A Key Performance Indicator Quality Model and Its Industrial Evaluation

Miroslaw Staron*, Wilhelm Meding†, Kent Niesel‡, Alain Abran§
*University of Gothenburg, Sweden
miroslaw.staron@gu.se
†Ericsson AB, Sweden
wilhelm.meding@ericsson.com
‡Volvo Car Group, Sweden
kent.niesel@volvocars.com
§Ecole de technologie superieure, Canada
alain.abran@etsmtl.ca

***Abstract*—Background: Modern software development companies increasingly rely on quantitative data in their decision-making for product releases, organizational performance assessment and monitoring of product quality. KPIs (Key Performance Indicators) are a critical element in the transformation of raw data (numbers) into decisions (indicators).**

Goal: The goal of the paper is to develop, document and evaluate a quality model for KPIs – addressing the research question of *What characterizes a good KPI?* In this paper we consider a KPI to be "good" when it is actionable and supports the organization in achieving its strategic goals.

Method: We use an action research collaborative project with an infrastructure provider company and an automotive OEM to develop and evaluate the model. We analyze a set of KPIs used at both companies and verify whether the organization's perception of these evaluated KPIs is aligned with the KPI's assessment according to our model.

Results: The results show that the model organizes good practices of KPI development and that it is easily used by the stakeholders to improve the quality of the KPIs or reduce the number of the KPIs.

Conclusions: Using the KPI quality model provides the possibility to increase the effect of the KPIs in the organization and decreases the risk of wasting resources for collecting KPI data which cannot be used in practice.

I. Introduction

Contemporary medium-to-large software development organizations often rely on quantitative information in monitoring their products and processes [17]. These companies use measures and indicators to both monitor the status and to plan long-term evolution of their business [12]. One of the tools used for this purpose is the notion of KPI – Key Performance Indicators.

In the ISO/IEC 15939:2007 [10] standard about measurement processes (Systems and Software Engineering – Measurement Processes), the notion of an indicator is defined in a broad sense – *Variable assigned a value by applying the analysis model to base and/or derived measures*. This definition requires the indicator to be composed of two parts – the value and the applied analysis model. In software engineering this kind of application of the analysis model usually results in using a color scale with the traffic light metaphor [11] and [21], [6].

However, there is also another notion close to the notion of the indicator – Key Performance Indicator (KPI), which is coined by another theoretical framework – Balanced Scorecard [5]. There, the KPI is defined as *Key performance indicators (KPIs) are customizable business measure utilized to visualize status and trends in an organization*. In contrast to a generic indicator the KPI has the link to the business operations (usually a link to the business strategy) and can be customized based on the changes in the business environment. Therefore in our work we focus on the KPIs in the latter sense, i.e. being a customizable business measure, as it allows us to place the numbers (measures and indicators) in the context of the organization defining and using them.

Although in theory the use of the KPIs should be straighforward, defining a high quality KPI requires good knowledge about the company, its business environment and about the data sets. Therefore, in this paper we define and evaluate a quality model for KPIs, essentially addressing the research question of

What characterizes a good KPI?

In short, in our work, we consider the KPI to be good when it is actionable and supports the organizational goals of the company in an objective and correct way. To address the research question we develop a quality model – KPI Quality model – which contains attributes of both the KPIs and the measurement procedures leading to the producing values of the KPIs. Using this model allows the companies to understand how well the KPIs are defined, aligned with the company's goals and how useful they are. It also allows the companies to understand what characterizes the proper infrastructure to support the collecting of the good KPIs.

As a basis for the development of the quality model we started with the ISO/IEC 9126 standard's distinction between the three types of quality – i) internal quality, ii) external quality and iii) quality in use. However, we have found that these three dimensions are too limiting as they do not capture the specifics of measurement processes – i.e. the measurement as data processing and the measurement as part of the organization and the field of studies. This distinction has been recognized in the practical applications of metrology in software engineering by Abran [1]. We extend this model by adding the quality attributes from the information quality based on the established frameworks [18], and [7]. Finally the we also add the organizational change adoption theories of Goodman and Dean [2] adapted to the notion of the indicator.

This KPI Quality model has been developed as part of an action research project conducted together with two companies – an infrastructure provider company, and an automotive OEM, which is a vehicle manufacturer. Since the companies are two different entities, their input to the evaluation increases the external validity of the results and allows the model to be generalized to other contexts in the software development industry.

The remaining of the paper is structured as follows. Section II presents the most relevant related work in the literature regarding the experiences of selecting dashboards. Section III describes the design of the action research project where the model was developed. Section IV contains the theoretical framing of the KPI quality model and the model itself. Section VI contains our experiences of applying the model to two industrial cases and section VIII summarizes and concludes the paper.

II. Related Work

We review work in two main areas – standardization of measures in software engineering and creating dashboards and visualizations.

A. Standardization

One of the main international standard in measurement is the common vocabulary in metrology – VIM [9]. The standard defines such concepts as measurement uncertainty, measurand and quantification. These concepts are important when setting up the measurement program and assessing its quality – in particular when considering the assessment of how the data should support the decisions at the company (e.g. whether the product is ready to be releases w.r.t. its quality, [22]).

