XIX IMEKO World Congress
Fundamental and Applied Metrology
September 6−11, 2009, Lisbon, Portugal

ACCURACY, TRUENESS, AND PRECISION: CONSIDERATIONS BASED ON THE INTERNATIONAL VOCABULARY OF METROLOGY (VIM, 3RD ED.) AND RELATED STANDARDS

Roberto Buccianti[1], *Marco Cibien*[2], *Luca Mari*[3], *Bruno I. Rebaglia*[4]

[1] CEI – Comitato Elettrotecnico Italiano, Milano, Italy, r.buccianti@alice.it
[2] UNI – Italian Organization for Standardization, Milano, Italy, marco.cibien@uni.com
[3] Università Cattaneo – LIUC, Castellanza (VA), Italy, lmari@liuc.it
[4] ITIA – CNR, Milano, Italy, b.rebaglia@itia.cnr.it

Abstract − The paper analyzes the concept of accuracy, and the related ones of trueness and precision, as aimed at characterizing the behavior of measurement processes, and highlights how such concepts, as currently defined by some relevant standards, should be reformulated to guarantee their consistency. Finally, some preliminary suggestions are proposed to this goal.

Keywords: measurement science; measurement process behavior; accuracy

1. INTRODUCTION

The recent publication of the third edition of the *International vocabulary of metrology — Basic and general concepts and associated terms* (VIM3 [1]), "intended to promote global harmonization of terminology used in metrology" {VIM3, Scope} (an internal reference to the referred standard), has generated a new interest in measurement science in general, and particularly in reaching a common understanding about the parameters by which the behavior of measurement processes can be characterized. At this regards three basic concepts can be acknowledged as fundamental: accuracy, trueness, and precision. The usage of such concepts is widespread in many metrological fields, but noteworthy theoretical and operational discrepancies can be found in their meaning adopted in different sectors, so that even giving a common definition of them can be a complex task. The presence of these terms in the VIM3 is then welcome, a document aimed at being "a common reference for scientists and engineers – including physicists, chemists, medical scientists – as well as for both teachers and practitioners involved in planning or performing measurements, irrespective of the level of measurement uncertainty and irrespective of the field of application" {VIM3, Scope}. Nevertheless, some issues still persist, in particular as far as the compatibility of the VIM3 with respect to other standards related to qualification of measurement, and specifically the ISO 5725 [2] and the ISO 3534 [3] series, is concerned. Such standards explicitly derive some of their definitions – but not the three mentioned ones – from the second edition of the VIM (VIM2 [4]). In view of a possible revision of these documents, a comparative analysis among them appears to be an important goal to guarantee the compatibility of their underlying conceptual frameworks. A concrete benefit of this compatibility would be the availability of a common terminological ground for all subjects involved in designing and performing measurements, leading to attribute the same operative meaning to the same concepts, and the related indicators and numerical values.

Starting from a general discussion on how to qualify the behavior of a generic measurement process, this paper addresses some of these issues, by analyzing the differences between the conceptual frameworks set by the VIM3 and the ISO 5725 / ISO 3534 series, and finally investigating the critical role of measurement accuracy.

2. BASICS OF MEASUREMENT PROCESS BEHAVIOUR EVALUATION

Whenever, under given conditions, a sample {3534-1, 1.3} of indications {VIM3, 4.1} is obtained (note that the VIM3 does not give a name to the operation by which an indication is acquired, actually a stage of a measurement process) from replicate measurements, several statistics can be computed on such a sample, and in particular:

- *location* (or "position") statistics, such as sample median {3534-1, 1.13} and, in the case of ratio scale evaluations, sample mean {3534-1, 1.15};
- *scale* (or "dispersion") statistics, such as p-quantiles {3534-1, 2.13} and, once more for ratio scale evaluations, sample standard deviation {3534-1, 1.17}.

Both location statistics and scale statistics (these terms are taken from {3534-1, 2.9 Note 2}) can be exploited to convey information on the behavior of the measurement process, as resulting from the available sample. To this goal such two categories exhibit however a basic asymmetry since:

- the synthesis performed by a scale statistic gives a sufficient information to be interpreted in terms of measurement process behavior; for example, the behavior decays as the sample standard deviation increases;

- on the other hand, information on measurement process behavior can be obtained from a location statistic only if the value of this statistic is compared with a reference value; if, as typical for ratio scale evaluations, this comparison is formalized as distance between the sample mean and the given reference value, the behavior decays as this distance increases.

