

Do We Really Know How to Measure Software Quality?

Vineta Arnicane^[0000-0003-3942-9229], Juris Borzovs^[0000-0001-7009-6384],
Anete Nesaule-Erina^[0000-0001-7244-2148]

Faculty of Computing, University of Latvia
Raina blvd. 19, Riga, Latvia
{vineta.arnicane, juris.borzovs}@lu.lv,
anete.nesaule@gmail.com

Abstract. Particular return to McCall’s et al. seminal work has been done in recent international standards ISO/IEC 25022:2016 and ISO/IEC 25023:2016. 122 lower level measures (metrics in McCall’s et al. terminology) were introduced. Authors of standards themselves classify quality measures as Highly Recommended, Recommended, Used at user’s Discretion and Generic, Specific. In this paper, quality measures are additionally classified also from the point of view of objectivity of measurement and of usage in industry.

Keywords: Software engineering · Software quality models · Quality characteristics · Quality metrics · Quality measures.

1 Introduction

We are not going to review software quality models because the authors of [12] already have done it. As James A. McCall, wrote in Wiley Online Library in 2002 [11]:

“All engineering disciplines have some basis in measurement. Software quality measurement frameworks have been developed to provide a measurement approach for software engineering. The frameworks have a common structure. At the highest level are quality factors, that comprise a definition of software quality and represent attributes or characteristics of the software, relate to its overall quality. The second level provides the criteria or software attributes that relate to the factors, and their existence provides the related characteristics of quality. The third level provides the “metrics” or measurements that measure the degree to which those attributes exist.

The use of software quality factors and metrics has become a common practice in the industry, although the applications are not consistent, systematic or typically applied across projects or organizations.” In their seminal paper [10] the authors introduced 11 quality factors, 23 criteria (i.e., sub-factors), and 42 metrics to measure software quality at the U.S. Air Force Electronic Systems Division’s and Rome Air Development Center’s mission to provide standards and technical guidance to software acquisition managers. Two types of metrics

were used. The first type, like a ruler, is a relative quantity measure. The second type is a binary measure which determines the existence (1) or absence (0) of something. The units of the metric were chosen as the ratio of actual occurrences to the possible number of occurrences. So, a total quality of a software product could be characterized as a sum of values of metrics.

For unknown reasons, idea of quantitative evaluation of software was not incorporated into international standard neither in its first version in 1991 [2], nor in its subsequent revisions [3–6]. This view is also shared by [1]: “The metrics in the lower level of the McCall’s, Boehm’s, Dromey’s and FURPS quality models are neither clearly nor completely defined and connected.” (However, to our opinion, it is not true regarding McCall’s lower level definitions.)

Particular return to McCall’s et al. seminal work has been done in recent standards ISO/IEC 25022:2016 [8] and ISO/IEC 25023:2016 [9]. 122 lower level measures (metrics in McCall’s et al. terminology) were introduced. In this paper, quality measures are additionally classified also from the point of view of objectivity of measurement and of usage in industry.

In Section 2 generally explains quality models provided by ISO/IEC 25022:2016 and ISO/IEC 25023:2016 standards. There is an investigation about objectivity of quality measures described in Section 3. The Section 4 provides insight into the capacity of measures to assess subcharacteristics and therefore characteristics of quality models. There is survey about usage of quality measures in the enterprises of Latvia elaborated in Section 5. Section 6 contains results of survey and Section 7 gives insight in conclusions and future work.

2 What do ISO/IEC 25022:2016 and ISO/IEC 25023:2016 offer?

ISO/IEC 25022:2016 and ISO/IEC 25023:2016 are the part of series of International Standards from ISO/IEC 25000 to ISO/IEC 25099 entitled *Systems and software engineering – Systems and software Quality Requirements and Evaluation* with the acronym *SQuaRE*. They give the measures which are supposed as appropriate to assess the extent or degree at what system or product satisfies each of subcharacteristics. There are two quality models defined in the standard ISO/IEC 25010:2011 [7] - *Quality in Use model* and *System/software product quality model*. Each of these models consists of quality characteristics and their subcharacteristics as shown in Fig. 3 and Fig. 4.

Each measure in ISO/IEC 25022:2016 and ISO/IEC 25023:2016 is classified as generic or specific. Generic are generally applicable, can be used in a wide range of situations while Specific are specialised for specific needs. For instance, generic measure is *MRe-1-G Reusability of assets - how many assets in a system can be reusable*. It is measured as ratio between assets which are designed and implemented to be reusable and assets in a system. As example for specific measure can be highlighted *REc-6-S Website visitors converted to customers - the proportion of visitors to a particular web page(s) who become customers*. It is measured using business analytics.

