. The difference between the adaptation for classes of products and single products is only 4 percentage points; thus, both kinds of adaptation are likely to happen in practice. We conclude from this that adaptation is a necessary task in using quality models independently of whether it is employed for a single product or for classes of products. It might even be an indicator that the existing standards are not sufficient for practical needs.

In detail, of the 125 responses, 11 ckecked "don't know". Therefore, only 12.0% do not adapt their quality models. The quality models of 79.2% of the participants are adapted, of which 41.6% are adapted for classes of products and 37.6% for each product.

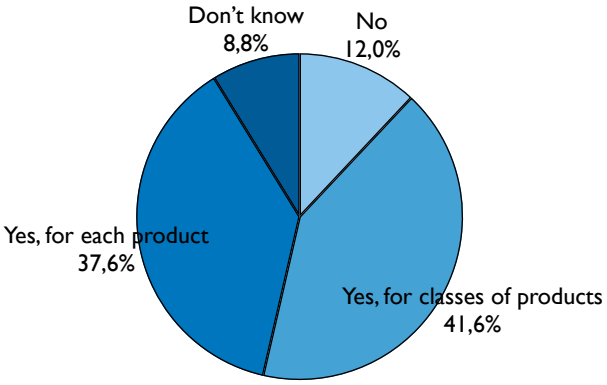

Figure 5 Adaptation of quality models

Satisfaction with quality models is moderate

After establishing what kinds of quality models are used and whether they are applied as-is or adapted, we ask how satisfied the respondents are with their employed quality models at present. Since we are interested in the satisfaction of the actually used quality models instead of a general opinion, we asked the following question:

Question 3 How satisfied are you with the quality models you use for your products?

The respondents could answer on a 10-point ordinal scale from "very satisfied" to "very dissatisfied". Alternatively, a "don't know" answer was possible. Fig. 6 shows the distribution of the answers on this scale including the absolute number of answers.

The average satisfaction with the used quality models tends to an undecided opinion with a slight shift in the direction of satisfaction. The quality model users seem not be completely unsatisfied with their models. However, a clear satisfaction is not observable as well. This suggests that the concept of quality models seems to deliver a basic level of satisfaction but it has still room for improvements.

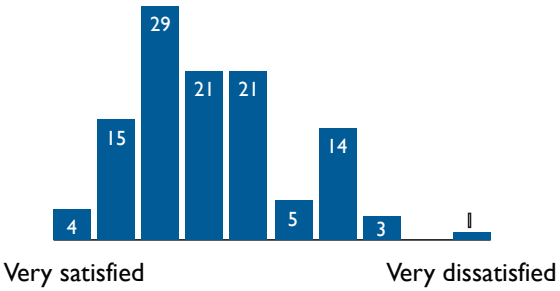

Figure 6 Satisfaction with quality models

Eight of the values are missing; thus, 117 values remained for further analysis. The answers are coded by means of a variable with a scale from 1 to 10, where 1 codes "very satisfied" and 10 "very dissatisfied". The whole range of values from 1 to 10 was used by the respondents. The median is 4, the mean is 4.21 with a variance of 3.480.

Importance of quality attributes differs

The fourth question examines what quality aspects are important. To this end we asked the respondents to rate quality attributes that we derived from ISO 25000 and the experiences of Phase 1. For all quality attributes, a short definition was available. The question is as follows:

Question 4 How important are the following quality attributes for your products?

For each quality attribute a 10-point scale from "very important" to "very unimportant" is given. Alternatively, for each quality attribute, a "don't know" answer was possible. Fig. 7 visualizes the central tendency and range for each quality attribute. The boxes represent the values from the lower to the upper quartile and the horizontal line marks the median. The upper and the lower whisker represent the minimum and the maximum, the circles and asterisks visualise outliers.

The analysis shows that the individual ranking of quality attributes varies and that the importance of quality attributes does not differ strongly. The former can be seen in the ranges of the quality attributes that use the whole spectrum of possible answers for all of the quality attributes. The latter is derived from the small distribution of medians, of which most are 7 or 8. As medians, the ranking "very important" (10) and "medium important" (6), only appear once. Hence, the distinction of importance between the attributes is not clear. However, functional suitability tends to be the most important quality attribute whereas portability seems to be the least important one. We conclude that all quality attributes are important. Depending on the context each attribute can be of most importance. Nevertheless, on most participants minds foremost the functional suitability of their software together with its reliability and performance. Portability and installability is only of high relevance in specific contexts, probably standard software.