VIM standardizes important concepts which influence measurement processes, for example:

- Measuring instrument: device used for making measurements, alone or in conjunction with supplementary device(s)
- Measuring system: set of one or more measuring instruments and often other devices, assembled and adapted to give measured quantity values within specified intervals for quantities of specified kinds

The VIM specifies the concepts, but does not prescribe any specific quality attributes of these. Our work presented in this paper contributes to developing a generic quality model of measurements – starting with the KPIs.

B. Dashboards and visualization

In our previous work we have studied the characteristics of effective and efficient measurement programs at large organizations [20]. Our results showed that there is a need for a more detailed model describing the characteristics of a good KPI in order to support the assessment of the quality of the content of measurement programs. During the discussions in the workshops in our previous study our industrial partners showed the need for a better description of what constitutes a good KPI – therefore the work presented in this paper addresses this need.

Staron and Meding [18] designed a set of principles of for assessing the reliability of information, which was the base for constructing one of the dimensions of the quality model. This method was proven to be useful when designing industrial measurement systems, e.g. for monitoring bottlenecks [19] and has been an expansion of the more generic quality model of information by Lee et al. [7]. This paper expands on both and adds the dimensions relevant to the measurement procedures as well as the information itself.

In our previous work we also studied how information visualization in form of models helps decision making in large companies – [8]. The results showed that the alignment of the type of model and the decision is one of the prerequisites for efficient software development and prevents waste.

III. Research design – Action Research

In this study we applied the principle of action research as advocated by Susman and Evered [23] and used in our previous studies with the same company [14], [15], [13]. The action research set-up provided us with a unique opportunity to be part of a project at an automotive OEM which aimed at a redesign of a large program status reporting tool. The tool was used to monitor the progress of car development projects and was divided into three parts – Key Performance Indicators, Milestone reporting and Risk monitoring. In our work we focused only on the Key Performance Indicators part as it was aligned with the researcher's competence and the company's interest.

The project provided us also with a possibility to work with the infrastructure provider when validating the model. We worked with the organization developing one of the infrastructure provider's products and consisting of over 100 engineers. The collaborating practitioners were part of the metric team who designs, deploys and operates over 4000 measurement systems. One of their responsibilities is to design and develop KPIs together with the stakeholders in the product development and maintenance organizations.

The research was organized in action research cycles, which is shown in table I.

Table I
ACTION RESEARCH CYCLES OF THIS PROJECT

Cycle	Goal	Outcomes
Project initialization	Understand the practices of using the tool	Plan for assessing the KPIs
Development of tools	Prepare research instruments	KPI quality model, dashboard selection model
Interviews	Collect the data	A set of dashboard selection models
Validation in another context	Validation of the method at the infrastructure provider company	Checking the applicability in another context
Metrological improvement	Adaptation to the metrological use of KPIs based on	Aligning the model with metrology [1]

In the first cycle we focused on refining the initial problem formulation – understanding how KPIs are used in the organizations and to study the documents explaining the usage of the KPIs, their definition and relation to company's strategic goals.

In the second cycle we prepared research instruments for defining the KPI Quality Model – preparing the dissemination patterns based on literature studies and discussions with the focus group at the company. In the third cycle we applied the model to a set of KPIs at the first case company – the automotive OEM. The application resulted in the first set of improvements – adding the ISO 25000 Data Quality attributes. After the application at the automotive OEM we applied it at the infrastructure provider to ensure the external generalizability of the model in the fourthe cycle. During this cycle we have also realized that the model required more alignment with the metrological foundations of measurement and therefore, in the fifth cycle, it was adapted to software metrology and evaluated again at the same infrastructure provider. The resulting model and its application are presented in this paper.

IV. FOUNDATIONS OF THE KPI QUALITY MODEL

A. Theoretical framing

Here we describe the theories used to design the model:

- ISO/IEC 15939 – measurement information model and its refinement [1],
- ISO/IEC 25000 – criteria for evaluation of measures,
- Organizational change adoption theory, and
- Information quality

1) ISO/IEC 15939 measurement information model: ISO/IEC 15939 is not the only framework available for structuring measures in measurement systems. The core component of the standard is the conceptual measurement information model which describes relationships between the main types of elements of measuring systems.

The information need is an insight necessary for a stakeholder to manage objectives, goals, risks, and problems observed in the measured objects. These measured objects can be entities like projects, organizations, software products, etc. characterized by a set of attributes. ISO/IEC 15939 includes the following definitions, which are relevant:

- Entity – object that is to be characterized by measuring its attributes.
- Attribute – property or characteristics of an entity that can be distinguished quantitatively or qualitatively by human or automated means.
- Base measure – measure defined in terms of an attribute and the method for quantifying it.
- Derived measure – measure that is defined as a function of two or more values of base measures.
- Decision criteria – thresholds, targets, or patterns used to determine the need for action or further investigation, or to describe the level of confidence in a given result.
- Indicator – measure that provides an estimate or evaluation of specified attributes derived from a model with respect to defined information needs.
- Information product – one or more indicators and their associated interpretations that address an information need.
- Stakeholder – a person who has the mandate and possibility to act upon the status of the indicator.