The system / software product quality model's measures are classified according to the recommendation level [9]. The authors of the standard ISO/IEC 25023:2016 advice to use highly recommended measures always. Accordingly recommended measures should be used as quality measures when it is appropriate. At the same time measures with unknown reliability are meant to be used at user's discretion as a reference when developing a new quality measure [9].

3 How objective are measures?

The measure can be applicable and reliable only if its measurements can be obtained objectively. We analysed measures mentioned in ISO/IEC 25022:2016 and ISO/IEC 25023:2016 and obtained assessment is shown in the Fig. 1. For

Ef-1-G	SPi-1-G	FCp-1-G	CIn-2-G	UEp-1-G	SCo-2-G	MMd-2-G	
Ef-2-S	SCo-1-G	FCr-1-G	CIn-3-S	UEp-2-S	SCo-3-S	MMd-3-S	
Ef-3-G	REc-1-G	FAP-1-G	UAp-1-G	UEp-3-S	SIn-1-G	MTE-1-G	
Ef-4-G	REc-2-G	FAP-2-G	UAp-2-S	UIn-1-S	SIn-2-G	MTE-2-S	
Ef-5-G	REc-3-G	PTb-1-G	UAp-3-S	UAc-1-G	SIn-3-S	MTE-3-S	
Ey-1-G	REc-4-G	PTb-2-G	ULe-1-G	UAc-2-S	SNo-1-G	PAd-1-G	
Ey-2-S	REc-5-S	PTb-3-G	ULe-2-S	RMa-1-G	SAC-1-G	PAd-2-G	
Ey-3-S	REc-6-S	PTb-4-G	ULe-3-S	RMa-2-G	SAC-2-S	PAd-3-S	
Ey-4-S	REc-7-S	PTb-5-G	ULe-4-S	RMa-3-G	SAu-1-G	Pln-1-G	
Ey-5-S	REc-8-G	PRu-1-G	UOp-1-G	RMa-4-S	SAu-2-S	Pln-2-G	
Ef-6-S	RHe-1-G	PRu-2-G	UOp-2-G	RAv-1-G	MMo-1-G	PRe-1-G	
Sus-1-G	RHe-2-G	PRu-3-G	UOp-3-S	RAv-2-G	MMo-2-S	PRe-2-S	Subjective (questionnaires, Likert, etc.)
Sus-2-G	RHe-3-G	PRu-4-S	UOp-4-S	RfT-1-G	MRe-1-G	PRe-3-S	Objective, but the result depends on the end-user qualification
Sus-3-G	REn-1-G	PCa-1-G	UOp-5-S	RfT-2-S	MRe-2-S	PRe-4-S	Objective for a specific product, but comparisons only relative
Sus-4-G	CCm-1-G	PCa-2-G	UOp-6-S	RfT-3-S	MAN-1-G		Objective, if industry criteria exist
Sus-5-G	CFI-1-S	PCa-3-S	UOp-7-S	RRe-1-G	MAN-2-S		Objective, can be used in comparison
Sus-6-G	CFI-2-S	CCo-1-G	UOp-8-S	RRe-2-S	MAN-3-S		
Str-1-G	CFI-3-S	CIn-1-G	UOp-9-S	SCo-1-G	MMd-1-G		

Fig. 1. Assessment of possibility to obtain objective measurements for measures.

instance, we assessed as objective measure *RAv-1-G Test function completeness - how completely are test functions and facilities implemented*. Measure is computed as as ratio between system operation time actually provided and system operation time specified in the operation schedule. This measure is objective because both parameters of ratio come from either from statistics or from schedule. The measurements can be used in order to compare two products.

Sus-2-G Satisfaction with features - the satisfaction of the user with specific system features is the example of subjective measure. It is measured using questionnaires. Judgement of people about system's features can greatly depend on their mood, level of tiredness, attitude to system, etc..

Measure *Ef-3-G Errors in task - the number of errors made by the user during a task* is the instance of measures which are measurable objectively but at the same time the result depends heavily on the end-user qualification in the field as well as the level of training to work with the system under test. That also is the reason why the obtained results are hardly applicable to comparison different systems or one system in the different stages of development.

