

Software Metrics & Software Metrology

Alain Abran

Chapter 10
Analysis of Quality Models and Measures
in ISO 9126

Agenda

This chapter covers:

- Introduction to ISO 9126
- The analysis models in ISO 9126 as standard reference models.
- The metrology related part of ISO 9126:
 - the base and derived measures.
- Example 1: Analysis of the designs of the measurement of Effectiveness in ISO 9126.
- Example 2: Analysis of the design of the measurement of Maturity in ISO 9126
- Summary & an improvement strategy for ISO 9126.

Agenda

This chapter covers:

- Introduction to ISO 9126
- The analysis models in ISO 9126 as standard reference models.
- The metrology related part of ISO 9126:
 - the base and derived measures.
- Example 1: Analysis of the designs of the measurement of Effectiveness in ISO 9126.
- Example 2: Analysis of the design of the measurement of Maturity in ISO 9126
- Summary & an improvement strategy for ISO 9126.



Introduction to ISO 9126

- 2001: first international consensus on the terminology for the quality characteristics for software product evaluation:
 - ISO 9126 – *Software Product Evaluation - Quality Characteristics and Guidelines for their Use.*

- These ISO 9126 quality models provide by far the most comprehensive coverage in software engineering on software quality:
 - Both national and international standards experts have invested a great deal of energy over the past 2 decades to improve them, and are still working to improve them further.
 - These ISO quality models are considerably more detailed than the quality models proposed since the 1980s by individual authors

Introduction to ISO 9126

- ISO 9126 series consists of:
 - 1 International Standard (ISO 9126 Part 1):
 - the structure of the ISO quality models for software products

 - 3 Technical Reports:
 - Part **2** proposes a list of derived measures of **external** software quality.
 - Part **3** proposes a list of derived measures of **internal** software quality.
 - Part **4** proposes a list of derived measures for the **quality in use** model.

Introduction to ISO 9126

- ISO 9126 is complemented by a set of guides in the ISO 14598 series:
 - Developers Guide
 - Evaluators Guide
 - Evaluation Modules, etc.

- ISO itself is putting a great deal of effort into further improvements.
 - For instance: within the 2-3 years, the ISO expects to issue a major update of this series:
 - Relabeled as the ISO 25000 series on software product quality.

Agenda

This chapter covers:

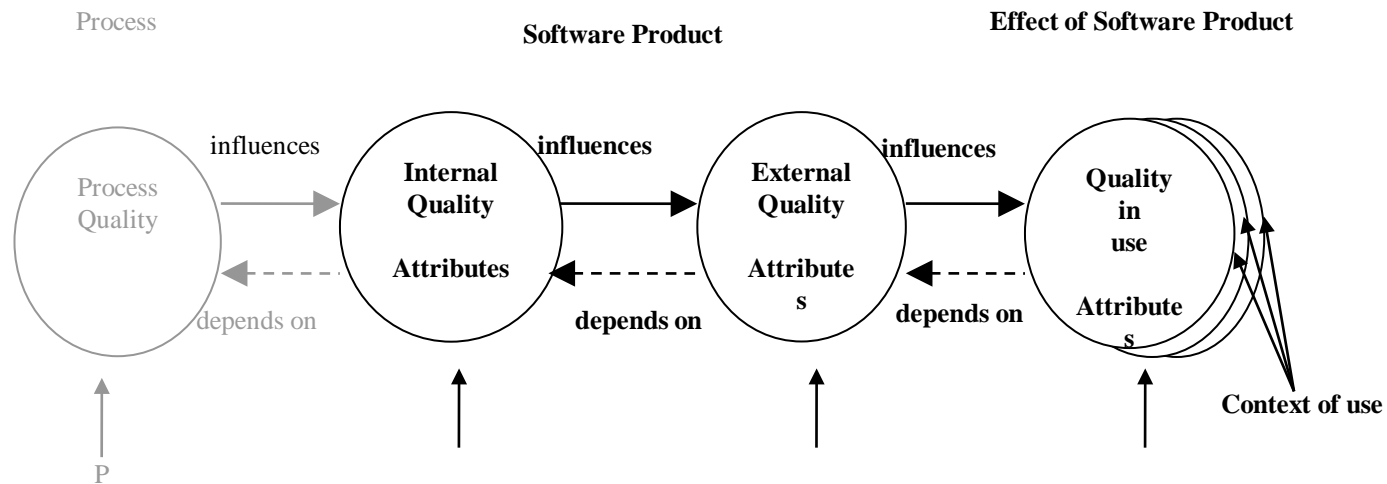
- Introduction to ISO 9126
- **The analysis models in ISO 9126 as standard reference models.**
- The metrology related part of ISO 9126:
 - the base and derived measures.
- Example 1: Analysis of the designs of the measurement of Effectiveness in ISO 9126.
- Example 2: Analysis of the design of the measurement of Maturity in ISO 9126
- Summary & an improvement strategy for ISO 9126.