In this regard, several typologies of reference values can be envisaged (see also {3534-2, 3.2.7}, that defines "accepted reference value" as "value that serves as an agreed-upon reference for comparison"):

- theoretical values based on scientific principles;
- values obtained from the collaborative experimental work of a scientific or technical group, typically by a peer inter-laboratory comparison process;
- values obtained from the experimental work of some national or international recognized organization;
- values materialized by (not necessarily traceable) working standards, typically agreed on by customers / users and suppliers / manufacturers;
- values computed from series of previous observations of the same system.

Whatever typology is adopted, the relation between location statistics and reference values is generally expressed in terms of *bias*, a concept in its turn defined in multiple and non-coinciding ways by the mentioned standards: "estimate of a systematic measurement error" {VIM3, 2.18}, "systematic error of the indication of a measuring instrument" {VIM2, 5.25}, "expectation of error of estimation" {3534-1, 1.33}, "difference between the expectation of a test result or measurement result and a true value" {3534-2, 3.2.7}.

Furthermore, the situation is made even more complex by the fact that the VIM3, in dealing with reference values, notes that "a reference quantity value can be a true quantity value of a measurand, in which case it is unknown, or a conventional quantity value, in which case it is known." {VIM3, 5.18 Note 1}, being important noting that the second case mentions quantities, and not specifically measurands, with the consequence that a conventional quantity value is not required to be obtained by measurement, as indeed the definition {VIM3, 2.12} highlights (the VIM3 itself acknowledges that a third typology – a "measured quantity value of a measurement standard of negligible measurement uncertainty", as listed in {VIM3, 5.17 Note 1} – should be added here). Accordingly, it appears that a further discrimination should be made, to prevent that an "unknown" reference value is chosen and therefore that no information on measurement process behavior can be obtained from a location statistic.

3. ACCURACY, TRUENESS, AND PRECISION: ANALYSIS

The previous analysis has been made by carefully avoiding any usage of the terms "accuracy", "trueness", "precision", and also "random measurement error" and "systematic measurement error". It is now time to take into account how such terms are defined in the VIM3 and the ISO 5725 and the ISO 3534 series, so to relate them to the analyzed concepts of measurement process behavior.

Table 1. Definitions of "accuracy".

5725-1 {3.6}	Closeness of agreement between a test result and the accepted reference value
3534-2 {3.3.1}	Closeness of agreement between a test result or measurement result and the true value
VIM3 {2.13}	Closeness of agreement between a measured quantity value and a true quantity value of a measurand

These definitions share the same concept that accuracy relates to the relation between a reference value and a quantity value obtained experimentally (for our current purposes the differences between "test result", "measurement result" – as defined by the VIM2 –, and "measured quantity value" can be neglected). On the other hand, they differ significantly.

For witnessing the current state of confusion around the concept of accuracy, some further definitions can be drawn from the IEC 60050 series [5] – the *International Electrotechnical Vocabulary* –:

- "closeness of the agreement between the result of a measurement and the conventionally true value of the measurand" {IEV, 394-40-35};
- "quality which characterizes the ability of a measuring instrument to provide an indicated value close to a true value of the measurand" {IEV, 311-06-08};
- "specified value of a parameter that represents the uncertainty in the measurement" {IEV, 415-05-12}.

The ISO 5725 standpoint is definitely the only clearly operational one, given that accepted reference values can be quantified (theoretically, experimentally, or by agreement / consensus, as mentioned before). On the other hand, in the ISO 3534 and the VIM3 accuracy is confined to be a qualitative concept ("The concept 'measurement accuracy' is not a quantity and is not given a numerical quantity value." {VIM3, 2.13 Note 1}), due to the fact that true (quantity) values {VIM3, 2.11}, {3534-2, 3.2.5} are taken as reference, so that accuracy inherits the controversial nature of true value:

- "quantity value consistent with the definition of a quantity" {VIM3, 2.11} – what does "consistence with a quantity definition" operatively mean?
- "value which characterizes a quantity perfectly defined in the conditions which exist when that quantity is considered" {3534-2, 3.2.5} – what does "perfect definition" for a quantity operatively mean?

The VIM3 also states that true values are unknowable {VIM3, 2.11 Note 1}, a standpoint that the ISO 3534 tries to overcome in its consequences when it states that "in practice, the accepted reference value is substituted for the true value" {3534-2, 3.3.3 Note 3}. Finally, according to {3534-2, 3.3.1 Note 3} "accuracy refers to a combination of trueness and precision" and this is almost the same point of view expressed in {5725-1, 3.3.6 Note 2}, i.e., accuracy cannot be expressed in terms of bias or standard deviation only. In the next Section this claim will be further discussed.

