

# Analysis of consistency of basic concepts for the expression of uncertainty in present International Written Standards and Guides

*Franco Pavese*

INRIM\*, Torino, Italy

*\*Until 31.12.2005, Istituto di Metrologia "G.Colonnetti" (IMGC of CNR)*

# Basic concepts on which consensus is vital

- **Type A, Type B uncertainty components**
- **Repeatability, intermediate precision, reproducibility, (im-)precision, accuracy, trueness**
- **True value, random errors, systematic effects (systematic error), bias**

## **An examination on weather**

- consensus is already achieved, at least in present International Written Standards and Guides, and
  - we are all talking the same language

**is the purpose of this talk.**

## **A preliminary Note: Replicated measurements**

In metrology one is generally performing not a single measurement but a “**series of measurements**” (**QUAM 2000.1**),

i.e. **replicating** measurements: **ISO 3534-1 (1993)**

*“replication is the determination of a value more than once”.*

**Replicated measurements can be repeated or not, depending on the measurement conditions and processes.**

E.g., in **A2LA Guide (2002)**: *“To verify control of precision, the laboratory may perform a number of replicate measurements under repeatability conditions”*

**It is not advisable in general to interchange the words ‘**repeated**’ and ‘**replicated**’**

## a) GUM general principles:

- “*The concept of uncertainty adopted in this Guide is based on the measurement result and its evaluated uncertainty **rather than on the unknowable quantities** ‘true’ value and error”, i.e. on “observable quantities”.*
- “*The term ‘**true value** of a measurand’ ... **is not used** in this Guide because the word ‘true’ is viewed as redundant”, because “the ‘true’ value of the measurand (or quantity) is simply the **value of the measurand** (or quantity)”*
- “*The **realized quantity**” is “an **approximation of the measurand**”, the latter being defined according to VIM (1993) “the quantity subject to measurement”.*

## Observed problems

GUM: “Neither the **value of the realized quantity** nor the **value of the measurand** can ever be known exactly”

→Also the value of the measurand and the value of the realized quantity are **unknowable**, as the ‘true’ value is.

GUM: “For all practical purposes associated with the measurement” the measurand “**value is unique**”, ...  
... “ a **unique ‘true value’** is only an idealized concept”.

→But they are said by GUM to carry the same meaning.

Anyhow, the measurement is only dealing with the **realized quantity**, an approximation to the measurand, for the unknowable value of which there is no reason to be unique.

(Ehrlich *et al* on **VIM3** (ACQUAL (2007) 201-218) say that GUM is using the **concept** of ‘true value’ but not the **term**.)

# Observed problems

GUM indicates that “*remaining error*”, difference to the “*value of the measurand*”, “*values of measurand due to incomplete definition*” are **unknowable** contributions. Therefore, they are **not** included in the “**result of measurement**”, and are only indicated to contribute to an so-called “final result of measurement”.

Ehrlich *et al* on **VIM3** (ACQUAL (2007) 201-218) say instead that, according to the GUM, the “*definitional uncertainty is to be included in the calculation of measurement uncertainty*”  
→ it should **not**, if **unknowable** contributions are to be out the scope of the GUM consideration.

## b) GUM general principles

- “*the result of the measurement of the **realized quantity** is **corrected** for the difference between that quantity and the measurand*”

### Observed problems

→ **Both** terms of the difference, “realized quantity” and “measurand”, are **not observable**.

The first because its value is as unknowable as that of the “quantity”

GUM: “*Neither the **value of the realized quantity** nor the **value of the measurand** can ever be known exactly*”

The second because the knowledge of the measurand can only be approximated.

## c) GUM general principles

- “*it is assumed that the results of a measurement have been corrected for all **recognized significant systematic effects***”
- “*after the correction, the **expectation** or expected value of the error arising from a □ systematic effect is **zero***”

### Observed problems

→ **No unrecognised systematic effects should exist** (assumed instead to exist –though unknowable– in the GUM Appendix D).