Analysis Models of ISO 9126: The (quantitative) models

■ A- The Standard Reference Model in ISO 9126

- This reference model includes 3 views of the quality of a software product:
 - Internal quality of the software
 - External quality of the software
 - Quality in use of the software



Analysis Models of ISO 9126: The (quantitative) models

- ISO 9126-1 presents 2 structures of quality models for software product quality:
 - a 1st structure for both the internal and external quality models. It includes:
 - 6 characteristics,
 - subdivided into 27 sub characteristics for internal and external quality.
 - a 2nd structure for the quality in use model:
 - 4 'quality in use' characteristics: effectiveness, productivity, safety, and satisfaction.

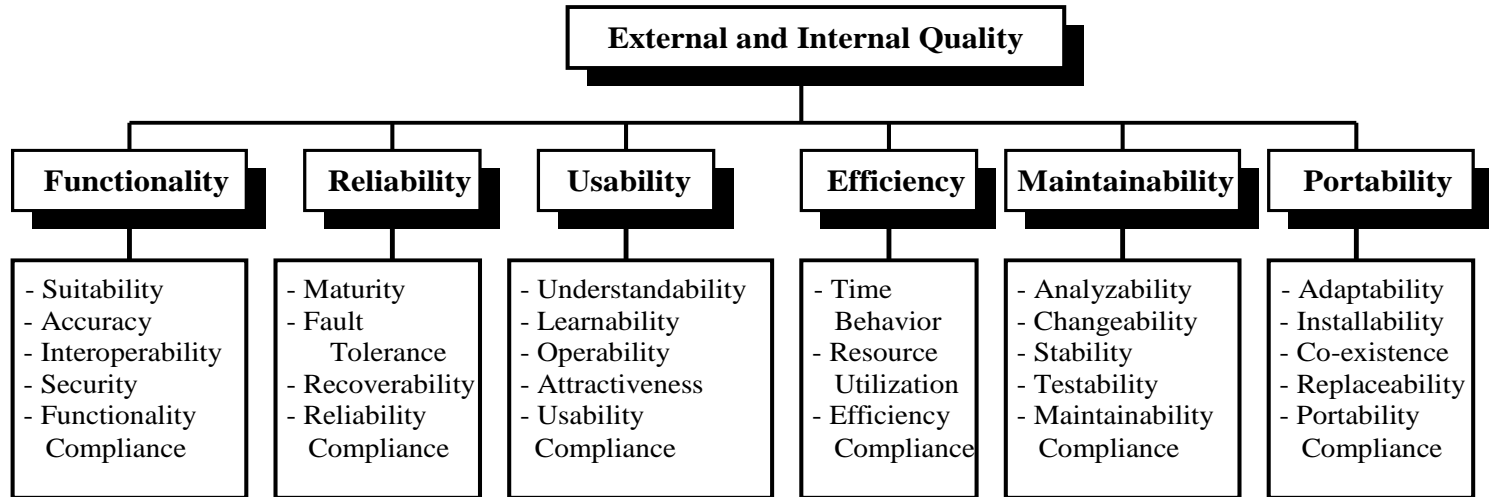


Figure 2: Quality model for External and Internal Quality: characteristics and sub-characteristics – ISO 9126-1