The "don't know" answers are counted as missing. The numbers of missing values range from 3 to 9; thus, 116 to 122 answers are available for each quality attribute.

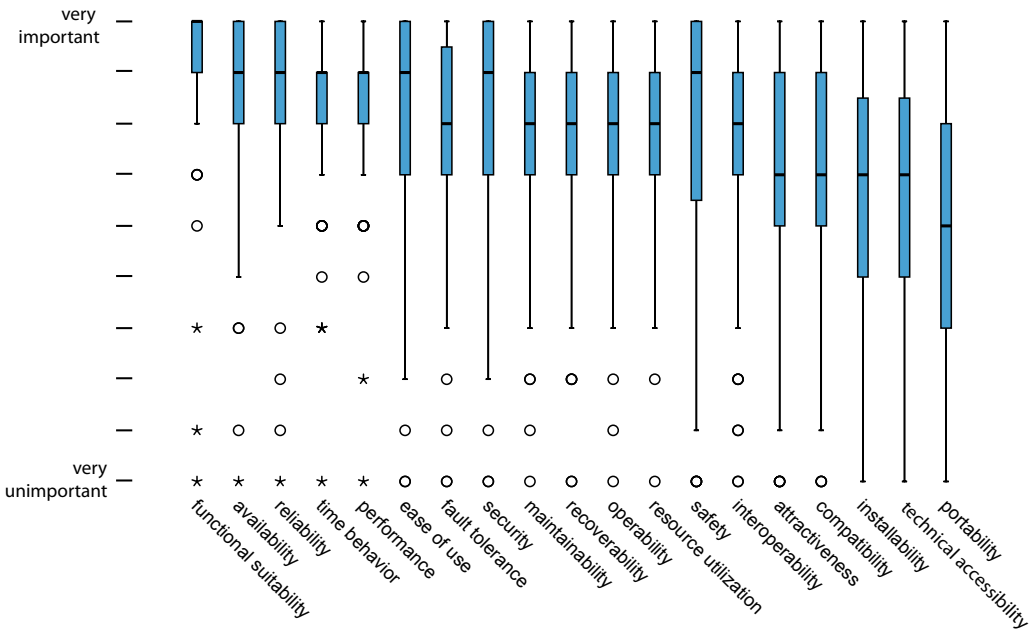

Figure 7 Importance of quality attributes

Models are used in all quality assurance activities

Existing quality models are used in different contexts supporting a variety of activities in the development and maintenance of software. Therefore, we consider in what development activities quality models are used. We asked the following question:

Question 5 In which development activities do you use quality models?

These activities were asked for: Requirements engineering, architecture and design, development of coding guidelines, informal reviews, formal reviews, tool-based code analysis, data collection and measurement, testing, and evaluation of customer feedback.

For each of these activities the respondents could give one of the following answers: The activity is not part of the development process ("activity does not exist"), quality models are used in this activity ("yes"), quality models are only partially applied ("partly"), or quality models are not used ("no"). Otherwise, the respondent could state a "don't know" answer. In addition, question 6, a free text question, gives the respondents the opportunity to state, for example, the motivation why they use a quality model in a development activity. Fig. 8 depicts the relative number of "yes" and "partly" answers for each activity.

Quality models are frequently used in all development activities. In particular, the testing activity is supported by quality models and, additionally, quality models are used in a thorough manner. This is reflected by the low number of participants who answered "partly" (7%). This is in contrast to architecture and design in which quality models often are used partially (37%). The slight dominance of testing may be explained by the important role testing plays in quality assurance in general (see next question).

For each activity, the respondents that answered "does not exist" or "don't know" were excluded. The three activities in which quality models are used most are testing (73% of the respondents answered "yes"), formal reviews (57%), and coding guidelines (56%). The three activities in which quality models are used least are informal reviews (22% of the respondents answered "no"), customer feedback (16%) and data collection (14%).