→ The correction should be **exact**.

**Both instances are unverifiable.**

(In the GUM the systematic effect is assumed to be uncertain)

*Note: the word “recognized” carries an ambiguity about the fact that, after identification, investigation bringing to the correction would consist of experimental determinations or of inference from modelling.*

## d) GUM general principles

- “*Implicit in this Guide is the assumption that a measurement can be **modelled mathematically** to the degree imposed by the required accuracy of the measurement*”
- “*The mathematical model of the measurement procedure that transforms the set of **repeated observations** into the measurement result is of critical importance since, in addition to the observations, it generally **includes various influence quantities that are inexactly known***”

iNRI

## Observed problems

A mathematical model is used and the uncertainty propagation law, in place of the variational method, to take into account the effect on the measurement result of the variability of the influence factors: the variational method is only recommended for checking model incompleteness.

→ **Is it alternative to the determination of reproducibility?**

Model incompleteness is assumed **not** to contribute to the variance of the “result of measurement”, but to be only a random unknowable contribution to that of the “final result of measurement”.

→ It should be considered as a **non-uniqueness**, since it is responsible for the multiplicity of the measurand values.

→ No mention of the fact that in **empirical** or **theoretical** models the values of the model parameters are determined from basically different processes.

## e) GUM general principles

- “**Type A** ... method of evaluation of a standard uncertainty by the statistical analysis of a **series of observations**” and that uncertainty component “is obtained from a Type A evaluation calculated from a series of **repeated observations**”
- **INC-1 (1980)**: “the uncertainty components in **category A**” are “those which are evaluated by applying statistical methods to a series of **repeated determinations**”

### Observed problems

GUM “Variations in repeated observations are assumed to arise because influence quantities that can affect the measurement result are not held completely constant”.

→ **The universally adopted definition of ‘repeated measurements’ exclude the variability of any influence quantity “over a short period of time”**.

# The concepts of *repeatability*, *reproducibility* and *accuracy* according to VIM (2007), ISO 5725, ISO 3534 and others

## Repeatability conditions

- a) **VIM (2007) definition:** “*condition of measurement in a set of conditions that includes the **same** measurement procedure, same operators, **same** measuring system, **same** operating conditions and **same** location, and replicate [means repeated?] measurements on the same or similar objects over a **short period of time**”*
- b) **ISO 3534-1 (1993), ISO 5725-1 (1994), VIM (until 2004 draft):** “*conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time*”

# The concepts of *repeatability*, *reproducibility* and *accuracy* according to VIM (2007), ISO 5725, ISO 3534 and others

## Repeatability conditions

- c) ISO 3534-2 (2006):** “*observation conditions where independent test/measurement results are obtained with the **same** method on **identical** test/measurement items in the **same** test or measuring facility by the **same** operator using the **same** equipment **within short intervals of time**”*”

iNRI

→ **The concept of “influence quantity” variability is not a characteristic of ‘repeatability’ (“repeated observations”) but of ‘reproducibility’**

→ **The concept of “systematic effects” is a characteristics of ‘accuracy’**

# **Type A and Type B uncertainty components concepts according to INC-1, GUM & VIM (2007)**

## **VIM (2007) definition:**

- **Type A** evaluation of measurement uncertainty arises from “*evaluation of a component of measurement uncertainty by a statistical analysis of quantity values obtained under **defined** measurement conditions*”, where the conditions can be “***repeatability** condition of measurement, **intermediate precision** condition of measurement and **reproducibility** condition of measurement*”
- **until 2004 draft**, from “*a statistical analysis of the quantity values obtained by measurements under **repeatability** conditions*”

# The concepts of *repeatability*, *reproducibility* and *accuracy* according to VIM (2007), ISO 5725, ISO 3534 and others

## Reproducibility conditions

It involves the concept of “**influence quantity**” ...

a) **GUM** (from **VIM** (1993)) “*quantity that is not the measurand but that affects the result of the measurement*”

b) **VIM** (2007) “*quantity that, in a **direct** measurement, does not affect the quantity that is actually measured, but affects the relation between the indication and the measurement result*”

⇒ The variability of the influence quantities is the entire and only potential source of the systematic effects (or error).

