

MEASUREMENT FUNDAMENTALS: A PRAGMATIC VIEW

L. Mari(1), P. Carbone(2), D. Petri(3)

(1) Università Cattaneo – LIUC, C.so Matteotti, 22, 21053 Castellanza (Va), Italy.

email: lmari@liuc.it, Tel. ++39 0331 572228

(2) DIEI –Università degli Studi di Perugia, Via G. Duranti, 93, 06125 Perugia, Italy.

email: carbone@diei.unipg.it, Tel. ++39 075 5853629

(3) DISI – Università degli Studi di Trento, Via Sommarive, 14, 38050 Povo –Trento, Italy.

email: petri@disi.unitn.it, Tel. ++39 0461 283902

Abstract

Measurements are more and more required in an increasing variety of human activities to acquire reliable information useful for effectively supporting decision making processes. Furthermore, the entities whose properties are to be measured and the measuring systems are becoming increasingly complex, hardly to be modeled and managed. In this rapidly evolving scenario several issues concerning the fundamentals of measurement science and technology arise. The purpose of this paper is to discuss some aspects of the crucial question about which evaluation processes can be considered measurements. Rather than focusing on formal conditions or technological constraints, we propose a pragmatic characterization of measurement, under the assumption that a better comprehension of the concept can be achieved by identifying and discussing the basic features which justify the reliability attributed to measurement results. With such an inter-disciplinary approach, this work aims at promoting a broad discussion among interested researchers, even working in different scientific disciplines, so to increase the synergies among different research areas and to improve the body of knowledge about measurement fundamentals.

1. INTRODUCTION

In many, if not all, human activities measurement is considered a fundamental process to obtain reliable information on the empirical world, in view of the Galilean motto of measuring what is measurable and making measurable what is not yet. Nowadays a systematic adoption of measurement is particularly solicited by the widespread application of techno-science in the social, industrial, economical, ... fields. Once specifically aimed at evaluating physical quantities, measurements are today more and more required in biology, medicine, economy, sociology, psychology, ... Furthermore, measuring systems are becoming complex entities, in which the customary measurement techniques have to be extended to multivariate measurands, and the measurement of physical quantities has to be often complemented by the measurement of non-physical properties, sometimes called “weakly defined measurement” [1], or simply “soft measurement”. An excellent reference on this matter is the document *Evolving needs for metrology in trade, industry and society and the role of the BIPM* [2], which states for example that currently “an estimated 80 % [of the world trade] is affected by standards and regulations” and that according to various studies “the cost to producers and service providers of complying with standards can be 10 % of production costs”. Of course, measurement is the basis to assess such compliance. The Bureau International des Poids et Mesures (BIPM) document lists some of the application areas where the role of measurement is increasingly critical: they include “transport; information technology, navigation and telecommunications; electronics and optics; electromagnetic and ionizing radiation; energy; climate change, environmental and pollution control; clinical chemistry and laboratory medicine; food safety; anti-doping; pharmaceuticals; forensics and security”.

This evolving scenario arises several significant issues for measurement science, not only at the operative level – for example about the possibility to apply the now standard procedures of uncertainty evaluation specified by the *Guide to the expression of uncertainty in measurement* [3] in such diverse fields – but also, and primarily, in reference to the state and the nature of measurement science itself. A fundamental, critical problem relates to the very concept of measurement: is its acceptance, as commonly understood in measurement of mechanical, optical, thermal, electrical, ... quantities, already adequate and directly applicable in this broader context? The question is not only lexical, i.e., whether a single entry in a vocabulary may accommodate all usages of the term, or only related to the, however important, goal of mutual understanding, which is the basic goal that drove a number of leading international organizations to gather into the Joint Committee for Guides in Metrology, JCGM, and to produce the *International Vocabulary of Metrology* (VIM3) [4]. Rather, an inquiry on a widely shared meaning of ‘measurement’ is

useful to give a convincing justification to the customary claim of the “special reliability” of measurement itself, which is surely not assumed in the case of, e.g., subjective judgment or guess: the public trust attributed to measurement results and the resource spending acceptable for measurement should not be approved for such other activities.