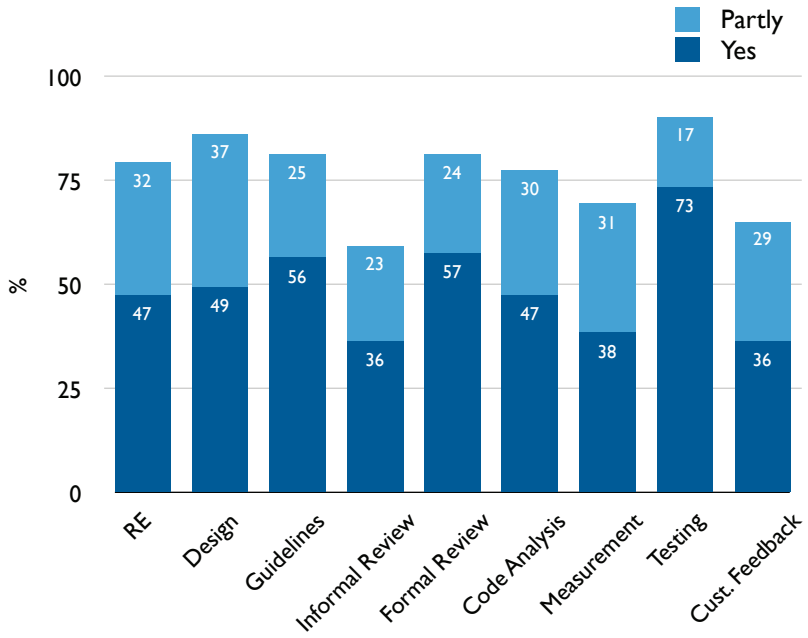

Figure 8 Quality assurance techniques that use quality models.

Testing is most important

The next two questions focus on two related aspects of quality assurance: What techniques are crucial in order to evaluate quality requirements and how often/regularly these techniques are used. Question 7 has the objective to determine which techniques are considered to be important.

Question 7 How important do you rate the following techniques for evaluating quality requirements?

The following techniques were considered: informal reviews, formal reviews, tool-based analyses, data collection and measurement, testing, and evaluation of customer feedback. Each of these techniques could be rated from "very important" to "very unimportant". If the respondents could not give an answer, they could check the "don't know" option. Fig. 9 shows the distributions of the importance ratings. For each of the techniques the median is marked by a black line. The whiskers give the quartiles and circles and asterisks denote outliers.

From the data, we conclude that testing is considered the most important technique for evaluating quality requirements. The techniques that were rated as most important after testing are formal reviews and customer feedback. The importance of a technique does not necessarily mean that quality models are used accompanying. This is shown by the technique customer feedback: The use of quality models for this activity is minor (36%), although the activity is considered important.

Looking at the distribution of the medians, similiar to the quality attributes, the importance does not differ strongly. The range of he medians contains only 3 of 10 scale points. Hence, overall all mentioned techniques are considered important.

The "don't know" responses were not further considered. However, the number of these responses was very low for most techniques (0-2), except for the technique "measurement" (11).

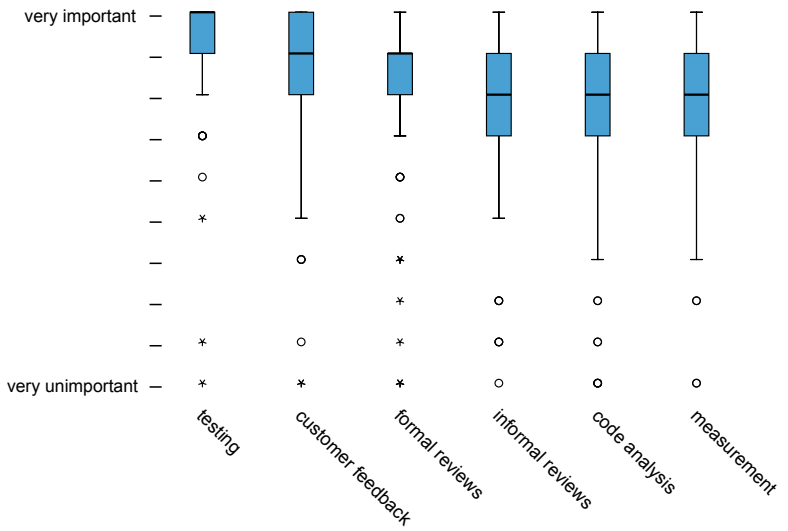

Figure 9 Importance of quality assurance techniques

Testing daily, customer feedback and formal reviews at milestones

Beyond the general importance, an important characterstic of a quality assurance technique is how often is used in the development and maintance of software systems. Therefore, the following question was asked:

Question 8 How often do you use the following quality evaluation techniques?

The same techniques as in the last question were used. For each of these techniques the following answers could be given: "daily", "weekly", "monthly", "at specic milestones", and "never". It was also possible to answer "don't know". For this question, multiple answers were allowed. That means it was possible to answer that an activity is performed once a month and also at certain milestones. Fig. 10 shows the relative results with colour-codes for the time intervals for each technique.

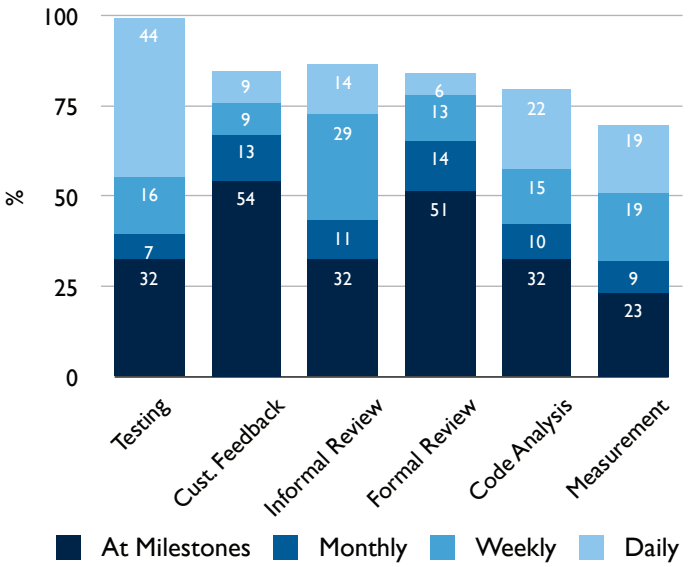

Figure 10 Usage frequency of quality assurance techniques

Techniques that can be carried out fully automatically, such as testing, are in favour when it comes to evaluation on a daily basis. Other techniques that are more time consuming yet important, such as formal reviews, are used predominantly at milestones. Monthly is in general not a very common interval for quality assurance. Likewise is "weekly" only popular for informal reviews.

The two techniques that obtain the most answers of "never" were formal review (27%) and code analysis (26%). The two techniques that are used most often at milestones are customer feedback (23%) and formal review (23%). The two techniques that are used most often in a monthly term are formal reviews (22%) and customer feedback (20%). The two techniques that are used most often on a weekly basis are informal review (29%) and data measurement (19%). The two techniques that are used most often on a daily basis are testing (41%) and code analysis (19%).

For this question a sanity check was performed to make sure that "never" as well as "don't know" did not occur together with any other option. In two cases this test failed. The corresponding inconsistent answers were considered missing as well as the "don't know" answers. These answers were not taken into consideration for the further analysis.

Functional suitability evaluated by most, testing for almost all quality attributes

Not all quality assurance techniques are suited to analyze all aspects of quality equally well. The following question focuses on the techniques used to evaluate specific quality attributes:

Question 9 How do you evaluate the specific quality attributes?

The answers are designed as a matrix with the specific quality attributes on the vertical and the evaluation alternatives on the horizontal level. For this question, multiple answers are possible. Besides the usage of a technique for a quality attribute, the participants could also state "don't evaluate" and "don't know". The results of this question are shown in fig. 11. The larger the bubble, the more often the corresponding combination of quality attribute and technique was checked.