Analysis Models of ISO 9126: The (quantitative) models

B - An Organizational Reference Context model: Interpretation of ISO 9126 for Evaluation & Decision Making

- For the application of the ISO quality models to specific software within an organization, an Organizational Reference Context would typically be set up and used.
 - How to do this is described in the ISO 14598 series, from various perspectives (developers, third party, etc.).
- The application of this analysis model (which corresponds to evaluation and decision making on the basis of decision criteria in ISO 9126) is usually performed as a 4-step quantification process:
 1. Identification of quality-related requirements:
 - selection of the parts of the ISO quality models that are relevant to a particular context of quality evaluation.

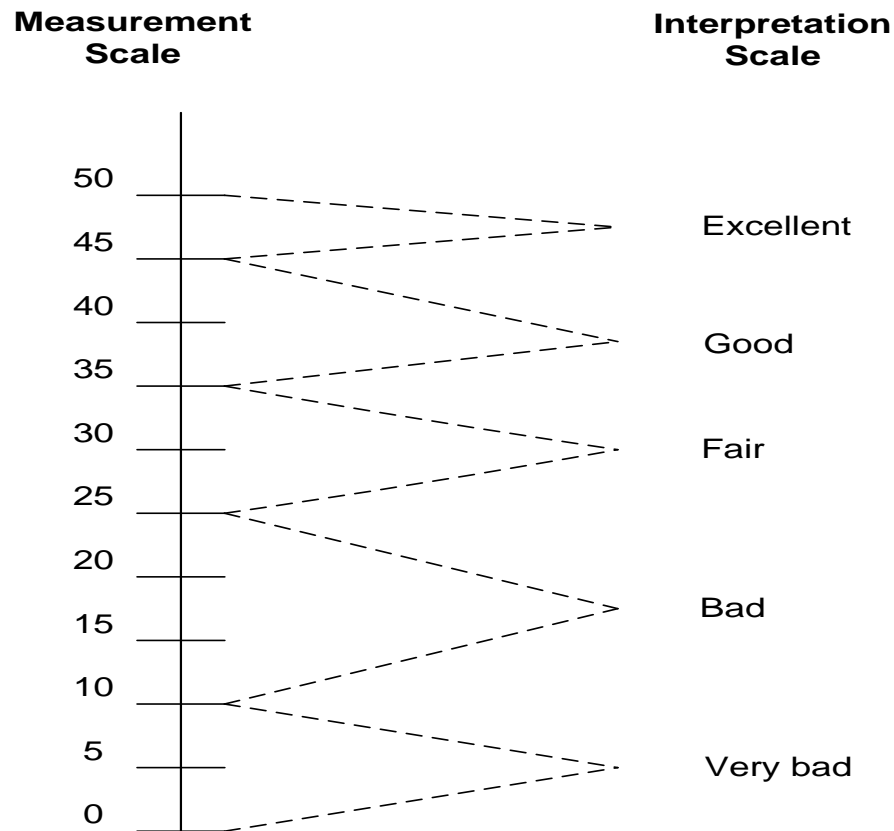
Analysis Models of ISO 9126: The (quantitative) models

2. Identification of the context of interpretation, that is:
 - the selection of reference values, such values being either generic or specific threshold values, or
 - the determination of targets specified for a particular context.
3. Use of the derived measures from the data preparation phase to fill out the instantiated quality model determined in step 1.
4. Comparison of the results of step 3 with either the set of reference values or the targets determined in step 2 to take a decision based on both the information provided and whatever relevant information is available to the decision maker.

Analysis Models of ISO 9126: The (quantitative) models

■ This is illustrated below with:

- reference values for the decision criteria (or evaluation criteria) for this **organizational context**
- target values (or an evaluation scale) **for the specific software to be measured**



Analysis Models of ISO 9126: The (quantitative) models

- For the set of relationships over the set of objects of interest for the Information Needs (see ISO 19593 & Chapter 4), the Analysis Model would typically either:
 - A. quantify a relationship which is well understood, well described over time, and for which there is a large consensus, or
 - B. attempt to 'quantify' a relationship (i.e. a concept) for which it is not yet well known how to capture it within a single measurement dimension and a single (base or derived) measure (with its corresponding single measurement unit, or set of such units).


