华东师范大学软件学院课程作业

课程名称：软件质量分析	年级：2023 级本科	姓名：张梓卫
作业主题：ISO 9126 翻译	学号：10235101526	实验日期：
指导老师：	组号：	

目录

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 that were outstanding 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.

ISO 9126-4 技术报告提出的度量，可用于阐释 ISO 9126 系列在 2000 年代后期尚未解决的一些与度量相关问题。ISO 第 4 部分中提出的许多问题同样应用于第 2 部分和第 3 部分。

In ISO 9126-4, 15 derived measures are proposed for the 4 quality characteristics of the ISO Quality in Use model —see Table 10.1.

在 ISO 9126-4 中，15 个派生度量被在 ISO "使用质量"模型的 4 个质量特征提出——见表 10.1。

The objective of the analysis is to identify the measurement concepts that were not tackled in the ISO 9126 series of documents, that is, their gaps in their measurement designs.

本分析的目标是理解并鉴别 ISO 9126 系列文档中未解决的度量概念，即度量设计中的欠缺与空白。

Table 10.1：Derived Measures in ISO 9124-4: Quality in Use

Quality Characteristic	**Derived Measures**
Effectiveness	— task effectiveness — task completion — error frequency
Productivity	— task time — task efficiency — economic productivity — productive proportion — relative user efficiency
Safety	— user health and safety — safety of people affected by use of the system — economic damage — software damage
Satisfaction	— satisfaction scale — satisfaction questionnaire — discretionary usage

- 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.

表 10.1 质量模型 ISO Quality in Use 中的派生度量

表 1: ISO 9124-4 中的派生度量：使用中的质量

质量特性	派生度量
有效性	— 任务有效性 — 任务完成度 — 错误频率
生产率	— 任务时间 — 任务效率 — 经济生产率 — 生产比例 — 相对用户效率
安全性	— 用户健康与安全 — 系统使用过程中受影响人员的安全 — 经济损失 — 软件损坏
满意度	— 满意度量表 — 满意度问卷 — 自主使用

- 每个在派生度量设计中的空白都代表着在即将发布的 ISO 更新（ISO 25000 系列）中改进度量的机会。
- 本次分析阐释了许多软件度量在工业界提出的改进需求。

10.4.2 Analysis of the Measurement of Effectiveness in ISO 9126

10.4.2 ISO 9126 中效率测量的分析

In ISO 9126 - 4 , it is claimed that the proposed three measures for the Effectiveness characteristic —see Table 10.1 — assess whether or not the task carried out by users achieved the specifi c goals with accuracy and completeness in a specifi c context of use.

This sub - section identifi es a number of issues with:

- 1. the base measures proposed,
- 2. the derived measures,
- 3. the measurement units,
- 4. the measurement units of the derived quantities, and
- 5. the value of a quantity for Effectiveness.

在 ISO 9126-4 中声称，有效性特征提出的三个度量标准为—（见表 10.1）—评估用户执行的任务是否在特定的使用场景下，以准确性和完整性达到了特定目标。

本小节指出了几个问题：

- 提出的基础度量，
- 派生度量，
- 度量单位，
- 派生量的度量单位，以及

- 有效性的数量值。

10.4.2.1 Identifi cation of the Base Measure of Effectiveness. The 3 Effectiveness - derived measures (task effectiveness, task completion, and error frequency) come from a computation of four base measures, which are them selves 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.

10.4.2.1 有效性基础度量的识别。有效性的 3 个派生度量（任务有效性、任务完成度和错误频率）来自对四个基础度量的计算，这些度量是直接收集/测量的，分别是：

- 任务时间，
- 任务数量，
- 用户犯错次数，以及
- 每个缺失或错误组件的比例值。

The first three base measures above refer to terms in common use (i.e. task time, number of tasks, and number of errors made by the user), but this leaves much to interpretation on what constitutes, for example, a task.

Currently, ISO 9124 does not provide a detailed measurement - related definition for any of them:

- In summary, it does not provide assurance that the measurement results are repeatable and reproducible across measurers or across groups measur ing 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.

上述前三个基础度量是常用术语（即任务时间、任务数量和用户犯错次数），但这给“任务”的定义留下了许多解释空间。目前，ISO 9124 并未为这些度量提供详细的与度量相关的定义：

- 总的来说，它未能保证测量结果在不同测量者或组织中对同一软件进行测量时具有可重复性和可再现性，不同组织对任务的理解可能有所不同，粒度层次也可能不同。
- 这种对这些基础度量的解释自由度，使其在内部和外部基准测试中的基础较为薄弱。

The third base measure, number 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 signifi cantly 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, specifi ed, or theoretically correct value or condition. For example: a difference of 30 meters between a computed result and the correct result.