Better situation are with next group of measures which can be measured objectively but the results are specific of product and can be used for comparison in the form of percents. For instance, measure *REc-2-G Time to achieve return on investment - the time taken to achieve the expected return on investment* which can be calculated by business analytics based on careful analysis, though, there is also a good dose of more or less subjective assumptions. Measure *MTe-*

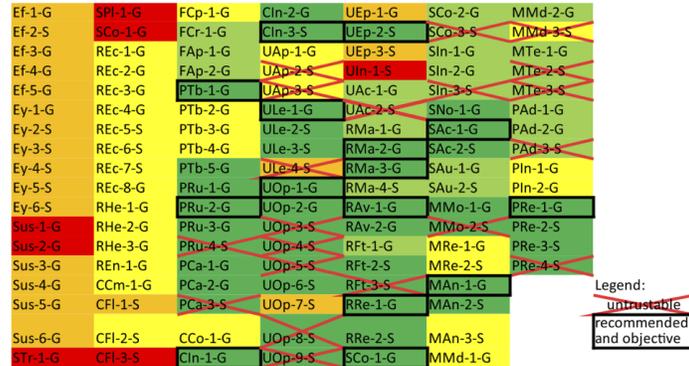


Fig. 2. Coloured quality characteristics for software or software product quality.

1-G Test function completeness - how completely are test functions and facilities implemented which is calculated as ratio between count of test functions implemented as specified and the count of test functions required, is a representative of group of measures which can be measured objectively if there exist industry criteria which help to understand how many test functions should be required. Otherwise this measure becomes very subjective.

4 Can measures given in ISO/IEC 25022:2016 and ISO/IEC 25023:2016 really assess the quality subcharacteristics?

Let us try to understand which of measures are usable from practical point of view - which are objective and reliable. If we put together our findings about objectivity of quality measures with recommendations of standard we would have obtained situation shown in the Fig. 2. We put red crosses on measures that the authors of standard ISO/IEC 25023:2016 considered as measures with unknown reliability. Next step - we put black box around those measures which are objective according to our analysis and are highly recommended by authors of standard.

If to look on Quality in use model's measures of subcharacteristics there are no any objectively measurable measure. That could be expected to happen

(see Fig. 3). Accordingly also quality characteristics cannot be well measured if their subcharacteristics are not such. The situation with Software product

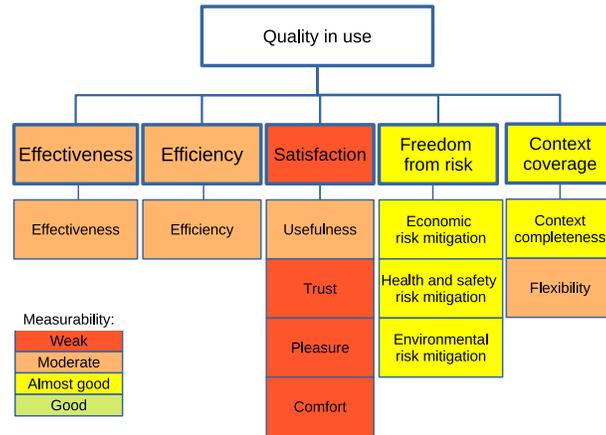


Fig. 3. Measurability of quality characteristics for Quality in use model.

quality model looks much better (see Fig. 4). The most difficult measurable are usability subcharacteristics while reliability, security and performance efficiency subcharacteristics looks measurable and reliable.

5 What measures are used in practise in Latvia?

5.1 Survey Objectives

The University of Latvia conducted a survey on measures used in the Latvia's industry of software development.

The primary objective of the survey was to determine measures of quality characteristics used in Latvia when they carry out software development and maintenance activities.

The aim was to find out which quality characteristics and subcharacteristics are currently used in the IT industry of Latvia and whether all quality measures included in ISO/IEC 25022:2016 and ISO/IEC 25023:2016 standards are well known by quality assurance specialists, testers and developers.

5.2 Survey Description

The survey targeted senior employees involved with quality assurance, testing or development in software development companies or departments of software development in enterprises with another kind of main business.