Analysis Models of ISO 9126: The (quantitative) models

- To use in practice any one of such relationships described textually in ISO 9126 represents an ‘attempt to quantify’ without a prescribed standard or organizational reference context nor empirically verified foundation [Abran 2005a, 2005b]
 - Neither of ISO 9126 and ISO 14598 (and the upcoming ISO 25000 series) proposes specific ‘reference models of analysis’ nor an inventory of ‘organizational reference contexts’ with reference values and decision criteria
 - Each organization has to define its own, thereby limiting possibilities of comparisons industry-specific wide or industry-wide.

- Abran, A.; Al-Qutaish, R.E.; Desharnais, J.M.; Habra, N. (2005a) ‘An Information Model for Software Quality Measurement with ISO Standards’, International Conference on Software Development - SWEDC-REK, Reykjavik, Iceland, University of Iceland, pp. 104-116
- Abran, A.; Al-Qutaish, R.E.; Desharnais, J.M. (2005b) ‘Harmonization Issues in the Updating of the ISO Standards on Software Product Quality’, in Metrics News, vol. 10, no. 2, December, 2005, pp. 35-44

Agenda

This chapter covers:

- Introduction to ISO 9126
 - The analysis models in ISO 9126 as standard reference models.
 - **The metrology related part of ISO 9126:**
 - **the base and derived measures.**
 - Example 1: Analysis of the designs of the measurement of Effectiveness in ISO 9126.
 - **Example 2: Analysis of the design of the measurement of Maturity in ISO 9126**
 - Summary & an improvement strategy for ISO 9126.
- 

The Metrology-related part of ISO 9126: base and derived measures

- The implementation of analysis models, including the ones from ISO 9126, has to be based on the data collection of base measures (& derived ones, where relevant).
- The measures available for the data collection for the ISO 9126 quality models are proposed and described in the 3 technical reports in the ISO 9126 series which propose:
 - an inventory of +250 derived measures for each quality characteristic or sub characteristic,
 - 80 base measures (and corresponding 80 attributes) which are used to build the above +250 derived measures,
 - explanations of how to apply and use them, and
 - some examples of how to apply them during the software product life cycle.

A key question, then, is:

- What is the quality, from a measurement perspective, of this extensive list of 250+ proposed measures?

The Metrology-related part of ISO 9126: base and derived measures

Derived Measures in ISO 9126 – 2001+

Each of the over 250 derived measures proposed in ISO 9126 is typically defined as a formula composed of base measures.

An example from ISO 9126 Part 4: Quality in use:

- Base Measure 1 (B1): Number of detected failures.
- Base Measure 2 (B2): Number of test cases performed.
- Derived Measure: $B1 / B2$ with the following measurement units:

$$\frac{B1}{B2} = \frac{\text{Number of detected failures}}{\text{Number of test cases performed}}$$

In this example from ISO 9126, the derived measure (with the corresponding measurement units above) is assigned the following name: “**Failure density against test cases.**”

However, while the numerical assignment rules (that is, the formula for the derived measures) for each derived measure are described as mathematical operations (here, a division), neither B1 nor B2 for these operations is described with sufficient clarity to ensure the quality (accuracy, repeatability, and reproducibility) of the measurement results of:

- these base measures, and, of course,
- the derived measures built upon them.

In general, this is a fairly weak foundation from a measurement perspective, and both practitioners and researchers should be very careful in the implementation, use, and interpretation of the numbers that result from the use of these ISO quantitative models.

Notwithstanding the above comments, these ISO documents remain at the forefront of the state of the art on software measurement, and efforts are under way to improve them.