This standard also defines a set of criteria for evaluating the information products (which include the indicators according to the definition above), these are:

- Confidence in the information product
- Evidence for the fitness of purpose of the information product
- Understandability of the information product
- Satisfaction of the assumptions of the information product
- Accuracy of the measurement procedure
- Repeatability of the measurement method
- Reproducibility of the measurement method

The model of ISO 15939 is focused on the measures and their relationships, but does not take into account the organizational aspects of the measures – e.g. whether a measure or an indicator is appropriate for the organization or how it should be interpreted in the organizational context. A refinement and extension of the measurement information model can be found in the metrological understanding of the measurement process presented by Abran [1], and is presented in figure 1.

On the right-hand side the model in figure 1 contains the elements of the ISO 15939 standard, grouped into three parts – data collection to describe the process of quantification of the measurand, data processing to describe the process of mathematical operations on the collected measures and data analysis to complement the mathematical analyses with the interpretation of the measures, in the context of an organization and the state-of-the-art reference values universal for specific areas of measurement. This process describes the natural use of metrology in software engineering – starting

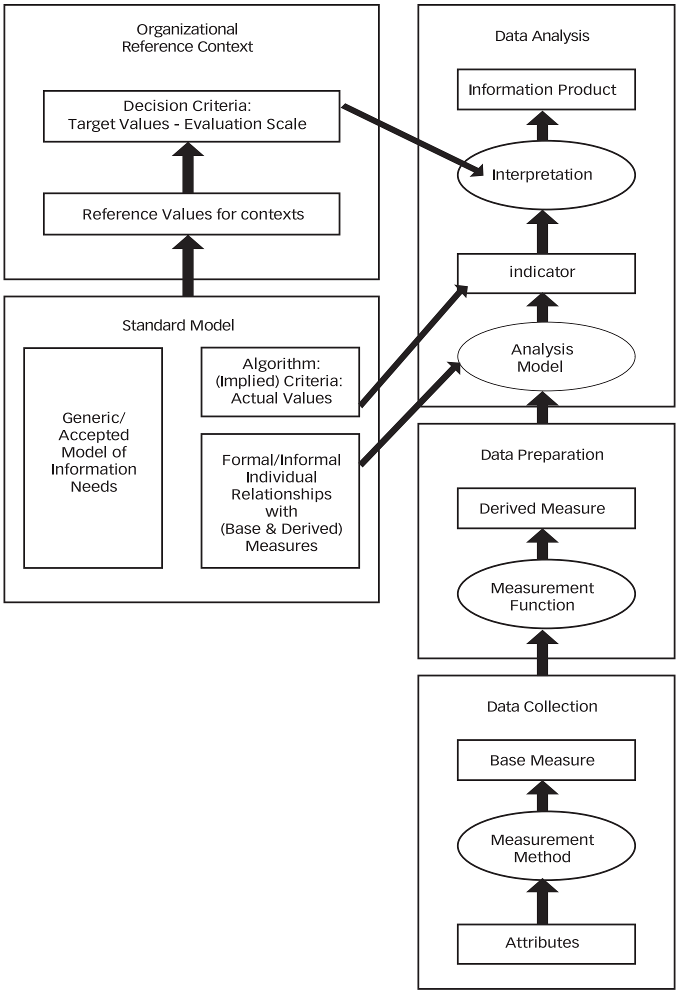

Figure 1. Refined analysis model of ISO/IEC 15939 with metrological standard reference model and organizational reference context. Adopted from [1], figure 4.4

from the quantification (transforming empirical properties into mathematical symbols) through the use of mathematical operations to make inferences and calculations to the interpretation of the numbers in the organizational context (transforming mathematical symbols into empirical facts).

In our KPI quality model we follow the same structure and grouping as presented in figure 1, grouping the quality attributes accordingly.

2) ISO/IEC 25000 criteria for assessing data quality: In the ISO 15939 standard the focus is on the process of measuring and not on the actual definition of measures (or which measures should be used). Therefore, it should be complemented by the ISO 25000 series of standards where the focus is on the definition of quality measures. One of the elements which are new in the ISO/IEC 25000 series of standards as compared to the ISO 9126 standard is the set of criteria for evaluating the quality of data (measures in our case are a subclass of data). These are listed as:

- Accuracy
- Completeness
- Consistency
- Credibility
- Currentness
- Accessibility
- Compliance
- Confidentiality
- Efficiency
- Precision
- Traceability
- Understandability
- Availability
- Portability
- Recoverability

As they partially overlap with the information quality attributes (e.g. Protection of privacy) we need to unify these two lists in order to create the KPI quality model. They are also an extension of the set of criteria from the ISO/IEC 15939:2007 standard as presented in the previous subsection. These could be complemented with the set of quality attributes which were present in earlier drafts of the model:

- Relevance to the prioritized information needs
- Repeatability and reproducibility of quality measure elements
- Predictive validity of software quality measure
- Feasibility of collecting the data in the organizational unit
- Availability of human resources to collect, analyze and manage data
- Ease of data collection
- Availability of appropriate tools
- Protection of privacy
- Number of potentially relevant indicators supported by the required quality measure elements
- Ease of interpretation by measurement users and measurement analysts
- Number of users or consumers of the information products utilizing the indicator
- Life cycle stage applicability
- Evidence (internal or external to the organizational unit) as to the measures fitness for purpose of the Information need
- Characteristics of the required quality measure elements (e.g., classification dimensions in ISO/IEC 25021)

3) Organizational change adoption theory: When discussing the effectiveness of the KPIs or their quality in use, we choose the organizational change adoption theory by Goodman et al. [3], the later extension by Goodman and Dean [4], and finally its adaptation to the metrics adoption [2]. The theory of change in organization evaluates how well a change is *institutionalized* in the organization – i.e. adopted

– and the impact it has on the organization.