Table 2. Definitions of "trueness".

5725-1 {3.7}	Closeness of agreement between the average value obtained from a large series of test results and an accepted reference value
3534-2 {3.3.3}	Closeness of agreement between the expectation of a test result or a measurement result and a true value
VIM3 {2.14}	Closeness of agreement between the average of an infinite number of replicate measured quantity values and a reference quantity value

As for accuracy, trueness relates to the relation between a reference value and a quantity value, that all definitions characterize as a location statistic obtained from experimental data (it should be noted that the request for such a statistic to be an expectation / average implies that trueness is defined only for quantities evaluated in algebraically rich scales). According to the VIM3 this location statistic should be computed on a sample of infinite elements, so that the concept is again a theoretical rather than operational one, as explicitly acknowledged: "Measurement trueness is not a quantity and thus cannot be expressed numerically" {VIM3, 2.14 Note 1}. Furthermore, {5725-1, 3.7 Note 3} and {3534-2, 3.3.1 Note 3} state that in this case closeness of agreement is usually expressed in terms of bias, thus making it explicit that trueness is, or is expressed by, a location parameter.

Table 3. Definitions of "precision".

5725-1 {3.12}	Closeness of agreement between independent test results obtained under stipulated conditions
3534-2 {3.3.4}	Closeness of agreement between independent test/measurement results obtained under stipulated conditions
VIM3 {2.15}	Closeness of agreement between indications or measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions

The definitions provided by the ISO 5725 and the ISO 3534 are practically identical, and the one given in the VIM3 is aligned as well. All of them clearly state that precision is a parameter entirely dependent on the sample of the available experimental data, and further note ({5725-1, 3.12 Note 10}, {3534-2, 3.3.4 Note 2}, {VIM3, 2.15 Note 1}) that in this case the generic concept of closeness of agreement is usually expressed in terms of standard deviation, i.e., by a scale statistic. This makes precision an unproblematic, well-defined scale parameter, also by noting that it is the only one defined also for indications, and therefore independently on any calibration requirement.

4. CONSIDERATIONS ON ACCURACY, TRUENESS, AND PRECISION

As the previous analysis has highlighted, all the considered standards define the three concepts of accuracy, trueness and precision in terms of "closeness of agreement", a concept in its turn not defined in the context. In order to further analyze it, let us write $c(r,e)$ to denote the relation / function of closeness of agreement c between a reference value r and an experimental value e. Hence, both accuracy and trueness, for all the three considered standards, can be expressed as $c(r,e)$. On the other hand, in the case of precision the closeness of agreement is assessed within a given sample of experimental values $\{e_1, \ldots, e_n\}$, and therefore the relation / function is $c(e_1, \ldots, e_n)$.

The term r has been already taken into account: it is in general a reference value, particularized in some definitions as a true value, or a conventional true value (a term explicitly discouraged by {VIM3, 2.12 Note 1}) sometimes in the IEC context. This latter option prevents an operational definition, but it is the only one which justifies the choice of calling $c()$ as "trueness". In the case of definitions given by the ISO 5725 and the VIM3, indeed, the definition of trueness does not explicitly mention any dependence on true values. On the other hand, this can also be considered a purely conventional lexical choice, and accepted as such. The important point here is however that the option r = true value seems to make the parameter (related to) $c()$ useless to characterize quantitatively any measurement process or measuring system, and makes even ordinal comparisons (this measurement process / measuring system is more accurate / true – is "true" the proper adjective of "trueness"? – than that one) doubtful to perform and interpret in their results.

As far as the term e is concerned, the position of the ISO 5725 (and the ISO 3534, under the condition that measurement is modeled so that its result is formalized as a function, required to compute the expectation on it) seems to be a clear and consistent one: accuracy is defined on the basis of the value of a *single* experimental operation, whereas trueness implies the relation with a value computed *from a sample*, i.e., a set of replicated experimental operations. Accordingly, trueness would appear to be a general concept that includes accuracy as a specific case, obtained when the sample size is 1. In other terms, of the three concepts under consideration accuracy is the only one whose definition does not require the measurement process to be replicated (it could be also noted that the VIM3 uses the adjective "replicate" for both measurement and, in elliptical way, quantity values).