Analysis of Derived Measures

■ Analysis of the derived measures in ISO 9126-4: Quality in use

- The ISO 9126-4 technical report on the measures proposed for the ISO **Quality in use** model is used to illustrate some of the metrology-related issues in the ISO 9126 series in the late 2000s.
 - Many of the measurement issues raised with respect to ISO Part 4 would also apply to Parts 2 and 3.
- In ISO 9126-4, 15 derived measures are proposed for the 4 quality characteristics of the ISO Quality in use model – see next table.

Analysis of Derived Measures

| Quality Characteristic | Derived Measures |
|-------------------------------|---|
| Effectiveness | <ul style="list-style-type: none">-task effectiveness-task completion-error frequency |
| Productivity | <ul style="list-style-type: none">-task time-task efficiency-economic productivity-productive proportion-relative user efficiency |
| Safety | <ul style="list-style-type: none">-user health and safety-safety of people affected by use of the system-economic damage-software damage |
| Satisfaction | <ul style="list-style-type: none">-satisfaction scale-satisfaction questionnaire-discretionary usage |

Derived Measures in ISO 9124-4: Quality in use


Analysis of Derived Measures

■ Objective of the analysis:

- Identify the measurement concepts not tackled in ISO 9126:
 - i.e. their gaps in their measurement designs.
 - Each of these gaps in the design of the derived measures represents an opportunity to improve the measures in the upcoming ISO update, which is the ISO 25000 series.
 - This analysis provides an illustration of the improvements that are needed to many of the software measures proposed to the industry.

Agenda

This chapter covers:

- Introduction to ISO 9126
- The analysis models in ISO 9126 as standard reference models.
- The metrology related part of ISO 9126:
 - the base and derived measures.
- **Example 1: Analysis of the designs of the measurement of Effectiveness in ISO 9126.** 
- Example 2: Analysis of the design of the measurement of Maturity in ISO 9126
- Summary & an improvement strategy for ISO 9126.

Example 1: Analysis of a Derived Measure - Effectiveness

■ Analysis of the measurement of Effectiveness in ISO 9126-4

- 9126-4 claims that the proposed 3 measures for the Effectiveness characteristic assess whether or not the task carried out by users achieved the specific goals with accuracy and completeness in a specific context of use.
- Issues identified within:
 - the base measures proposed,
 - the derived measures,
 - the measurement units,
 - the measurement units of the derived quantities, and
 - the value of a quantity for Effectiveness.

Example 1: Analysis of a Derived Measure - Effectiveness

■ Identification of the base measure of Effectiveness

- The 3 Effectiveness-derived measures (task effectiveness, task completion, and error frequency) come from a computation of 4 base measures, which are themselves collected/measured directly, namely:
 - task time,
 - number of tasks,
 - number of errors made by the user, and
 - proportional value of each missing or incorrect component.

Example 1: Analysis of a Derived Measure - Effectiveness

■ A- The first 3 base measures:

- refer to terms in common use (i.e. task time, # of tasks, and # of errors made by the user)
- but this leaves much to interpretation on what constitutes, for example, a task.

■ ISO 9124-4 does not provide a detailed measurement-related definition for any of these base measures:

- In summary, it does not provide assurance that the measurement results are repeatable and reproducible across measurers or across groups measuring the same software, or across organizations either, where a task might be interpreted differently and with different levels of granularity.
 - This leeway in the interpretation of these base measures makes a rather weak basis for both internal and external benchmarking.

Example 1: Analysis of a Derived Measure - Effectiveness

B- The base measure, # of errors made by the user, is defined in Appendix F of ISO TR 9126-4 as:

- an “instance where test participants did not complete the task successfully, or had to attempt portions of the task more than once.”
- This definition diverges significantly from the one in the IEEE Standard Glossary of Software Engineering Terminology, where:
 - the term “error” has been defined as “the difference between a computed, observed, or measured value or condition and the true, specified, or theoretically correct value or condition.”
 - Example: a difference of 30 meters between a computed result and the correct result.

Example 1: Analysis of a Derived Measure - Effectiveness

C- The base measure, the “proportional value of each missing or incorrect component” in the task output is based, in turn, on another definition, whereas:

- each “potential missing or incorrect component” is given a weighted value A_i based on the extent to which it detracts from the value of the output to the business or user.
- This embedded definition itself contains a number of subjective assessments for which no repeatable procedure is provided:
 - the value of the output to the business or user,
 - the extent to which it detracts from that value,
 - the components of a task, and
 - potential missing or incorrect components.