The theory of organizational change uses the notion of degree of institutionalization of the change, organized in five facets when conceptualizing the institutionalization:

1) knowledge of the behavior – concerns the extent to which an individual possesses the knowledge of the change in order to be able to perform the new tasks; in terms of the measurement program this relates to how much the organization knows about the measurement program,
2) performance of the behavior (which we rename to "performance of the measurement" in the paper) – concerns the performance of the new tasks; in terms of the measurement program this relates to whether the measurement activities are performed in the organization (for example by a dedicated group of individuals),
3) preferences of the behavior – concerns the attitude to the change; in terms of the measurement program this means that it is perceived as "good" by the organization,
4) normative consensus – concerns the general acceptance of the change – the knowledge that others perform the activity and consensus about the appropriateness of them; in terms of the measurement program this means how well the measurement results from the program are used (for example for making decisions), and
5) values of the behavior – concerns social consensus on the values relevant to the behavior; in terms of measurement systems this means that there is a consensus on what ought to and ought not to be measured and why.

This theory allows us to understand the dimensions of how a good KPI is used in the organization – thus providing the solid theoretical foundation of the quality in use dimension of the quality model. This approach has also been used previously in the context of assessing the robustness of measurement programs [20].

4) Information quality: Information quality is important to assess whether the information provided by measurement systems can be trusted or not [18]. The annex D of ISO/IEC 15939 with the methods for assessing the quality of measurement systems provides a basic set of criteria. However, we use the AIMQ framework [7] instead, as it provides a quality model of information quality which is more extensive and measurable. The AIMQ framework defines 15 quality attributes of information: accessibility, appropriate amount, believability, concise representation, consistent representation, ease of operation, interpretability, objectivity, relevancy, reputation, understandability, timeliness, free of error, completeness, and security. The information quality defines the quality of the information, e.g. whether a given metric measures what it is supposed to measure.

5) Summary of collected attributes: In total we can observe that the above sources contribute with 56 quality attributes. The attributes relate to the categories of the quality model – Standard reference model, organizational reference model, data collection, data preparation and data analysis, as we show in figure 2.

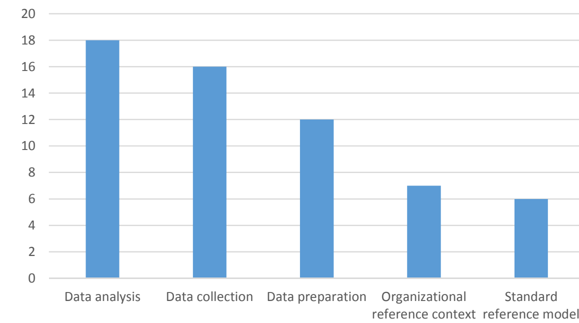

Figure 2. Collected quality attributes per quality dimension

In the list, however, there are attributes which come from multiple sources, but are overlapping (e.g. repeatability in AIMQ framework and repeatability of the measurement method in ISO/IEC 25000). Therefore the number of attributes in the figures is higher than the number of attributes of the quality model.

During the discussion of these attributes with our industrial partners we found that these attributes need to be complemented with a set of attributes that are important in practice:

- Transparency – it should be transparent to the stakeholders how the indicators are calculated, even the provenance, origins and flow of the data should be transparent. In terms of business intelligence the KPI should be described using the technical meta-data.
- Actionable – the stakeholders should be able to take actions based on the indicators, and thus linking the KPIs to the source data in a straightfoward way is preferred.
- Traceable – the KPIs should be traced to the strategy of the organization – the stakeholders should have the possibility to point how the KPIs contribute to the fulfillment of the strategy of the organization. In terms of business intelligence the KPI should be described in terms of the business meta-data.

This provided us with the 59 quality attributes in our model.

V. KPI Quality model

The overview of the model and its attributes are presented in figure 3. In order to operationalize the KPI quality model we changed the wording of the original attributes and organized each of the quality dimensions into sub categories based on the figure 1 from section IV-A.