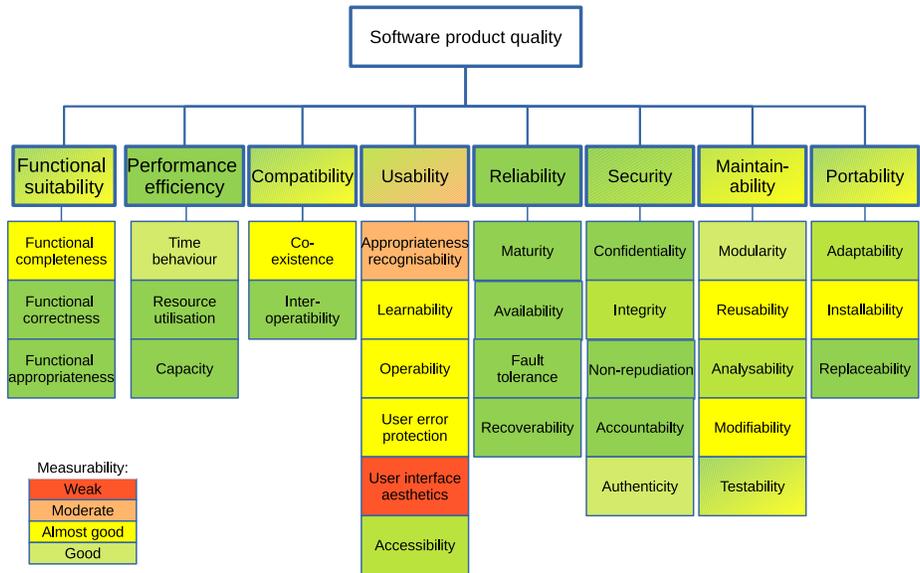


Fig. 4. Measurability of quality characteristics for software or software product quality.

There were three parts in the survey. The first part was an introductory section where the aims of survey and way how to fill it was described. The second part of the survey comprised questions about the company. The third part was the main part of the survey – 122 measures of 42 quality subcharacteristics according to standards ISO/IEC 25022:2016 and ISO/IEC 25023:2016.

The respondents of the survey were invited to complete the survey online. Each measure had predefined answers: 1)Yes; 2)No; 3)Other (*open type answer to give some specific answer*). Respondents were informed that survey terms are taken from standards ISO/IEC 25022:2016 and ISO/IEC 25023:2016. For each measure the definition was given, formulae was explained and the method of measuring was described. Confidentiality and privacy were assured to all respondents and the organization that they represented.

6 Analysis and Summary of Survey Findings

As a result, a total of 17 companies participated in the survey, representing more than 4000 IT employees including almost 800 testers. The survey results show which measures of system and software product quality as well as measure of quality in use are used in the IT industry of Latvia.

Summary of survey’s results is shown in the Fig. 5. Measures are color coded in five groups. The red ones are measures that are used rarely - in less than 20% of organizations, while yellow ones are used moderately and green ones frequently or very frequently - even in more than 80% of organizations.

Ef-1-G	SPI-1-G	FCp-1-G	CIn-2-G	UEp-1-G	SCo-2-G	MMd-2-G	
Ef-2-S	SCo-1-G	FCr-1-G	CIn-3-S	UEp-2-S	SCo-3-S	MMd-3-S	
Ef-3-G	REc-1-G	FAP-1-G	UAp-1-G	UEp-3-S	SIn-1-G	MTe-1-G	
Ef-4-G	REc-2-G	FAP-2-G	UAp-2-S	UIn-1-S	SIn-2-G	MTe-2-S	
Ef-5-G	REc-3-G	PTb-1-G	UAp-3-S	UAc-1-G	SIn-3-S	MTe-3-S	
Ey-1-G	REc-4-G	PTb-2-G	ULe-1-G	UAc-2-S	SNo-1-G	PAd-1-G	
Ey-2-S	REc-5-S	PTb-3-G	ULe-2-S	RMA-1-G	SAC-1-G	PAd-2-G	
Ey-3-S	REc-6-S	PTb-4-G	ULe-3-S	RMA-2-G	SAC-2-S	PAd-3-S	
Ey-4-S	REc-7-S	PTb-5-G	ULe-4-S	RMA-3-G	SAu-1-G	Pln-1-G	
Ey-5-S	REc-8-G	PRu-1-G	UOp-1-G	RMA-4-S	SAu-2-S	Pln-2-G	
Ef-6-S	RHe-1-G	PRu-2-G	UOp-2-G	RAv-1-G	MMo-1-G	PRe-1-G	
Sus-1-G	RHe-2-G	PRu-3-G	UOp-3-S	RAv-2-G	MMo-2-S	PRe-2-S	
Sus-2-G	RHe-3-G	PRu-4-S	UOp-4-S	RFT-1-G	MRe-1-G	PRe-3-S	
Sus-3-G	REn-1-G	PCa-1-G	UOp-5-S	RFT-2-S	MRe-2-S	PRe-4-S	Legend:
Sus-4-G	CCm-1-G	PCa-2-G	UOp-6-S	RFT-3-S	MAN-1-G		< 20%
Sus-5-G	CFI-1-S	PCa-3-S	UOp-7-S	RRe-1-G	MAN-2-S		20-40%
Sus-6-G	CFI-2-S	CCo-1-G	UOp-8-S	RRe-2-S	MAN-3-S		41-59%
STr-1-G	CFI-3-S	CIn-1-G	UOp-9-S	SCo-1-G	MMd-1-G		60-80%
							> 80%

Fig. 5. Results - at what extent measures are used in IT industry of Latvia.