For this question two main results are relevant: Functional suitability is the quality attribute evaluated most intensively and testing is a very important evaluation technique for all quality attributes except portability and maintainability. In addition, the results show that costumer feedback is relevant for functional suitability, reliability, performance and operability. Furthermore, measurement as well as tool-based code analysis are rarely used for any quality attribute with the exception of reliability and performance.

The combination performance and testing was mentioned most often. Of the respondents, 79% use testing to analyze the performance of their software systems. The second and third highest frequencies have testing for functional suitability (74%) and testing for reliability (73%). The lowest frequency has the combination tool-based code analysis and installability with 2%. Portability analyzed by tool-based code analysis (6%) and operability and tool-based code analysis (6%) are mentioned similarly infrequently. In general tool-based code analysis (193 answers in total, 14%) and data collection & measurement (187, 14%) are mentioned least frequently. Informal review (317, 23%), customer feedback (334, 24%), and formal review (355, 26%) are mentioned similarly overall. Testing is mentioned by far the most often with 784 (57%).

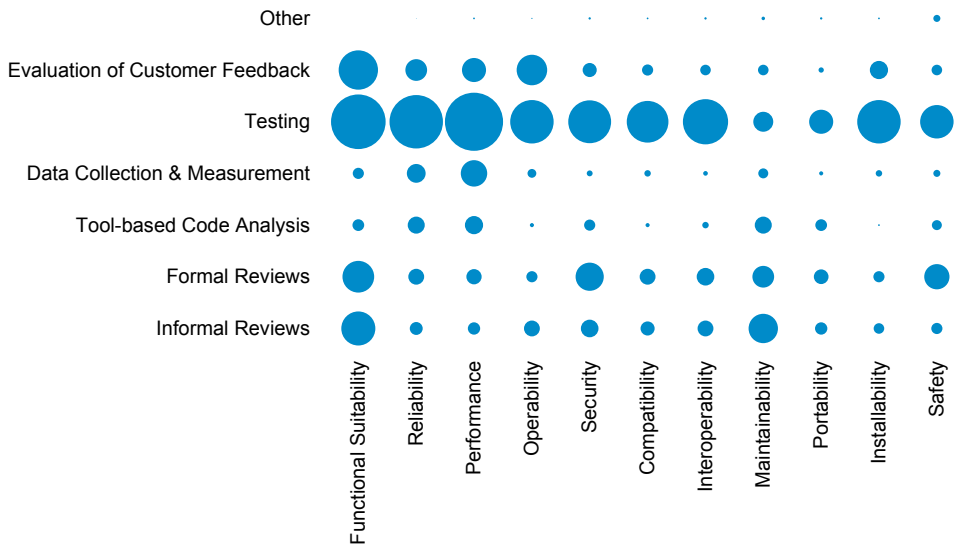

Figure 11 Quality assurance techniques per quality attribute

Some companies have a system-wide quality statement

It is a large difference whether isolated quality analyses are performed or whether these are aggregated into an overall quality statement about a software system. Therefore, in this question we asked, if a general quality statement based on the quality evaluation results is made by the respondent:

Question 10 Do you aggregate the quality evaluation results into an overall quality statement using a quality model?

A description of the "overall quality statement" is provided in the questionnaire to assure that all respondents have a common understanding of this term. The answer categories for this question that are possible are "yes", "no" and "don't know". For the analysis, all mentions of "don't know" are considered as missing values.

According to the findings, there is no tendency in favour of or against the aggregation into an overall quality statement. For the interpretation of this question we take question 11 into account that allowed additional textual comments. Some respondents used the opportunity to give a comment related to question 10. One respondent who aggregates the quality evaluation results says that the aggregation is made for the understanding of the management. Other respondents note that the aggregation is used for visualization. Opinions against the aggregation are that the data for the aggregation is not available. Suitability for daily use is another reason against aggregation as commented by a respondent. We cannot clarify completely which reasons are responsible for or against an aggregation. For further analyses it may be expedient to ask explicitly for that.

Ten respondents are not able to answer this question within the two categories "yes" or "no". For this reason, the number of valid values is reduced to 115. The frequencies reveal that there is no major difference between "yes" and "no".

Many improvements with high potential

For further developments of quality models, it is paramount to understand the potentails of improvements. Hence, in this question several improvements are examined for their potential:

Question 12 Which improvement potentials do you see in the following fields?