⇒ Time is almost never an influence quantity in itself, but in time the influence quantities can show a variability.

# The concepts of *repeatability*, *reproducibility* and *accuracy* according to VIM (2007), ISO 5725, ISO 3534 and others

## Reproducibility conditions

... and the concept of “**systematic effect**” or “**error**”:

**c) VIM (2007)** “*Component of measurement error that in replicate measurements remains constant or varies in a predictable way*”.

**VIM (1993)** was stating “*Mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions minus a true value of the measurand*”

(similarly **QUAM 2000.1**)

**d) DIN 1319 Part I (1995)** indicates that the absolute systematic error comprises two components:

- one covers the known systematic measurement error,
- the other the unknown systematic measurement error.

# The concepts of *repeatability*, *reproducibility* and *accuracy* according to VIM (2007), ISO 5725, ISO 3534 and others

## Reproducibility conditions

- **VIM (2007)** “*measurement precision under **reproducibility conditions** of measurements*” are those “*condition of measurement in a set of conditions that includes different locations, operators, measuring systems, and **replicate** measurements on the same or **similar** objects*”, noting that “*The different measuring systems may use **different measurement procedures***”.

## ***Reproducibility conditions:***

**b) ISO 3534-2 (2006) and ISO 5725-1 (1994) for testing state the same, except the last condition (*different procedures*). ISO 5725-1 also includes the effects of calibration and time (same A2LA 2002).**

## ***Contrasting definition:***

**QUAM (2000.1) “variability obtained when different laboratories analyse the same sample”,**

## f) GUM general principles

- “**Type B** ... method of evaluation of a standard uncertainty by **means other than the statistical analysis of a series of observations**”
- “... founded on a priori distributions”
- **INC-1 (1980)**: “the uncertainty components in **category B**” are “those which are evaluated by other means”

### Observed problems

Type B components of uncertainty are said to be based on methods other than statistical analysis, i.e., on *prior* knowledge not always based on experimental verification.

→ Does it always reconcile with the GUM statement on **knowable quantities**?

# *Type A and Type B uncertainty components concepts according to INC-1, GUM & VIM (2007)*

## **VIM (2007) definition:**

- **Type B** evaluation of measurement uncertainty arises from “*evaluation of a component of measurement uncertainty determined by means other than a Type A evaluation of measurement uncertainty*”,

with examples given:

- “*evaluation based on information associated with authoritative published quantity values;*
- *associated with the quantity value of a certified reference material;*
- *obtained from a calibration certificate;*
- *about drift;*
- *obtained from the accuracy class of a verified measuring instrument;*
- *obtained from limits deduced through personal experience”*

## a) IEC 60359 general principle

- The concept of **true value** is not only unknowable, but is said to be also **unnecessary**
  - The stated objective of the measurement is said to be to obtain measurement results that are **compatible** with each other
  - **VIM (2007): metrological compatibility** is the “*property of all pairs of measurement results for a specified measurand, such that the absolute value of the difference of the measured quantity values is **smaller than some chosen multiple** of the standard measurement uncertainty of that difference*”
- **How this reconciliates with GUM ?**

## a) Other VIM basic definitions

### Systematic measurement error

*“Component of measurement error that in replicate measurements remains **constant** or **varies in a predictable manner**”*

### Observed problems

- A systematic error **constant in value** (stable influence quantities) is only assumed for **repeated** measurements (*“over a short period of time”*)
- What about its **unpredictable variability** ?

