



www.gelog.etsmtl.ca

***Investigation of the Metrology Concepts in
ISO 9126
on Software Product Quality Evaluation***

Alain Abran

Rafa E. Al-Qutaish

Juan J. Cuadeado-Gallego

Agenda

Introduction

Metrology: An Overview

ISO 9126 & Quality in Use Metrics

Analysis of the Quality in Use Metrics

Effectiveness
Metrics

Productivity
Metrics

Safety
Metrics

Satisfaction
Metrics

Conclusion and Suggestions

Introduction

- ❖ In the field of software engineering, the term “metrics” is used in reference to multiple concepts:
 - ◆ Quantity to be measured
 - ◆ Measurement procedures
 - ◆ Measurement results
 - ◆ Models of relationships across multiple measures, or of the objects themselves

- ❖ In the software engineering literature, the term is applied, for instance, to:
 - ◆ A measure of a concept (e.g. McCabe cyclomatic complexity)
 - ◆ Quality models (ISO 9126 – software product quality)
 - ◆ Estimation models (e.g. Halstead's equations and COCOMO I and II estimation models)

- ❖ This has led to many curious problems, among them:
 - ◆ a rapid growth of numerous publications on metrics for concepts of interest, but with a very low rate of acceptance and use by either researchers or practitioners.
 - ◆ a lack of consensus on how to validate so many proposals.
- ❖ It is not seen to be economically feasible for either industry or the research community to investigate each of the hundreds of alternatives proposed metrics to date.

- ❖ Metrology has a long tradition of use in physics and chemistry.
- ❖ It is rarely referred to in the software engineering literature:
- ❖ Carnahan and others (1997) are among the first authors to identify this gap in what they referred to as “IT metrology”.
 - ◆ They have proposed logical relationships between metrology concepts, consisting of four steps:
 - ◆ Defining quantity/attribute,
 - ◆ Identifying units and scales,
 - ◆ Determining the primary references and
 - ◆ Settling the secondary references.

- ❖ Gray (1999) discusses the applicability of metrology to information technology from the software measurement point of view.
- ❖ Abran (1998) has highlighted some high-level ambiguities in the domain of software measurement:
 - ◆ He proposed substituting the appropriate metrology terms for the current ambiguous and peculiar software metrics terminology.

- ❖ The ISO 9126 series of documents on software product quality evaluation proposes a set of 120 metrics for measuring the various characteristics and subcharacteristics of software quality.
- ❖ The set of so-called metrics in ISO 9126 refers to multiple distinct concepts which, in metrology, would have distinct labels (or naming conventions, e.g. terms) to avoid ambiguities.

Metrology: An Overview

- ❖ Metrology has been defined as:
 - ◆ The field of knowledge dealing with measurement.
 - ◆ That portion of measurement science used to provide, maintain, and disseminate a consistent set of units to provide data for quality control in manufacturing.
- ❖ Each of the different interpretations of software metrics is associated to a related distinct metrology term with related metrology criteria and relationships with other measurement concepts.

- ❖ Three editions of the ISO International Vocabulary of Basic and General Terms in Metrology (VIM).
- ❖ The second VIM edition on metrology presents 120 terms in six categories and in increasing order of complexity, and describes each term individually in textual format:
 1. Quantities and Units (22 terms)
 2. Measurements (9 terms)
 3. Measurement Results (16 terms)
 4. Measuring Instruments (31 terms)
 5. Characteristics of the Measuring Instruments (28 terms)
 6. Measurement Standards – Etalon (14 terms)

- ❖ Two of the above six categories of terms deal with some aspects of the design of measurement methods:
 1. Quantities and units
 2. Measurement standards – etalon
- ❖ In this paper, we will use the first category.

ISO 9126 & Quality in Use Metrics

- ❖ In 1991, the ISO published its first international consensus on the terminology for the quality characteristics for software product evaluation.
- ❖ From 2001 to 2004, the ISO published an expanded four-part version, containing both the ISO quality models and inventories of proposed measures for these models; that is:
 - ◆ Software Product Quality Model (ISO 9126-1)
 - ◆ Software Product External Quality Metrics (ISO 9126-2)
 - ◆ Software Product Internal Quality Metrics (ISO 9126-3)
 - ◆ Software Product Quality in Use Metrics (ISO 9126-4)

- ❖ In ISO 9126-4, fifteen metrics have been proposed for the software product quality in use. They have been classified into four collections of metrics based on the characteristics presented in ISO 9126-1:
 - ◆ Effectiveness: task effectiveness, task completion and error frequency.
 - ◆ Satisfaction: task time, task efficiency, economic productivity, productive proportion and relative user efficiency.
 - ◆ Safety: user health and safety, safety of people affected by use of the system, economic damage and software damage.
 - ◆ Productivity: satisfaction scale, satisfaction questionnaire and discretionary usage.

- ❖ These fifteen metrics are analyzed using a metrology concept structure from the VIM category, *Quantities and Units*, based on four characteristics, that is:
 1. system of quantities.
 2. dimension of a quantity.
 3. unit of measurement.
 4. value of a quantity.
- ❖ Such analysis will contribute to identifying the measurement concepts that have not yet been tackled in the ISO 9126 series of documents.
- ❖ And it will represent an opportunity for improvement in the design and documentation of the measures proposed in ISO 9126-4.