On a 10-point scale from "very low potential" to "very high potential" the respondents had to specify how they estimate the prospects of improvement suggestions. The respondents who did not require a certain improvement could choose a separate category. The fields for improvment given were "defining practically usable quality models" , "adapting quality models to specific application contexts" , "integrating quality models into life-cycle processes", "transparent definition of quality" , "quantifying quality" , "defining evaluation criteria", "economical analyses of quality improvements", "standardization of quality models" , "aggregation to quality statements" , and "product quality certification" . Fig. 12 shows the distributions of the answers for each improvement field.

The findings show that for all fields a high improvement potential exists. More advanced topics such as qualitay statements, standardization, and certification may be rated with lower potential due to issues besides the use of quality models. For example, standardization has the likely problem that it takes long to finish the standard and that many different interests have to be incorporated. Looking at the fields with the most ratings of the two highest ranks, defining evaluation criteria, quantifying quality, and transparent definition of quality are the top 3. We therefore suggest to improve these fields first.

For most of the improvement potentials a few outliers exist. For this descriptive analyses they are not excluded. For further hypotheses it may be reasonable to exclude them. The boxplots demonstrates that three mentioned fields are not evaluated with a high potential, all of the others show almost a high potential. In these three fields the dispersion is also the highest. In the analyses "don't know" answers were not considered.

Functional suitability should be improved

Another aspect of improvement is the choice of quality attributes to be improved. Therefore, in the following question, we asked which quality attributes should be improved first, when developing a quality improvement strategy:

Question 13 If you were in a position to decide on a quality improvement strategy, which three attributes would you address first?

Respondents could choose three quality attributes from a list with pre-defined quality attributes and rank them. As answers, a reduced set of quality attributes was given that comprises "functional suitability", "reliability", "performance", "operability", "security", "compatibility", "interoperability", "maintainability", "portability", "installability", and "safety". Fig. 13 shows the relative number of answers of attributes on the first, second, and third rank.

As a result, we conclude that functional suitability, reliability, and performance are the three quality attributes rank most often. They should be part of most quality improvement strategies. On the other side, installability was not chosen; portability and compatibility were rarely chosen. They do not seem to be in need of urgent improvement.