However, this is not the rationale on which the ISO 5725 series is based, for which accuracy is stated as consisting in a combination of two well-structured quantitative parameters based on location and scale statistics, expressing trueness and precision, respectively. Although in a less specified sense, also the VIM3 states that measurement accuracy "is related to both" trueness and precision {VIM3, 2.13 Note 2}. Nevertheless, although these parameters are generally expressed in the same unit, they are not quantities of the same kind {VIM3, 1.2} and thus they should not be combined analytically according to any additive logic. Hence, the issue of the polysemic nature of the above definitions of accuracy (perhaps partly due to the fuzzy meaning of the syntagm "closeness of agreement") requires further consideration.

By taking into account a generic case of "closeness of agreement" $c(r,e)$, it can be noted that the reference value r can be obtained:

a) for the *same* measurand on which e is evaluated;

b) for a quantity that *is not* the one on which e is evaluated, a typical operational case.

From the standpoint of measurement process behaviour evaluation such two cases are markedly different, since an observed difference (bias) between r and e:

- in case a) such a difference is reasonably related to the measurement process only, as it could happen if the experimental value e is compared to the average value r of the previously obtained samples of the same process, so to assess whether the process itself is "under control";

- in case b) such a difference is related both to the measurement process and the specific conditions in which the reference value r is obtained.

While nothing in the above mentioned definitions of accuracy prevents the applicability of the concept to either case, only case a) seems to provide an unambiguous, inter-subjective, interpretation of it. Hence, as far as the concept of accuracy, together with those of trueness and precision, is aimed at characterizing the behavior of measurement processes, the scope of the related definitions must be narrowed to case a).

5. RETHINKING THE CONCEPTS

On the basis of this analysis, the definitions given by the VIM3 appear to be critical for at least two reasons:

- they do not point out in sufficiently clear way that the concepts specifically relate to the measurement process;

- while they are explicit on the relations between the general concept and its possible quantitative expression, the differences in the three cases (accuracy "is not a quantity and is not given a numerical quantity value" {VIM3, 2.13 Note 1}; trueness "is not a quantity (…) but measures for closeness of agreement are given (…) {VIM3, 2.14 Note 1}; precision "is usually expressed numerically by measures of imprecision, such as (…) " {VIM3, 2.15 Note 1}) are not clearly justified, and in particular the statement on trueness is ambiguous, as it is the concept of closeness of agreement used in it.

Our aim here is to try overcoming this limitation, by introducing a clear-cut distinction between the general concepts and their possible quantitative expressions, thus under the hypothesis that, in particular, *a concept* of accuracy can be defined and, in some given cases, can be quantified by *an indicator* of accuracy, as the VIM3 already acknowledges for precision. The reason of what follows can be simply stated: *a well-founded, unambiguous definition of the properties by which a measurement process can be characterized is very important for both theoretical measurement science* (for example for studying general dependence relations between accuracy of measurement processes and uncertainty of their results) *and all its applications* (for example for standardizing the way measuring instrumentation features are specified).

Hence, according to the VIM3:

- accuracy is defined as $c_{acc}(r_{true}, e)$, being r_{true} a "true quantity value of a measurand" and e a generic measured quantity value;

- trueness is defined as $c_{tru}(r, e_{ave})$, being r a generic reference quantity value and e_{ave} the "average of an infinite number of replicate measured quantity value";

- precision is defined as $c_{pre}(e_1, \ldots, e_n)$, exactly as mentioned before.

Under the hypothesis that c_{pre} and c_{tru}, and then c_{acc}, are aimed to be properties specific of a measurement process *mp*, a basic constraint to be enforced is that the arguments of such relations / functions must be related to the *mp* only (as in the case a) above). This operatively means that these properties must be defined so to forbid the situation in which their value changes while the considered *mp* has not changed.

While precision already satisfies this condition, trueness, as currently defined, does not¸ since no relation is imposed between the replicate measured quantity values, and thus e_{ave}, and the chosen reference quantity value r. Indeed, since the choice of a reference quantity value is not (necessarily) part of a *mp*, $c_{tru}(r, e_{ave})$ could change even while the considered *mp* remains the same. In other terms, the VIM3 does not define the trueness of a *mp*, but the "trueness of a *mp to a reference*". On the other hand, since the quantity of which r is a value is not definitionally constrained to be the same of which e_{ave} is computed, the latter quantity being admitted in principle to remain completely unknown, $c_{tru}(r, e_{ave})$ does not convey, rigorously speaking, *any* information on the conveyed *mp*. Finally, relating to accuracy our analysis already showed the basic issue of the definition: since the term e in $c_{acc}(r_{true}, e)$ is not required to be obtained, as instead in the case of trueness, as e_{ave}, the mentioned claim for accuracy to be related to trueness and precision {VIM3, 2.13 Note 2} is, in general, not justified.