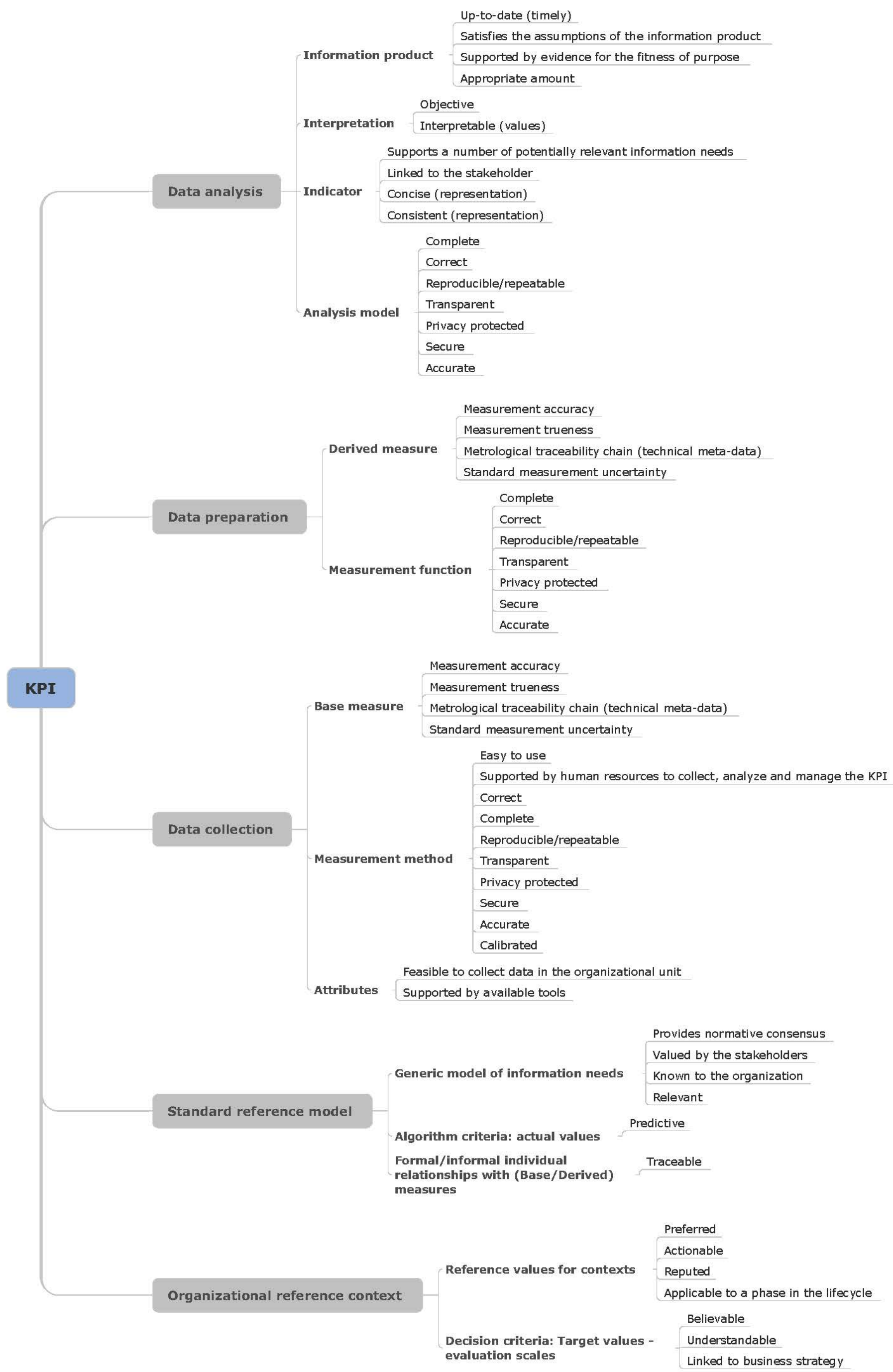

Figure 3. KPI quality model

A. *Data analysis*

This set of quality attributes are related to the data analysis and the interpretation of the measures in their context as presented in table II. Here we use four sub-categories:

1) Information product – grouping the quality attributes describing the information product
2) Interpretation – grouping the quality attributes related to how the indicators are interpreted, thus capturing the mapping from the mathematical symbols domain to the empirical domain
3) Indicator – grouping the quality attributes related to the number and the assigned status
4) Analysis model – grouping the quality attributes re-

lated to how the status of the indicator is assessed using the predefined criteria

These four sub-categories describe the properties a good indicator should have in terms of its construction. These are related directly to the next category – organizational reference context.

Table II
DATA ANALYSIS ATTRIBUTES

Sub-category	Attribute	Explanation
Inf. product	Up-to-date	The KPI value is timely, i.e. updated according to the schedule.
Inf. product	Satisfies the assumptions of the information product	The KPI is linked to the business goals of the organization.
Inf. product	Supported by the evidence for the fitness of purpose	There is evidence that the KPI is well fit for the stakeholder's information need.
Inf. product	Appropriate amount	The KPI contains the appropriate amount of information to fulfill the stakeholder's information need
Interpretation	Objective	The KPI objectively quantifies the measured entities
Interpretation	Interpretable	The KPI can be interpreted in the organization based on the data collected and the situational context.
Indicator	Supports...	The KPI can support (after adjustment) more than one information need of the organization.
Indicator	Linked to the stakeholder	The KPI is linked to a specific stakeholder who has the mandate and ability to act upon the information provided by the KPI
Indicator	Concise	The representation of the KPI is concise and does not include unnecessary details.
Indicator	Consistent	The KPIs representation is consistent with the goals of the KPI.
Analysis model	Complete	All steps of the measurement procedure are performed.
Analysis model	Correct	All the steps of the measurement procedure are performed correctly.
Analysis model	Reproducible/ repeatable	The measurement procedure is either documented or automated so that it is possible to reproduce the results.
Analysis model	Transparent	It is clear how the measurement procedure is performed and how the results are obtained.
Analysis model	Privacy protected	The procedure is created in such a way that it protects the privacy of the measured entities.
Analysis model	Secure	The procedure is created in such a way that it is protected from unauthorized tempering.
Analysis model	Accurate	The procedure truly reflects the measured attributes of the measured entities.

B. Data preparation

In the second phase of data processing the model groups elements related to data preparation – a set of mathematical operations on the measures transforming base measures into derived measures. Table III presents these attributes.