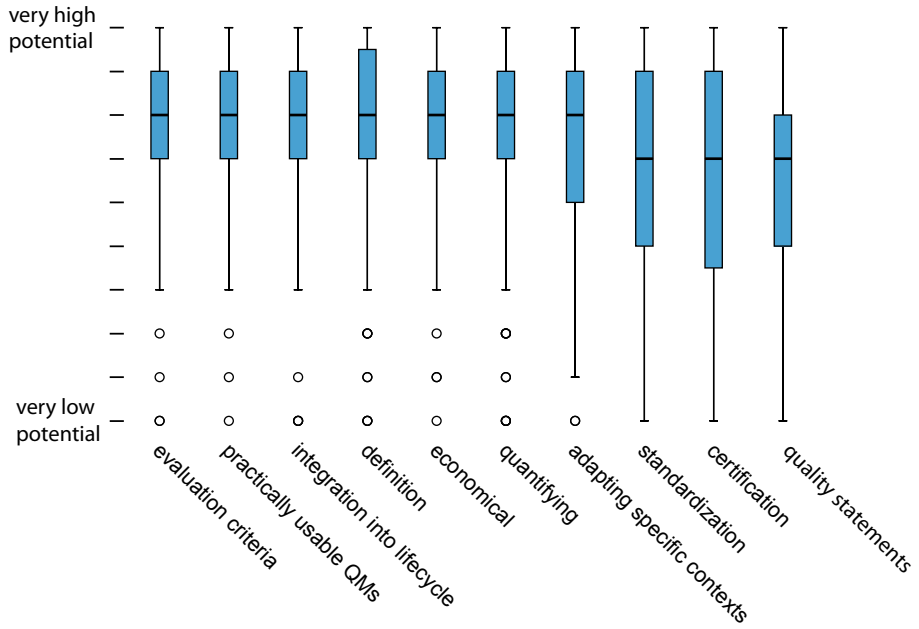

Figure 12 Improvement potentials for quality models

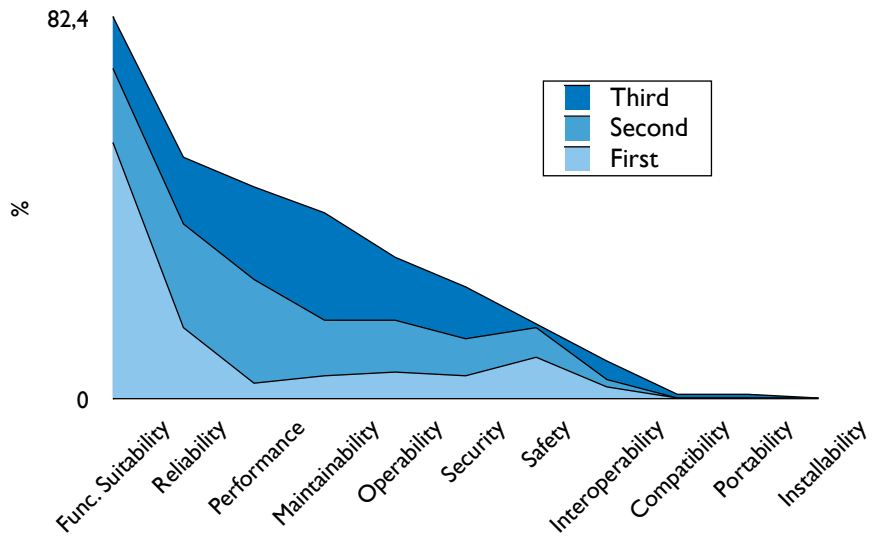

Figure 13 Quality attributes to be improved

Bibliography

1 R. Berntsson Svensson, T. Gorschek, B. Regnell
Quality Requirements in Practice: An Interview Study in Requirements Engineering for Embedded Systems
In: Proceedings of REFSQ 2009, LNCS 5512, Springer-Verlag, 2009

2 C. J. Davis, A. C. Gillies, P. Smith, and J. B. Thompson
Current practice in software quality and the impact of certification schemes
Software Quality Journal, 2(3):145–161, 1993

3 F. Deissenboeck, E. Juergens, K. Lochmann, S. Wagner
Software Quality Models: Purposes, Usage Scenarios and Requirements
In: Proceedings of the Seventh Workshop on Software Quality, IEEE Computer Society, 2009

4 T. Hall
What do developers really think about software quality?
In: Software Quality Management III, vol. 1: Quality Management, Witpress, 1995

5 H.-W. Jung, S.-G. Kim, and C.-S. Chung
Measuring software product quality: A survey of ISO/IEC 9126
IEEE Software, 21(5):88–92, 2004.

6 M. Klaes, J. Heidrich, J. Muench, A. Trendowicz
CQML Scheme: A Classication Scheme for Comprehensive Quality Model Landscapes
In: Proceedings of the 35th Euromicro SEAA Conference, 2009

7 S. Ng, T. Murnane, K. Reed, D. Grant, and T. Chen
A preliminary survey on software testing practices in Australia
In: Proceedings of the 2004 Australian Software Engineering Conference, IEEE Computer Society, 2004

8 S. Wagner, K. Lochmann, S. Winter, A. Goeb, M. Klaes
Quality Models in Practice: A Preliminary Analysis
In: Proceedings of the 3rd International Symposium on Empirical Software Engineering and Measurement, IEEE Computer Society, 2009

Quamoco

The Benchmark for Software Quality

Research Project "Quamoco"
http://www.quamoco.de/

Capgemini sd&m AG
Fraunhofer Institut für Experimentelles Software Engineering
itestra GmbH
SAP AG
Siemens AG
Technische Universität München

Contact:

Markus Großmann
Capgemini sd&m AG
Löffelstraße 46, 70597 Stuttgart
Tel.: +49 711 7 83 24-135
Fax: +49 711 7 83 24-150
Email: markus.grossmann@sdm.de

Supported by the German Federal Ministry of Education and Research (01 IS 08023)

Bundesministerium
für Bildung
und Forschung