If the generic process of assigning a quantity value to a quantity of interest is termed “evaluation” (as in the abstract case of ‘function evaluation’), so that measurement, subjective judgment, and guess are all examples of evaluations, the problem may be then formulated: *what does it characterize measurement as a specific kind of evaluation?*

Since the problem is not conceptually new (although perhaps this formulation is), some answers have been already proposed in the past. At least three well known, general standpoints can be mentioned.

First standpoint: according to Euclid (Elements, Book V, definitions 1-3), “a magnitude is a part of a magnitude, the less of the greater, when it measures the greater; the greater is a multiple of the less when it is measured by the less; a ratio is a sort of relation in respect of size between two magnitudes of the same kind”. The hypothesis is assumed here that a quantity is measurable because it can be represented as the ratio of (integer) numbers. While this standpoint has shaped the traditional concept of measure, the lack of any specifications on the way values are obtained (“according to my experience, I can see that this object is 1,2 m long” expresses in fact a ratio of two “magnitudes”) makes it useless for our purpose. Furthermore, even in the case of physical quantities it is today customarily accepted that ordinal properties can be measurable (the VIM3 calls them “ordinal quantities” and mentions four examples, Rockwell C hardness, octane number for petroleum fuel, earthquake strength on the Richter scale, subjective level of abdominal pain on a scale from zero to five). This makes “Euclidean quantities” a specific, although very important (quantity calculus applies only to them), case of a broader class. The above considerations highlight that the issue of measurability relates in principle to the very general concept of property, for which some more or less synonymous terms can be used, such as “feature”, “aspect”, “characteristic”, ..., but whose definition, if even possible is not a matter of measurement science.

Second standpoint: measurement is what is performed by a (physical) measuring system, i.e., a properly calibrated and operated instrument realizing a physical transduction effect. Despite its operational roughness, this standpoint has shown its effectiveness for centuries, and it can be considered the (default) acceptance in physical sciences and engineering. On the other hand, bounding measurement to physical instrumentation removes our problem without solving it [5]: the lexical move of making “measurement” and “physical measurement” synonymous is plainly inappropriate in the BIPM perspective of the above mentioned “evolving needs” for metrology and towards soft measurement.

Third standpoint: as the outcome of a critical analysis on the possibilities of applying measurement in social sciences, measurement has been axiomatized as a morphic mapping from quantities to quantity values, and actually and more generally from properties to property values. Morphisms are structurally constrained mappings which require that, in assigning a property value to a property, empirical observable relations between properties are preserved in abstract relations between property values [6]. This condition, which is in principle trivial, allows expressing the empirical relations occurring between instances of properties by means of properly chosen linguistic entities. For example, a morphic constraint in the ordinal case is: if a resistor x_1 experimentally behaves by exhibiting a higher resistance than a resistor x_2 , then the resistance value assigned to x_1 must be greater than the resistance value assigned to x_2 . Hence, the derived so-called representational theories of measurement [7] generalize the Euclidean standpoint by releasing some algebraic constraints, yet ensuring that measurement results actually reflect the behavior of the examined property. However, representational theories alone are still unable to discriminate between, e.g., measurement and subjective morphic judgments [8].

As a synthesis, our claim is that the traditionally available positions are not adequate to found an encompassing but scientifically well-defined concept of measurement, able to cope with the current “evolving needs”: the new challenges require a new characterization. In the following Section measurement is presented as an evaluation process and, despite this still generic description, some of its basic features are introduced. In Section 3 measurement is characterized as an objective and inter-subjective evaluation, according to a well-defined, technical meaning of the concepts of objectivity and inter-subjectivity. This allows arguing, in Section 4, on the twofold nature of measurement, being at the same time a model-based and an operative evaluation, whose implications are discussed in Sections 5 and 6 with respect to measurement uncertainty, also in reference to the critical case study of the measurability of research quality.