Example 1: Analysis of a Derived Measure - Effectiveness

■ **The derived measures of Effectiveness**

- The proposed 3 derived measures for the Effectiveness characteristic, which are defined as a prescribed combination of base measures, inherit the weaknesses of the base measures of which they are composed

■ In summary:

- there is no assurance that the measurement results of the derived measures are repeatable and reproducible across measurers, across groups measuring the same software, or across organizations either, where a task might be interpreted differently and with different levels of granularity.

Example 1: Analysis of a Derived Measure - Effectiveness

■ The measurement units of the base measures

- Of the 4 base measures, only 'task time', has:
 - an internationally recognized standard measurement unit:
 - the second, or a multiple of this unit
 - a universally recognized corresponding symbol:
 - 's' for the second as a measure of time.

Example 1: Analysis of a Derived Measure - Effectiveness

- The next 2 base measures (tasks and errors) do not refer to any international standard of measurement and must be locally defined. This means that:
 - They are not reliably comparable across organizations.
 - They are also not reliably comparable within a single organization when measured by different people, unless local measurement protocols (i.e. measurement procedure) have been clearly documented and rigorously implemented.

- The last base measure (proportional value of each missing or incorrect component) is puzzling:
 - it is based on a given weighted value (number), and
 - it has no measurement unit.

Example 1: Analysis of a Derived Measure - Effectiveness

■ Measurement units of the derived measures

- Task effectiveness: In ISO 9126-4, this derived unit that depends on a given weight:

$$\text{Task effectiveness} = 1 - (\text{a given weight})$$

- Therefore, its derived unit of measurement is unclear and undefined.
- Task completion: The derived measure is computed by dividing one base measure by the other (task/task) with the same unit of measurement. The measurement results is a percentage.
- Error frequency: The definition of the computation of this derived measure provides 2 distinct alternatives for the elements of this computation. This can lead to 2 distinct interpretations:
 - errors/task, or
 - errors/second.

Example 1: Analysis of a Derived Measure - Effectiveness

Of course,

- this, in turn, leads to 2 distinct derived measures as a result of implementing 2 different measurement functions (formulae) for this same derived measure;
- and leaves open 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 directly added or multiplied.
- In software measurement, who cares about this mixing of units?
 - Should you care as a software manager?
 - Should you care as a software engineer?

Example 1: Analysis of a Derived Measure - Effectiveness

■ Value of a quantity for Effectiveness

- The 4 types of metrology values of a quantity are:
 - true value,
 - conventional true value,
 - numerical value, and
 - conventional reference scale.
- In the measurement of Effectiveness with ISO 9126-4, for each base measure numerical values are obtained on the basis of the defined data collection procedure:
 - For each derived measure, numerical values are obtained by applying their respective measurement function:

Example 1: Analysis of a Derived Measure - Effectiveness

- For instance, the derived measures task effectiveness and task completion are expressed as %, and are interpreted as the effectiveness and completion of a specific task respectively.
 - For task effectiveness in particular, it would be difficult to figure out both a true value and a conventional true value.
 - For task completion and error frequency, the true values would depend on locally defined and rigorously applied measurement procedures, but without reference to universally recognized conventional true values (as they are locally defined).


Example 1: Analysis of a Derived Measure - Effectiveness

Finally, in terms of the metrological values:

- 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 measures in these derived measures of effectiveness refers to a conventional reference scale, or to a locally defined one.