Table III
DATA PREPARATION ATTIBUTES

Sub-category	Attribute	Explanation
Derived measure	Measurement accuracy	The measurement error should be as low as possible
Derived measure	Measurement trueness	The value should reflect the true value of the measurand
Derived measure	Metrological traceability chain	There is a meta-data describing the technical details of the measure (e.g. its data type)
Derived measure	Standard measurement uncertainty	The measurement error of the measure is provided
Measurement function	Complete	All steps of the measurement procedure are performed.
Measurement function	Correct	All the steps of the measurement procedure are performed correctly.
Measurement function	Reproducible/ repeatable	The measurement procedure is either documented or automated so that it is possible to reproduce the results.
Measurement function	Transparent	It is clear how the measurement procedure is performed and how the results are obtained.
Measurement function	Privacy protected	The procedure is created in such a way that it protects the privacy of the measured entities.
Measurement function	Secure	The procedure is created in such a way that it is protected from unauthorized tempering.
Measurement function	Accurate	The procedure truly reflects the measured attributes of the measured entities.

C. Data collection

The rationale behind this group of attributes is to describe the quality of the process of quantification of the measurand to a number. According to ISO 15939 this happens by defining and applying a measurement method to a measurand and the process results in a value of a defined base measure. Table IV presents the quality attributes. They are group into three logically distinct sub-categories:

1) Base measure – describing the quality of the measure (result of the measurement)
2) Measurement method – describing the quality of the measurement procedure
3) Attribute – describing the quality of the attribute being measured (measurand)

The reason for using these three sub-categories is that they have disctinct logical characteristics – mathematical (measure), camputational (method) and empirical (attribute).

D. Organizational reference context

This category of the quality model describes the quality of the definition and reference to the organizational context of the KPI. In particular the relation to its stakeholders, ability to stimulate action or credibility of the KPI. The set of quality attributes is divided into two sub-categories – reference values for the context and the decision criteria as shown in table V.

The organizational reference context describes the use of the KPI in the organization, which is often organization-specific. However, the KPIs often relate to established normative references for the fields of study – e.g. budget allocations in software projects have to correspond to the rules of budgeting in the economical domain. Therefore it

Table IV
DATA COLLECTION ATTRIBUTES

Sub-category	Attribute	Explanation
Base measure	Measurement accuracy	The measurement error should be as low as possible
Base measure	Measurement trueness	The value should reflect the true value of the measurand
Base measure	Metrological traceability chain	There is a meta-data describing the technical details of the measure (e.g. its data type)
Base measure	Standard measurement uncertainty	The measurement error of the measure is provided
Meas. method	Complete	All steps of the measurement procedure are performed.
Meas. method	Correct	All the steps of the measurement procedure are performed correctly.
Meas. method	Reproducible/ repeatable	The measurement procedure is either documented or automated so that it is possible to reproduce the results.
Meas. method	Transparent	It is clear how the measurement procedure is performed and how the results are obtained.
Meas. method	Privacy protected	The procedure is created in such a way that it protects the privacy of the measured entities.
Meas. method	Secure	The procedure is created in such a way that it is protected from unauthorized tempering.
Meas. method	Accurate	The procedure truly reflects the measured attributes of the measured entities.
Meas. method	Calibrated	The measurement method is calibrated to the type of the measurand
Meas. method	Easy to use	The measurement procedure is easy to use according to its users.
Meas. method	Supported by human resources to collect, analyse and manage the KPI	There is a measurement team supporting the data collection, analysis and management for the KPI
Attributes	Feasible to collect in the organization	It is possible to objectively collect the data either manually or automatically.
Attributes	Supported by available tools	There are tools which support the measurement procedure (e.g. scripts measuring the size of the software)

Table V
ORGANIZATIONAL REFERENCE CONTEXT ATTRIBUTES

Sub-category	Attribute	Explanation
Reference values for context	Preferred	The stakeholders have the preference for the KPI (i.e. judge it as useful).
Reference values for context	Actionable	The KPI allows the stakeholders to take concrete actions (i.e. there is an action plan linked to the KPI).
Reference values for context	Reputed	The KPI has a good reputation in the organization to lead to the right decisions and actions.
Reference values for context	Applicable...	It is clear how the KPI is applicable to one or more stages of the product lifecycle.
Decision criteria...	Believable	The KPI is believable.
Decision criteria...	Understandable	The KPI is understandable by the stakeholders and by the organization.
Decision criteria...	Linked to the business strategy	The KPI is linked to a specific business strategy of the organization.

is important that the organizational reference context is an adaptation of the standard reference model.

E. *Standard reference model*

In the last category of our quality model describes the quality of the links of the KPIs to the standard reference model as shown in table VI. It is divided into three subcategories – one regarding the generic model of information needs (which addresses the question why this KPI is to be defined in terms of standard needs of the field), the algorithm criteria (how it should be calculated in general) and finally the quality of the relation of the KPI towards the measures which are used in order to calculate the KPI.