2. MEASUREMENT AS EVALUATION

The background assumption that we are introducing – measurement is a specific kind of evaluation – is still

very generic, and still it throws some light on some fundamental features that measurement shares with any other evaluation. Pointing out that measurement is an *evaluation* abstractly characterizes it according to a black box model, as a process realizing a functional transformation of an input entity to an output entity, called in this case the *measurement result*.

This concept of evaluation as transformation is very general. For example, the mathematical transformation which takes a real number x and produces the real number $2x$ is an evaluation, in the usual sense that x , the input entity, is evaluated by means of the transformation and $2x$ is the resulting value. Output entities of evaluations are not required to be (real) numbers: they could be vectors or tensors, but also subsets or more complex entities such as probability density functions, and even linguistic (ordered or non-ordered) labels. On the other hand, not any transformation is an evaluation, as it occurs, e.g., in the case of chemical reactions. Evaluation outputs, and measurement results in particular, are information (and not empirical) entities – let us call them *symbols*, according to a customary terminology, adopted in particular in the representational theory of measurement – although of course a physical support is required for conveying (transferring, storing, presenting, ...) them. Hence, measurement is a map between the empirical world and the symbolic world, aimed at associating symbols to the (still unspecified, for the sake of generality) empirical entity under measurement in order to describe it.

A dependence relationship is assumed between measurement process input and output, typically interpreted according to a *causality principle*: outputs are effects of inputs. This is particularly the conceptual basis for all the developments related to measurement uncertainty evaluation techniques, which explain time variability of output values by assuming that at least one component of the empirical world (the entity under measurement, the measuring system, or the surrounding environment) had changed, due to the presence of so-called hidden variables, introduced to maintain causality (it is well known that in quantum physics the relations between causality and measurement uncertainty are much more complex, but we will not consider this subject here). No further constraints – and in particular no specifications on the structure of the process – are imposed on evaluations as such.

Moreover, measurement is an *informative evaluation*, aimed at acquiring and conveying information on objects of the empirical world. According to the Shannon's theory of information, this implies that the set of possible values includes at least two elements, and that a prior probability distribution can be associated to such set so that at least two elements (taking into account the discrete case) have a non-null probability to be chosen as measurement result. This excludes constant functions from the candidates to measurement, and highlights that measurement itself can be thought of as a process of selection of an element from a previously chosen set. Furthermore, the mathematical theory of information guarantees that extending the set of values does not reduce – and usually in fact increases – the average quantity of information, evaluated as an entropy, conveyed by an evaluation [9].

A further general condition characterizing measurement as a specific kind of evaluation relates to the entity under measurement. In fact, very diverse entities can appear as the input of the process: the VIM3 writes about “phenomena, bodies, or substances”, and by widening the context also pieces of software, individuals, industrial processes, organizations, and any perceivable thing can be considered as such. Let us introduce the term *object under measurement* to denote them. As a basic fact, the relation between objects under measurement and measurement results is generally many-to-many, i.e., the same object can be evaluated to different results, and different objects can be evaluated to the same result. The first condition is interpreted by assuming that any object has multiple aspects, i.e., properties, and that subject to measurement is not the object as such, but a property of it. Hence, measurement is an informative *property evaluation*.

This basic assumption is fully consistent with the functional modeling as far as such *general properties* (such as electrical resistance, pain intensity, attitude to scientific research) are superposed to the mappings, according to the logic:

$$property_i(object_j) = result_{ij}$$

where $property_i(object_j)$ represents the single instance of a general property (such as the resistance of this resistor, the intensity of the pain felt by this patient, the attitude to scientific research of that PhD student) for the given object. Therefore, measurement assigns a symbol – the measurement result – to the object under measurement, so that the symbol intends to provide descriptive information on the current state of the object with respect to one of its properties, called the *measurand*, i.e., the property intended to be measured [4].