Analysis of the Quality in Use Metrics

Effectiveness Metrics

- ❖ The three *Effectiveness Metrics* assess whether or not the task carried out by users achieved the specific goals with accuracy and completeness in a specific context of use.
 - ◆ **System of quantities for Effectiveness:**
 - ◆ **Base quantities:** 4 base metrics →
 1. Task Time
 2. Number of Tasks
 3. Number of Errors Made by the User
 4. Proportional value of each missing or incorrect

- ◆ The **first three of these base metrics** refer to terms in common use, but this leaves much to interpretation on what they mean.
 - ◆ Example 1: the term “task” has different meanings by different users which leads to different results
 - ◆ Example 2: the “error” term has two different definitions in ISO 9126-4 and IEEE Std. 610.12
- ◆ The **fourth base metric** has also different definitions contain a number of subjective assessments for which no repeatable procedure is provided.
- ◆ **Derived quantities:** 3 derived quantities →
 - 1.Task Effectiveness
 - 2.Task Completion
 - 3.Error Frequency

- ◆ The proposed **three Effectiveness Metrics**, which are defined as a prescribed combination of these base quantities, are therefore derived quantities.
- ◆ The **ranges** of the results obtained from implementing the corresponding measurement function are:
 - ◆ Task Effectiveness → between 0 and 1.
 - ◆ Task Completion → between 0 and 1.
 - ◆ Error Frequency → positive integer.
- ◆ These quantities **inherit the weaknesses** of the base quantities of which they are composed.

◆ Dimension of a quantity for Effectiveness

- ◆ Emerson states that the concept of dimension is particularly applicable to the derived quantities.
- ◆ task effectiveness and task completion, can have values between 0 and 1, and would be considered as dimensionless quantities, since a ratio of quantities with the same dimensions is itself dimensionless.

◆ Units of measurement for Effectiveness

◆ Base Units:

- ◆ Only the “**task time**” has an internationally recognized standard base unit (the second, or a multiple of this unit).
- ◆ The **next two base units** (tasks and errors) do not refer to any IS of measurement, and must be locally defined (which means that they fit poorly, for comparison purposes, when measured by different people)

- ◆ The **fourth base quantity**, proportional value of each missing or incorrect component, is puzzling because it is based on a given weighted value (number) and has no measurement unit.

- ◆ **Derived Units:**

- ◆ The “**task effectiveness**” leads to a derived unit that depends on a given weight. Therefore, like the base unit, its derived unit of measurement is **unclear**.
- ◆ The “**task completion**” is computed by dividing two base quantities (task/task) with the same unit of measurement.
- ◆ The definition of the “**error frequency**” provides two distinct alternatives for the elements of this computation.
 - ◆ This can lead to **two distinct interpretations**, i.e. **errors/task** or **errors/second**. this gives the possibility of **misinterpretation** and **misuse of measurement results** when combined with other units: for example, measures in centimeters and measures in inches cannot be added or multiplied.

◆ Value of a quantity for Effectiveness

- ◆ The four types of metrology values of a quantity are: true value, conventional true value, numerical value and conventional reference scale.
 - ◆ Numerical values are obtained for each base quantity based on the defined data collection procedure.
 - ◆ True values would depend on locally defined and rigorously applied measurement procedures for both “task completion” and “error frequency”. For “task effectiveness”, anyone would be hard pressed to figure out both a true value and a conventional true value.
 - ◆ Only “task time” refers to a conventional reference scale, that is, the international standard-etalon for time, from which the second is derived. None of the other base quantities in these effectiveness metrics refers to a conventional reference scale, or to a locally defined one

Analysis of the Quality in Use Metrics

Productivity Metrics

- ❖ The five productivity metrics assess the resources that users consume in relation to the effectiveness achieved in a specific context of use. In this standard, the time required to complete a task is considered to be the main resource to take into account
 - ◆ **System of quantities for Effectiveness:**
 - ◆ One of the five proposed productivity metrics is a **base quantity** (task time) while the other **four are derived quantities** (task efficiency, economic productivity, productive portion and relative user efficiency).
 - ◆ These derived quantities are themselves based on **five base quantities**: task time, cost of the task, help time, error time and search time.