第三个基础度量“用户犯错次数”在 ISO TR 9126-4 的附录 F 中被定义为“测试参与者未能成功完成任务，或必须多次尝试部分任务的实例。”

该定义与 IEEE 软件工程术语标准词汇表中的定义有显著不同，后者将“错误”定义为“计算、观察或测量值或条件与真实、指定或理论正确值或条件之间的差异”。

The fourth base measure, referred to as 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:

第四个基础度量，即任务输出中“每个缺失或错误组件的比例值”，基于另一种定义，即每个“潜在缺失或错误组件”根据其削弱输出对业务或用户价值的程度，赋予一个加权值。

该嵌入定义本身包含了许多主观评估，且没有提供可重复的操作程序：

- 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.
- 输出对业务或用户的价值，
- 它对该价值的削弱程度，
- 任务的组成部分，以及
- 潜在的缺失或错误组件。

10.4.2.2 The Derived Measures of Effectiveness. The proposed three derived measures for the Effectiveness characteristic, which are defined as a prescribed combination of the base measures mentioned above, 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 repeat able and reproducible across measurers, across groups measuring the same soft ware, or across organizations either, where a task might be interpreted differently and with different levels of granularity.

10.4.2.2 有效性的派生度量。对有效性特征提出的三个派生度量，由上述基础度量的组合定义，继承了组成它们的基础度量的弱点。总的来说，这些派生度量的测量结果在不同测量者、同一测量软件的不同群体或组织之间没有可重复性、复现性的保证，其中任务可能会以不同的粒度层次进行解释。

10.4.2.3 The Measurement Units of the Base Measures. Of the four base measures, a single one, i.e. task time, has:

10.4.2.3 基础度量的度量单位。在四个基础度量中，只有一个度量，即任务时间，具有：

- 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.
- 一个国际公认的标准度量单位：秒或其倍数；
- 一个普遍公认的符号：“s”表示时间的度量单位秒。

The next two base measures (tasks and errors) do not refer to any international standard of measurement and must be locally defined. This means that:

接下来的两个基础度量（任务和错误）没有任何国际标准度量，必须在本地定义。这意味着：

- 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.
- 它们在不同组织间无法进行可靠的比较。
- 它们在同一组织内也无法进行可靠的比较，除非清楚地记录并严格执行了本地的测量协议（即测量程序）。

The fourth base measure (proportional value of each missing or incorrect com ponent) is puzzling:

第四个基础度量（每个缺失或错误组件的比例值）令人困惑：

- it is based on a given weighted value (number), and

- it has no measurement unit.
- 它基于一个给定的加权值（数字），
- 它没有度量单位。

10.4.2.4 Measurement Units of the Derived Measures. *Task effective-ness*: In ISO 9126 - 4, this derived measure leads to a derived unit that depends on a given weight:

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

10.4.2.4 派生度量的度量单位。 任务有效性：在 ISO 9126-4 中，该派生度量的计算单位取决于给定的权重：

$$任务有效性 = 1 -（给定的权重）$$

Therefore, its derived unit of measurement is unclear and undefined.
因此，其派生度量单位不清楚，也未定义。

Task completion: The derived measure is computed by dividing one base measure by the other (task/task) with the same unit of measurement. The mea surement results is a percentage.

Error frequency: The definition of the computation of this derived measure provides two distinct alternatives for the elements of this computation. This can lead to two distinct interpretations:
任务完成度：该派生度量通过将一个基础度量除以另一个基础度量（任务/任务）来计算，其度量结果是一个百分比。
错误频率：该派生度量的计算定义提供了两种不同的计算要素替代方案，这可能导致两种不同的解释：

- errors/task, or
- errors/second.
- 错误/任务，或
- 错误/秒。

Of course, 当然，

- this, in turn, leads to two distinct derived measures as a result of implementing two 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.
- 这将导致两种不同的派生度量，因为对同一派生度量应用了两种不同的测量公式；
- 并且在与其他单位组合时，可能会导致测量结果的误解和误用。例如：厘米和英寸的度量不能直接相加或相乘。

```
In software measurement, who cares about this mixing of units?
  Should you care as a software manager?
  Should you care as a software engineer?
```

```
在软件评估里，谁在意这些部分的混合呢？
  作为一个软件管理工程师，你应该在意吗？
  作为一个软件工程师，你应该在意吗？
```

10.4.2.5 Value of a Quantity for Effectiveness. The five types of metrology values of a quantity are 2:

- Numerical quantity value
- Quantity - value scale
- Ordinal quantity - value scale

- Conventional reference scale
- Ordinal value

10.4.2.5 有效性数量的值。度量值的五种类型是：

- 数值量值
- 量值尺度
- 序数量值尺度
- 约定参考尺度
- 序数值

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:

在 ISO 9126-4 中，针对每个基础度量，依据定义的数据收集程序获得数值：

For each derived measure , numerical values are obtained by applying their respective measurement function. For instance, the derived measures task effectiveness and task completion are expressed as percentages, and are interpreted as the effectiveness and completion of a specifi c task respectively.

针对每个派生度量，通过应用其各自的测量函数获得数值。例如，任务有效性和任务完成度的派生度量以百分比表示，分别解释为特定任务的有效性和完成度。

- For task effectiveness in particular, it would be diffi cult to fi gure 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).
- 对于任务有效性，难以确定一个真实值和约定真实值。
- 对于任务完成度和错误频率，真实值将取决于本地定义并严格执行的测量程序，但不会参照国际公认的约定真实值（因为它们是本地定义的）。

Finally, in terms of the metrological values:

最后，关于度量值：

- 只有任务时间引用了一个约定的参考尺度，即国际标准（基准）时间单位秒。
- 在这些有效性的派生度量中，其他基础度量均未引用约定的参考尺度，或本地定义的尺度。

See the Advanced Readings section for an additional example.

请参阅“高级阅读”部分，获取更多示例。