While this general characterization leads to some non trivial consequences, what has been presented so far hardly can be accepted as sufficient to define measurement: some further conditions are required.

3. MEASUREMENT AS OBJECTIVE AND INTER-SUBJECTIVE EVALUATION

The fundamental intuition that measurement is a property evaluation whose results convey reliable

information on the measurand is not related to the nature of the object under measurement or of the measurand, nor to the algebraic structure of the set of property values, and in principle not even to the structure of the measurement process itself. Rather, such reliability can be characterized in terms of two general features expected for measurement results, which are supposed to convey:

- information specific to the measurand, and therefore to a given property of the object under measurement. This means that the provided information should be independent of any other property of the object or the surrounding environment, which includes both the measuring system and the subject who is measuring. This corresponds to guaranteeing that measurement results actually provide information about the measurand and not of some other property. It is a sometimes trivial, sometimes very complex to satisfy, condition about the appropriate attribution of information to its claimed object: hence, it is a requirement of *objectivity*;
- information interpretable in the same way by different users in different places and times. This corresponds to guaranteeing that measurement results are expressed in a form independent of the specific context and only referring to entities which are universally accessible, so that the meaning of a measurement result is unambiguous and can be easily reconstructed in principle by anyone, possibly on the basis of suitable conventions: hence, this feature expresses a requirement of *inter-subjectivity*.

Physical measurements usually embed these features directly in the structure of the measuring instrument, designed and operated so to behave as a transducer whose empirical output, called the instrument *indication*, ideally should depend only on the measurand, or a property functionally related to it, thus assuring that the information it provides relates specifically to the object under measurement, and therefore confirming objectivity. Moreover, the instrument indication should be mapped to measurand values through instrument calibration, which makes the instrument output traceable to a primary measurement standard. Hence, different instruments traced to the same standard provide comparable information so assuring inter-subjectivity. On the other hand, such conditions of objectivity and inter-subjectivity do not imply any specific constraint on the realization of the measuring process, so that in principle they can be introduced as requirements also in the case of the evaluation of non-physical properties.

These two features are independent with each other, in the sense that an evaluation might produce objective but non inter-subjective results (as in the case of the usage of an uncalibrated measuring system), or vice versa inter-subjective but non objective results (e.g., if they are expressed in the customary format for quantity values, i.e., number times measurement unit, but they have been obtained just at random). This is the reason why both of them are considered to be required here ([10] proposes to call “pre-measurement” an objective but still non inter-subjective evaluation).

For the condition of inter-subjectivity to be fulfilled, measurement results are required to convey information not only on the measurand value, but also, reflexively, on the actual trust attributed to that value. The classical assumption that quantity values are real-valued numbers (based on the hypotheses that “numbers are in the world”, as Kepler once wrote along the Pythagorean tradition, and that many physical phenomena are continuous or at least their variation is such) hides this requirement, under the wishful, but wrong, thinking that progressively increasing the quality of measurement would reduce, at then eliminate, all measurement errors, and the “true value” of the measurand would be discovered accordingly. Anyway, even if measurement might be hindered by errors (e.g., using a measuring instrument which is no longer calibrated), an appropriate inter-subjective communication of measurement results calls for a statement about the *degree of belief* that the measurer attributes to the provided information. On this basis the position emerged that a measurement result should specify the amount of information it is aimed at providing, and thus that it is not complete if the *uncertainty* of the stated measured property value is not included. Specifically, the VIM3 defines ‘measurement uncertainty’ as a “non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used”. Measurement uncertainty is thus an overall concept, which synthesizes the effect on the measurement result of multiple contributions (including, where applicable, measurement errors), from both the experimental and the modeling steps of the measurement process, into a single piece of information.