To obtain an operative and operatively useful definition of accuracy, trueness and precision, coherent with the assumption that they should be properties characterizing a *mp*, a detail in the VIM3 must be mentioned once more:

- while in a note to the definition of "reference quantity value" *two* cases are listed as possible ("a reference quantity value can be a true quantity value of a measurand, in which case it is unknown, or a conventional quantity value, in which case it is known" {VIM3, 5.18 Note 1}),

- in a different context the VIM3 itself acknowledges that *three* cases of reference quantity values are allowed ("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" {VIM3, 2.17 Note 1}).

The second case, "the missing one" – a measured quantity value of a measurement standard of negligible measurement uncertainty – is exactly what is typically exploited to qualify a *mp*. When employed to this goal, such a reference quantity value is actually dealt with as an *operative true value*, and as such we propose to call it. The adjective "operative" emphasizes the fundamental feature that, under the assumption that metrological traceability is guaranteed, this value can be assumed as *to be known*.

Let us denote r_{meas} this value and e_{meas} any experimental value obtained for the (assumed) same quantity by means of a given *mp*. On this basis, three properties of a *mp* can be defined:

P1: $c_{\mathrm{P1}}(e_1, \ldots, e_n)$, the closeness of agreement between measured quantity values (or indications) under specified measurement (repeatability / reproducibility) conditions;

P2: $c_{\mathrm{P2}}(r_{\mathrm{meas}}, e_{\mathrm{meas}})$, the closeness of agreement between an operative true quantity value of a measurand and a measured quantity value of the (assumed) same measurand;

P3: $c_{\mathrm{P3}}(r_{\mathrm{meas}}, e_{\mathrm{ave}})$, the closeness of agreement between an operative true quantity value of a measurand and a measured quantity value of the (assumed) same measurand, obtained as average of a suitable number of replicate measured quantity values.

It should be noted that, as explicitly acknowledged for sensitivity {VIM3, 4.12 Note 1}, these properties can possibly depend on the value of the quantity being measured, and therefore, generally speaking, they characterize a *mp* as operating in a given measuring sub-interval (see {VIM3, 4.7}).

It is important to consider that the proposed definitions of P1, P2, and P3:

- do not imply any idealization of the *mp*, and therefore they are operative;
- characterize general concepts that can be quantified by means of suitable indicators.

A comparative analysis of P1 and P2 highlights their role of complementary properties:

- P1 accounts for scale information only, being independent of any position information; as such, P1 can be evaluated independently of any reference quantity value;
- P2 accounts for position information only, being independent of any scale information; as such, P2 can be evaluated independently of any repeatability condition.

Although we are not concerned here with lexical issues, it can be noted that:

- P1 has exactly the role attributed by the VIM3 to precision, and as such it could be called the *measurement process precision*;
- P2 is defined as the closeness of agreement between a product of a *mp* and an operative true quantity value, and therefore it could be properly called the *measurement process trueness*;
- P3 is a property taking into account both scale and position information, and as such it can be interpreted as an overall property of a *mp*, related to both P1 and P2. This is the role the VIM3 attributes to accuracy: hence, P3 could be called the *measurement process accuracy*.

6. CONCLUSIONS

An inter-subjectively defined set of properties aimed at qualifying measurement process and measuring systems and instruments is of tremendous importance for the industrial practice. Such properties should be defined as general, operative concepts, admitting to be quantified by means of consistently defined indicators. As the current standards are in this regard incomplete, and sometimes in contradiction with each other, in this paper we have suggested some preliminary hypotheses to (re)define some basic metrological concepts to this aim.

REFERENCES

[1] JGCM 200:2008, International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM), Sèvres, France, 2008; available online: http://www.bipm.org/en/publications/guides/vim.html

[2] ISO 5725, Accuracy (trueness and precision) of measurement methods and results, Geneva, Switzerland, different publication dates for the various parts.

[3] ISO 3534, Statistics – Vocabulary and symbols, Geneva, Switzerland, different publication dates for the various parts.

[4] VIM:1993, International vocabulary of basic and general terms in metrology, Geneva, Switzerland, 1993.

[5] IEC 60050, International Electrotechnical Vocabulary, Geneva, Switzerland, different publication dates for the various parts; available online (as Electropedia: The World's Online Electrotechnical Vocabulary): http://www.electropedia.org