Agenda

This chapter covers:

- Introduction to ISO 9126
 - The analysis models in ISO 9126 as standard reference models.
 - The metrology related part of ISO 9126:
 - the base and derived measures.
 - Example 1: Analysis of the designs of the measurement of Effectiveness in ISO 9126.
 - **Example 2: Analysis of the design of the measurement of Maturity in ISO 9126**
 - Summary & an improvement strategy for ISO 9126.
- 

Example 2: Analysis of a Derived Measure - Maturity

- In ISO 9126, there is an implicit claim that the proposed (base or derived) measures capture the intended concept to be quantified:
 - However, there is no explicit statement about how much each of the proposed measure quantitatively captures the concept intended to be quantified.
 - At best, each proposed measure in ISO 9126 is considered as an unspecified contributor to such a quantification for evaluation and decision-making. Put differently, they are:
 - not considered as a contributor to concept measurement in terms of a specific scale and a standard etalon,
 - but only as a contributor (in some proportion) to its quantification (without an agreed-upon and widely recognized reference scale and a standard etalon).

Example 2: Analysis of a Derived Measure - Maturity

In other words,

- in the ISO 9126 analysis models, the practical solution (when a decision criterion cannot be measured directly) is to quantify through an association or mapping (direct or indirect), with, it is hoped, better understood concepts that have been quantified.
- An example is presented next with some of the ISO 9126 base measures combined into derived measures which are associated with other concepts (i.e. the quality characteristics and sub-characteristics within the ISO model of software product quality).

Example 2: Analysis of a Derived Measure - Maturity

- This section illustrates the missing links between the quantitative models proposed in ISO 9126 and the detailed measures proposed as the contributors to this quantification.
 - Example selected: the **Maturity** sub-characteristic and of the 7 derived measures proposed for this sub-characteristic.
- For the **Maturity** sub-characteristic:
 - The relationships between the attribute being measured by the derived measure ‘failure density against test cases’ & the ‘maturity’ sub-characteristic and the ‘reliability’ characteristic are not identified, neither described.
 - No model of such relationships is proposed either.

Example 2: Analysis of a Derived Measure - Maturity

■ Analysis of the Measurement of 'Maturity'

- The model proposed by ISO to quantify the external quality of a software product contains 6 quality characteristics and 27 sub-characteristics:
 - One of the 6 quality characteristics is 'Reliability' which is itself composed of 4 sub-characteristics.
 - One of these 4 sub-characteristics is 'Maturity' which is chosen here as an example.

Example 2: Analysis of a Derived Measure - Maturity

- To quantify this single **Maturity** sub-characteristic, ISO 9126-3 proposes a list of 7 distinct derived measures:

1. Failure density against test cases
2. Estimated latent fault density
3. Fault Density
4. Fault removal
5. Mean time between failures (MTBF)
6. Test coverage
7. Test maturity

■ Related ISO 9126 Definitions

Reliability:

The capability of the software product to maintain a specified level of performance when used under specified conditions

Maturity:

The capability of the software product to avoid failure as a result of faults in the software.

Example 2: Analysis of a Derived Measure - Maturity

A- Metrology perspective

- Each of these 7 derived measures is presented only at a fairly abstract level as formulae composed from a set of base measures, themselves lacking detailed descriptions, including the attributes they are attempting to measure.
 - This leaves each of them highly susceptible to individual interpretation: neither the base measures for these operations, nor the corresponding attributes, have been described with sufficient clarity to ensure the quality of the measurement results:
 - They are not documented at a detailed enough level to provide sufficient guidance to ensure the accuracy, repeatability, and repetitiveness of measurement results, in the event that the same software is measured by different measurers, which in turn leads to potentially significantly different values.

Example 2: Analysis of a Derived Measure - Maturity

- Each one of the 7 proposed derived measures are described individually as illustrated in the side box with the 'Failure density against test cases' as an example.

- **Example: 'Failure density against test cases'**

Purpose of this derived measure in ISO 9126:

- How many failures were detected during defined trial period?
- Method of application for this derived measure:
 - Count the number of detected failures and performed test cases.
- However, none of the embedded base measures are defined precisely in ISO 9126, including failure and test cases.

Example 2: Analysis of a Derived Measure - Maturity

B- Perspective of the Analysis Model in ISO 9126

- There is no ISO defined context or values for such context. Each group within each organization has to build its own set of values for analysis within a specific context.
 - What is the specific contribution of any one of the above 7 derived measures to the Maturity sub-characteristic?
 - Are there some overlaps across the relationship of any of these 7 derived measures, and if so, to what extent?
 - If not all 7 derived measures are mandatory-necessary, which one or which ones are the most representative of the Maturity sub-characteristic, and if so, to which extent?