## a1) Other VIM basic definitions

### Systematic measurement error

*“NOTE 1 — A reference quantity value for a systematic measurement error is a **true quantity value**, or a **measured quantity value** of a measurement standard of negligible measurement uncertainty, or a **conventional quantity value**.”*

### Observed problems

In order to talk of “measurement error” the two terms of the difference producing the error should be comparable.  
→ The **first** value is unknowable, the **second** value has an uncertainty associated to it, the **third** value is not directly comparable with the measured value and may be affected by an uncertainty.

## a2) Other VIM basic definitions

### Systematic measurement error

“*NOTE 2 — Systematic measurement error, and its causes, can be **known** or **unknown**. A correction can be applied to compensate for a known systematic measurement error.*”

### Observed problems

→ **How this reconciliates with the present GUM?**

iNRI

## a3) Other VIM basic definitions

### Systematic measurement error

“NOTE 3 — *Systematic measurement error equals measurement error minus random measurement error.*”

### Observed problems

→ This wording is consistent with the VIM definition (“constant”), **but**, instead, the random measurement error, for the **GUM** (and, more generally, for non-repeated observations), **includes** a contribution due to the **variability of systematic effects**.

Until the VIM 2004 draft, “*mean that would result from an infinite number of measurements of the same measurand carried out under **repeatability** conditions minus a true value of the measurand*”

## b) Other VIM basic definitions

### Repeatability conditions

a) VIM (2007) definition: “*condition of measurement in a set of conditions that includes the same measurement procedure, same operators, same measuring system, same operating conditions and same location, and replicate [means repeated?] measurements **on the same or similar objects** over a short period of time*”

### Observed problems

- The wording “similar objects” is **generic**.
- Objects?**
- It can be correctly used in testing while in metrology its use should be **excluded for artefact** standards.

## a) Other ISO general principles

### True value, trueness

Testing is generally intrinsically different from metrology.

A “**true value**” can be assigned to the measurand (*‘Conventional Value Hybrid Approach’ –CVHA*), as a “reference value” or a “consensus value” and one can define “**trueness**”:

**ISO 5725-1** (1994) “*The closeness of agreement between the average value obtained from a large series of test results and an **accepted reference value***”, adding

“*The measure of trueness is usually expressed in terms of **bias***”.

## a) Other ISO general principles

### True value

→ **ISO underlying definition looks not consistent with that of VIM (2007)** “*quantity value consistent with the definition of a quantity*” (a tautology?)

noting, however, a variety of ‘schools of thinking’:

“ *In the **Error Approach** to describing measurement, a **true quantity value** is considered **unique and, in practice, unknowable**.*

*The **Uncertainty Approach** is to recognize that, owing to the inherently incomplete amount of detail in the definition of a quantity, there is **not a single true quantity value** but rather a set of true quantity values consistent with the definition. However, this set of values is, **in principle and in practice, unknowable**.*

***Other approaches dispense altogether with the concept of true quantity value and rely on the concept of metrological **compatibility** of measurement results for assessing their validity**”*

## d) Other ISO general principles

### Bias in testing

- **ISO 5725-1** (1994), **3534-1** (2003) and **A2LA Guide** (2002)  
*“The difference between the expectation of the test results and an **accepted reference value**”*
- **ISO 3534-2** (2006) *“difference between the expectation of a test result or measurement result and a **true value**”*
- c) **EA-4/16** (2003) *“The bias of a test method is usually determined by studying relevant **reference materials or test samples** ... The uncertainty associated with the determination of the bias is an important component of the overall uncertainty”*
- d) **QUAM** (2000.1) *“Where the **bias is significant** compared to the combined uncertainty, additional action is required”*

## c) Other ISO general principles

### Bias in metrology

- **VIM (2007)** “*Estimate of a systematic measurement error*”

but **VIM 2006** draft: “*Systematic measurement error or its estimate, with respect to a reference quantity value*”.

- b) GUM** “*it is assumed that, after correction, the expectation or expected value of the error arising from a systematic effect is zero*”, **i.e., zero bias**

→ These documents are assumed to be valid also for testing.

# Thank you for your attention