As an example of rarely used measure in surveyed organizations can be mentioned *CIn-1-G Data formats exchangeability - what proportion of the specified data formats is exchangeable with other software or systems*. It is measured as ratio between data formats exchangeable with other software or systems and data formats specified to be exchangeable.

Next group of measures is colored in orange - these measures are used in 20-40% of surveyed organizations. As representative of moderately used measures colored in yellow are used in 41-59% of surveyed organizations. Green color indicates that measure is used frequently. Ones colored in light green are used in 60-89% of surveyed organizations.

The most frequently used measures are colored in dark green. There are only two measures used by more than 80% of surveyed organizations. The one is *PTb-1-G Mean response time* and the other is *Sus-1-G overall satisfaction - the overall satisfaction of the user* measured using questionnaires for users of system or software product.

7 Conclusions and Future Work

The authors of standards ISO/IEC 25022:2016 and ISO/IEC 25023:2016 have done hard work trying to define measures which could allow to assess subcharacteristics and therefore quality characteristics for Software in use and Software product quality models. They divide measures in two sets - generic for usage in wide range of situations while specific for specialised needs. For the Product quality model they give also classification of measures into sets of highly recommended, recommended and measures with unknown reliability.

If measures of both quality models are assessed from the objectivity viewpoint then we see that Quality in use measures all are more or less subjective while the most of Product quality model's measures are objective.

Survey about actual usage of quality measures in IT industry of Latvia shows that only two measures are used in more than 89% of surveyed enterprises. At the

same time one third part of all measures mentioned in standards is used in less than 20% of enterprises. Surprisingly, used measures are not mandatory reliable and objective. This fact urges to continue research in order to understand if it is enough for enterprises with those measures which they are using and, if not, why they do not use other measures.

8 Acknowledgements

The work described in this paper was supported by the project “Innovative information technologies” at the University of Latvia. We are grateful to all respondents of the survey. Any errors or omissions are the faults of the authors.

References

1. Al-Qutaish, Rafa E. Quality Models in Software Engineering Literature: An Analytical and Comparative Study. *Journal of American Science* **6**(3), 166–175 (2010)
2. ISO/IEC 9126:1991 Software engineering – Product quality. ISO/IEC, (1991)
3. ISO/IEC 9126-1:2001 Software engineering – Product quality – Part 1: Quality model. International Organization for Standardization, Geneva, Switzerland (2001)
4. ISO/IEC TR 9126-2:2003. Software Engineering - Product Quality - Part 2: External Metrics. International Organization for Standardization, Geneva, Switzerland (2003)
5. ISO/IEC TR 9126-3: 2003. Software Engineering - Product Quality - Part 3: Internal Metrics. International Organization for Standardization, Geneva, Switzerland (2003)
6. ISO/IEC TR 9126-4: 2004. Software Engineering - Product Quality - Part 4: Quality in Use Metrics. International Organization for Standardization, Geneva, Switzerland (2004)
7. ISO/IEC 25010:2011 Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models. International Organization for Standardization, Geneva, Switzerland (2011)
8. ISO/IEC 25022:2016 Systems and software engineering – Systems and software quality requirements and evaluation (SQuaRE) – Measurement of quality in use. International Organization for Standardization, Geneva, Switzerland (2016)
9. ISO/IEC 25023:2016 Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Measurement of system and software product quality. International Organization for Standardization, Geneva, Switzerland (2016)
10. McCall, J., Richards, P., Walters, G.: *Factors in Software Quality*, three volumes, NTIS AD-A049-014, 015, 055 (1977)
11. McCall, James A.: *Quality Factors*. In: *Wiley Online Library* (2002), <https://doi.org/10.1002/0471028959.sof265>. Last accessed 26 Feb 2020
12. Miguel, José P., Mauricio, D., Rodríguez, G.: A Review of Software Quality Models for the Evaluation of Software Products. *International Journal of Software Engineering Applications (IJSEA)*, **5**(6), 31–53, (2014)