◆ Dimension of a quantity for Productivity

- ◆ All the productivity metrics, except task time, are dimensionless quantities

◆ Units of measurement for Effectiveness

- ◆ There are five base units and no explicit derived units.
- ◆ the measurement unit for “task effectiveness” is not completely clear, since it depends on an ill-defined “given weight”

$$\begin{aligned} \text{'task efficiency' unit} &= \frac{\text{'task effectiveness' unit}}{\text{second}} \\ &= \frac{1 - \text{'a given weight' unit}}{\text{second}} = \frac{?}{\text{second}} \end{aligned}$$

1

- ♦ The measurement **unit** of “**economic productivity**” depends on the measurement **unit** of “**task effectiveness**”, which is unknown.

$$\begin{aligned}
 \text{'economic productivity' unit} &= \frac{\text{'task effectiveness' unit}}{\text{currency unit}} \\
 &= \frac{1 - \text{'a given weight' unit}}{\text{currency unit}} \\
 &= \frac{?}{\text{currency unit}}
 \end{aligned}$$

2

- ♦ The “**productive proportion**” has the same measurement unit in both the numerator and the denominator, the result is a percentage:

$$\text{'productive proportion' unit} = \frac{\text{second}}{\text{second}}$$

3

- ♦ The “relative user efficiency” also has no measurement unit, the result of this derived quantity is also a percentage.

$$\begin{aligned}
 \text{'relative user efficiency' unit} &= \frac{\text{'task efficiency' unit}}{\text{'task efficiency' unit}} \\
 &= \frac{\frac{\text{'task effectiveness' unit}}{\text{second}}}{\frac{\text{'task effectiveness' unit}}{\text{second}}} \\
 &= \frac{1 - \frac{\text{'a given weight' unit}}{\text{second}}}{1 - \frac{\text{'a given weight' unit}}{\text{second}}} \\
 &= \frac{\frac{?}{\text{second}}}{\frac{?}{\text{second}}}.
 \end{aligned}$$

4

Analysis of the Quality in Use Metrics

Safety Metrics

- ❖ The **safety metrics** claim to assess the level of risk of harm to people, businesses, software, property or the environment in a specific context of use.
- ❖ **Four derived quantities** must be quantified to evaluate the safety characteristics of a software product: **user health and safety, software damage, economic damage** and **the safety of people affected by use of the system**.
- ❖ Each of these derived quantities is the result of a computational formula, which consists of a combination of pre-collected base quantities: **number of usage situations, number of people, number of occurrences of software corruption, number of occurrences of economic corruption** and **number of users**.

- ❖ It can be observed that the resulting values of all the derived quantities should be **between 0 and 1**.
- ❖ All the safety metrics are **dimensionless quantities**.
- ❖ There are **five base units** and **two derived units** for these quantities.
- ❖ **Two of the derived quantities** have no measurement units: **user health and safety** and **safety of people affected by use of the system**.
- ❖ **none** of the measurement units has a **symbol**.

Analysis of the Quality in Use Metrics

Satisfaction Metrics

- ❖ The **satisfaction metrics** claim to assess the user's attitudes towards the use of the product in a specific context of use.
- ❖ **All three proposed satisfaction metrics are derived quantities: satisfaction scale, satisfaction questionnaire and discretionary usage.**
- ❖ These derived metrics depend on **four base quantities: population average, number of responses, number of times that specific software function / application / systems are used and number of times that specific software function/application/systems are intended to be used.**
- ❖ **Two of the proposed satisfaction metrics are dimensionless quantities: satisfaction questionnaire and discretionary usage.**

- ❖ It can be observed that the resulting values of all the derived quantities should be **between 0 and 1**.
- ❖ All the safety metrics are **dimensionless quantities**.
- ❖ There are **five base units** and **two derived units** for these quantities.
- ❖ **Two of the derived quantities** have no measurement units: **user health and safety** and **safety of people affected by use of the system**.
- ❖ **None** of the measurement units has a **symbol**.
- ❖ There are **four base units** and **no derived units**.
- ❖ The measurement unit, **satisfaction scale**, is not clear, since it depends on a **“questionnaire producing psychometric scales”**:

$$\text{'satisfaction scale' unit} = \frac{\text{psychometric scale unit}}{\text{people}}$$

5

Conclusion and Suggestions

- ❖ This paper has presented an analysis of the ISO 9126-4 Technical Report on quality in use metrics, and has investigated the extent to which it addresses the metrology criteria found in classic measurement. Based on the analysis in this paper, the following comments and suggestions can be made:
 - ◆ Identifying and classifying the quality in use metrics into base and derived quantities makes it easy to determine which should be collected (base quantities) to be used subsequently in computing the other (derived) quantities.
 - ◆ Based on equations (1) and (3 to 5), some of the derived units are ambiguous, since they depend on other quantities with unknown units.

- ◆ None of the quality in use metrics refers to any system of units, coherent (derived) unit, coherent system of units, international system of units (SI), off-system units, multiple of a unit, submultiple of a unit, true values, conventional true values or numerical values.
- ◆ None of the base and derived quantities, except for task time, has symbols for their measurement units.
- ◆ It is to be noted that the ranges of the results of many of the derived metrics in ISO 9126-4 are between 1 and 0.
 - ◆ Therefore, it is easy to convert them to percentage values. However, from our point of view, these results will be easier to understand if they are ranked in terms of qualitative values.
- ❖ The analysis methodology developed to investigate ISO TR9126-4 could also be of use to analyze the metrological strengths and weaknesses of close to 120 metrics proposed by the ISO in TRs 9126-2 and -3.



www.gelog.etsmtl.ca

***Thank
You !***



Questions?



alain.abran@etsmtl.ca

rafa.al-qutaish.1@ens.etsmtl.ca

jjcg@uah.es