Example 2: Analysis of a Derived Measure - Maturity

- We now ask how well the derived measure, ‘Failure density against test cases’, fully describes (i.e. quantifies) the hierarchy of concepts of the ISO 9126 model of software product quality:
 - How much of the Maturity sub-characteristic is captured by it.
 - In practice, the derived measure, Failure density against test cases, is only a contributor, that is, an indicator within that piece of the hierarchy of quality-related concepts.
 - How can we quantify this contribution?
 - The answer is not documented in ISO 9126-4, or even identified as an issue.

Example 2: Analysis of a Derived Measure - Maturity

Example: A derived measure proposed for quantifying Maturity in ISO 9126-2: Failure density against test cases

| Step (ISO 15939) | Description |
|---------------------|--|
| Data Collection: | <ul style="list-style-type: none">• Base Measure 1 (B1): Number of detected failures.• Base Measure 2 (B1): Number of test cases performed. |
| Data Preparation | <ul style="list-style-type: none">• Algorithm of Derived Measure: $B1 / B2$• Name of Derived Measure: Failure density against test cases. |
| Analysis | <p>The derived measure, Failure density against test cases, is associated * with (mapped to):</p> <ul style="list-style-type: none">• The quality sub-characteristic: Maturity• Within the quality characteristic: Reliability• Within the quality sub-model: External Quality – ISO 9126-2 <p>* Associated: there is not an explicit description in ISO 9126 of how much of the proposed measure quantitatively captures the concept intended to be quantified.</p> |


Example 2: Analysis of a Derived Measure - Maturity

■ C- Missing links

- None of the expected links between this (weak) metrology basis for the measurement of the base and derived attributes and the quantification of the quality sub-characteristic (eg. Maturity) and characteristic (e.g. Reliability) is described in ISO 9126.
 - Hopefully, such links will be described but it will take years of research and development to gain insights into this problem and to develop analysis models based on solid empirical evidence

Agenda

This chapter covers:

- Introduction to ISO 9126
 - The analysis models in ISO 9126 as standard reference models.
 - The metrology related part of ISO 9126:
 - the base and derived measures.
 - Example 1: Analysis of the designs of the measurement of Effectiveness in ISO 9126.
 - Example 2: Analysis of the design of the measurement of Maturity in ISO 9126
 - **Summary & an improvement strategy for ISO 9126.**
- 

Summary and Improvement Strategy for ISO 9126

- As has been illustrated, the base measures and their current lack of adherence to metrology principles and characteristics constitute one of the major impediments to the overall applicability and deployment of ISO 9126.
- It is also recognized that, to properly analyze, verify, and correct the design of all 80 base measures will require considerable time and work on the part of a single, isolated research team.
 - To tackle this challenge, a larger, multi-group research team has to be set up, its members to work concurrently on this issue, including research groups at universities across the world, all of whom in coordination with the work in progress of ISO/IEC JTC1/SC7 Working Group 6.

Summary and Improvement Strategy for ISO 9126

- In particular, improvement work should focus on 5 of the 80 attributes in ISO 9126 which appear in more than 10 derived measures; that is, function (38 occurrences), duration (26), task (18), case (16), and failure (11) – See the Advanced Readings 2 section.
- Work on the detailed design of the base measures and on the definitions of the attributes should leverage relevant measurement definitions from other international standards wherever possible.
 - , the ISO 9Even definitions from existing standards still need further refinement to facilitate their use in operational procedures from a measurement viewpoint.
 - Finally126 standard also includes a number of qualifiers of the base measures which will require further clarification from a measurement viewpoint.

Summary and Improvement Strategy for ISO 9126

- In conclusion, much work remains to be done:
 - To define the base measures in detail, even those identified as requiring priority attention.
 - To identify the links across the quality models and their corresponding base and derived measures proposed in ISO 9126.