Table VI
STANDARD REFERENCE MODEL ATTRIBUTES

Sub-category	Attribute	Explanation
Generic model of inf. needs	Provides normative consensus	The KPI should provide the consensus that improving the KPI will lead to improvements in the organization.
Generic model of inf. needs	Valued by the stakeholders	The KPI is perceived to bring value to the stakeholders and their information needs.
Generic model of inf. needs	Known to the organization	The KPI is known (disseminated) in the organization.
Generic model of inf. needs	Relevant	The KPI is relevant to the prioritized information needs
Algorithm criteria: actual values	Predictive	The KPI is valid from the empirical perspective and can be used for making predictions.
Formal/information rel...	Traceable	The results obtainable by using the measurement procedure are able to be traced back to the sources of the data.

VI. EVALUATION – APPLICATION OF THE KPI QUALITY MODEL

We evaluated the KPI quality model in every cycle of the action research described in section III, and in this paper we report only on the last cycle in order to show how the evaluation was conducted in practice. Each evaluation was planned in the same way:

1) Identification of KPIs to assess using the model
2) Assessment (researchers together with a senior measurement program leader at the company)
3) Visualization and discussions (researchers together with the senior measurement program leader)

The outcome of each evaluation was a set consisting of one diagram showing how well the KPI fulfills the quality attributes, the list of improvements and feedback from the senior measurement program leader.

For the final evaluation we chose four KPIs from the infrastructure provider which are used to monitor the performance of one of their development organizations. These KPIs are:

- KPI1 - Speed
- KPI2 - Quality
- KPI3 - Customer
- KPI4 - Market situation

After assessing their quality we counted the number of attributes possessed by these KPIs and divided the number by the total number of attributes. This ratio is presented per category in figure 4.

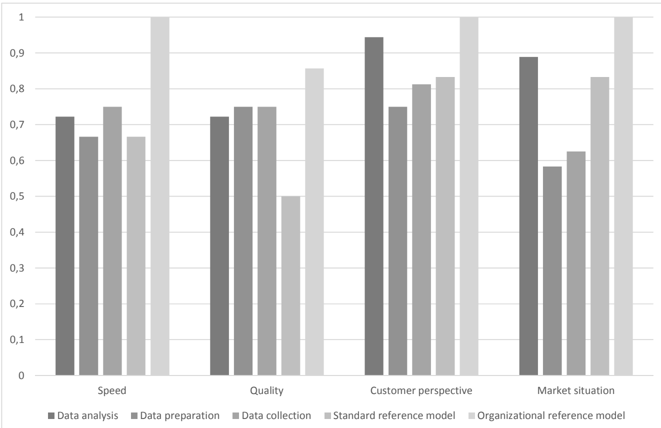

Figure 4. Evaluation results

The figure shows that all of the KPIs have all the attributes in the category of *Organizational reference model*. Since these KPIs are defined by a mature measurement organization the stakeholders are aware of the needs for the KPIs and their role in the organization. They also have well-defined information needs, which are satisfied by these KPIs.

However, since the field of software engineering does not have the same standardized set of KPIs the category of *Standard reference model* has lower number of fulfilled quality attributes.

Finally, we asked the senior measurement program leader who participated in the evaluation for the quote and he summarized it with the following quote:

"Having the method in place gives me the confidence that collecting, setting up and following up the KPIs guarantees that I will achieve the goals that I set for these KPIs. The reason for that is two fold – for the first, I have an clear understanding of the back-end of the KPI (infrastructure, visualization, standardization) and for the second I have the understanding how to use it correctly when it is in place."

This evaluation shows that the KPI quality model is applicable in industrial settings and that the results of its application reflect the senior measurement program leader's opinion about the quality of their KPIs.

VII. Threats to validity

In this paper we used the framework presented by Runeson et al [16] for evaluating the validity of empirical studies in software engineering. We recognized and addressed the following threats to validity.

The main threat to the *construct validity* in this study is the fact that we use different theoretical foundations to construct the model – e.g. Goodman's organizational change theory and the ISO 15939 standard. These are written with different purposes in mind and therefore can cause differences in interpretation in the KPI context. In order to minimize this we performed the evaluation of the model in industry.

The main threat to the *internal validity* is the evaluation of the model in the same context where it was designed, namely at the infrastructure provider company. Since in this study we had the possibility to work with two companies, we could evaluate the model at the automotive OEM with similar results, which minimizes this threat.

Since we do not use inferential statistics during our evaluation we naturally have the threat of low *conclusion validity*. In order to minimize this threat we visualized the results using the diagram and discussed these results with the stakeholder in order to avoid misinterpretation.

Finally, as in every action research project, there is a risk that our results are only applicable to the context where they were designed (i.e. the *external validity*). We believe that bootstrapping our work in international standards and using metrology as the foundation minimizes the risk of using company-specific quality attributes in the model. This increases the generalizability and external validity of the results.

VIII. Conclusions

In this paper we revisited the notion of Key Performance Indicator – a KPI. This notion becomes increasingly important in software engineering as it allows to provide interpretation of the measures in the context of organizational business goals, standard reference models and is aligned with the international standards on measurements, such as ISO/IEC 15959.

In particular we have reviewed the KPI concept from the perspective of software metrology, measurement processes, organizational contexts and the view of practitioners. We have developed a model of a good KPI that fulfills 59 quality attributes of the model presented in this paper. Fulfilling of these quality attributes makes the KPI aligned with the standard reference models of the field of software engineering (whenever such exists), the organizational context of the KPI, the quality of the KPI as a data point and the ability of the KPI to stimulate action in the organization.

Our results are important for contemporary medium-to-large software development organizations, which often rely on quantitative information in monitoring their products and processes [17]. These kind of companies use measures, indicators and KPIs to both monitor the status and to plan long-term evolution of their business [12].

In our further work we intend to further develop the KPI quality model and collect more data to enable statistical analyses whether there are correlations between quality attributes in practice.

Acknowledgment

This work has been partially supported by the Swedish Strategic Research Foundation under the grant number

SM13-0007. This research has been partially carried out in the Software Centre, University of Gothenburg, Ericsson AB, and Volvo Car Group.

REFERENCES

[1] Alain Abran. *Software metrics and software metrology*. John Wiley & Sons, 2010.

[2] Paul Goodman. *Practical implementation of software metrics*. International software quality assurance series. McGraw-Hill, London, 1993. lc92042989 Paul Goodman.

[3] Paul S Goodman, Max Bazerman, and Edward Conlon. Institutionalization of planned organizational change. In *Research in Organizational Behavior*, pages 215–246. JAI Press,Greenwich, 1980.

[4] Paul S Goodman and James W Dean Jr. Creating long-term organizational change. In *Change In Organizations*. Carnegie-Mellon Univ Pittsburgh, PA, Graduate School of Industrial Administration, 1982.

[5] Robert S Kaplan and David P Norton. *The balanced score-card: translating strategy into action*. Harvard Business Press, 1996.

[6] Ludwik Kuzniarz and Miroslaw Staron. Inconsistencies in student designs. In *Proceedings of the 2nd Workshop on Consistency Problems in UML-based Software Development*, pages 9–17, 2003.

[7] Yang W Lee, Diane M Strong, Beverly K Kahn, and Richard Y Wang. Aimq: a methodology for information quality assessment. *Information & management*, 40(2):133–146, 2002.

[8] Niklas Mellegård and Miroslaw Staron. Characterizing model usage in embedded software engineering: a case study. In *Proceedings of the Fourth European Conference on Software Architecture: Companion Volume*, pages 245–252. ACM, 2010.

[9] International Bureau of Weights and Measures. *International vocabulary of basic and general terms in metrology*. International Organization for Standardization, Genve, Switzerland, 2nd edition, 1993.

[10] International Standard Organization and International Electrotechnical Commission. Software and systems engineering, software measurement process. Technical report, ISO/IEC, 2007.

[11] Kosta Pandazo, Arisa Shollo, Miroslaw Staron, and Wilhelm Meding. Presenting software metrics indicators: a case study. In *Proceedings of the 20th International Conference on Software Product and Process Measurement (MENSURA)*, volume 20, 2010.

[12] David Parmenter. *Key performance indicators (KPI): developing, implementing, and using winning KPIs*. John Wiley & Sons, 2010.

[13] Rakesh Rana, Miroslaw Staron, Christian Berger, Jörgen Hansson, Martin Nilsson, Fredrik Törner, Wilhelm Meding, and Christoffer Höglund. Selecting software reliability growth models and improving their predictive accuracy using historical projects data. *Journal of Systems and Software*, 98:59–78, 2014.

[14] Rakesh Rana, Miroslaw Staron, Claire Berger, Jorgen Hansson, Martin Nilsson, and Fredrik Torner. Evaluating long-term predictive power of standard reliability growth models on automotive systems. In *Software Reliability Engineering (ISSRE), 2013 IEEE 24th International Symposium on*, pages 228–237. IEEE, 2013.

[15] Rakesh Rana, Miroslaw Staron, Niklas Mellegård, Christian Berger, Jörgen Hansson, Martin Nilsson, and Fredrik Törner. Evaluation of standard reliability growth models in the context of automotive software systems. In *Product-Focused Software Process Improvement*, pages 324–329. Springer, 2013.

[16] Per Runeson, Martin Host, Austen Rainer, and Bjorn Regnell. *Case study research in software engineering: Guidelines and examples*. John Wiley & Sons, 2012.

[17] Miroslaw Staron. Critical role of measures in decision processes: Managerial and technical measures in the context of large software development organizations. *Information and Software Technology*, 54(8):887–899, 2012.

[18] Miroslaw Staron and Wilhelm Meding. Ensuring reliability of information provided by measurement systems. In *Software Process and Product Measurement*, pages 1–16. Springer, 2009.

[19] Miroslaw Staron and Wilhelm Meding. Monitoring bottlenecks in agile and lean software development projects–a method and its industrial use. *Product-Focused Software Process Improvement*, pages 3–16, 2011.

[20] Miroslaw Staron and Wilhelm Meding. Mesram–a method for assessing robustness of measurement programs in large software development organizations and its industrial evaluation. *Journal of Systems and Software*, 113:76–100, 2016.

[21] Miroslaw Staron, Wilhelm Meding, Jörgen Hansson, Christoffer Höglund, Kent Niesel, and Vilhelm Bergmann. Dashboards for continuous monitoring of quality for software product under development. *System Qualities and Software Architecture (SQSA)*, 2013.

[22] Miroslaw Staron, Wilhelm Meding, and Klas Palm. Release readiness indicator for mature agile and lean software development projects. In *Agile Processes in Software Engineering and Extreme Programming*, pages 93–107. Springer, 2012.

[23] Gerald I Susman and Roger D Evered. An assessment of the scientific merits of action research. *Administrative science quarterly*, pages 582–603, 1978.