In practice neither objectivity nor inter-subjectivity are Boolean (i.e., yes-no) features. Indeed, the result of any measurement is affected by the way the measuring system was calibrated and it conveys information not only on the measurand but also, unavoidably, on other properties of the surrounding environment. The fact that these two features admit intermediate levels emphasizes the need of introducing a further pragmatic component in measurement: the threshold over which the evaluation results are “sufficiently” objective and “sufficiently” inter-subjective to be considered measurement results is set in reference to the expected use of the provided information, typically the support to a decision making process. This perspective agrees with the so called goal-question-metric paradigm [11] and considerations reported in the ISO 9000:2005 standard [12]: any measurement process must first consider the intended use of the experimentally obtained

information, in order to satisfy requirements predefined by the final user of that information. It is worth noticing that the evaluation of measurement uncertainty alone does not include the definition of thresholds and thus it may lead to different interpretations by varying the context: as an example, while measurement uncertainty of 10 kg could be considered unacceptable in several contexts, it might be fit when it comes to extraction operations in mineral engineering.

In particular, expressing uncertainty as a suitable standard deviation, as recommended by the GUM, allows comparing the quantity of available information to the minimum quantity assumed by design as needed for effectively supporting the decision making, which is called the *target measurement uncertainty* (“measurement uncertainty specified as an upper limit and decided on the basis of the intended use of measurement results”, according to the VIM3). Hence, target measurement uncertainty might be thought of as a quantitative means to formalize the concept of “sufficient” objectivity and inter-subjectivity with respect to the decision under consideration. In a favorable context, in which decisions can be actually made with the support of measurement results, the relations among the chained concepts of defined-by-specification (as expressed, e.g., by a nominal value and a tolerance for the property under consideration), required-for-decision, and experimentally-obtained quality of information can be then depicted as in Fig.1.

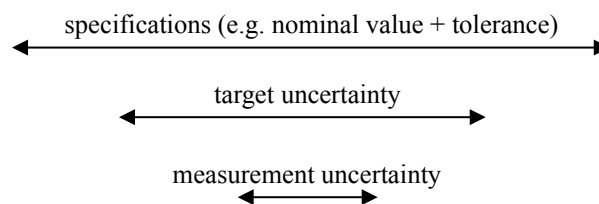


Fig.1. Measurement uncertainty as an essential tool for decision making

As a consequence, the pragmatic component of measurement could be stressed up to a point that an evaluation producing results whose uncertainty is greater than the required target measurement uncertainty would not be acknowledged to be a valid measurement, because in fact its results are not useful to support its intended use. When measurement results are exploited at a later time and possibly in a different context, information about their intended use might not be available at measurement time. This conclusion is still applicable and measurement uncertainty still provides a means to judge the quality of a measurement by comparison with pre- or post-defined thresholds: also in these cases, if targets of uncertainty thresholds cannot be met, measurement results cannot be used.

Concerning the expression of measurement uncertainty, it could be also noted that the extension from quantities with unit to generic properties generates a (mainly still open) issue: indeed, the GUM framework does not apply to ordinal quantities, nor as a consequence to nominal properties, since mean value and standard deviation are not empirically meaningful for them [13]. While other, algebraically weaker, statistics are available and well-known (e.g., median and percentiles in the ordinal case), a widespread agreement on this subject is still to be reached. A plausible backgrounder might be the mentioned Shannon’s theory of information, which builds on the assignment of probability density functions to value sets and allows interpreting uncertainty in selection of values by means of an entropy function, from singletons (no uncertainty, null entropy) to uniform distributions (complete uncertainty, maximum entropy).

4. MEASUREMENT AS A MODEL-BASED AND OPERATIVE EVALUATION

The customary structure of a physical measuring system assumes that by means of measurement an empirical entity (the measurand) is represented by an information entity (the measurand value including its uncertainty). This effect of bridging two different “worlds” (as for example K.R. Popper called them, the “World 1, of physical entities” and the “World 3, of products of human minds” [14]) results from a two-stage process:

- a *modeling* stage, that is a set of conceptual activities needed to define the measurand and to model its relationship with the measurement result; this is achieved through a proper modeling of the whole experimental environment, called *measurement context* in the following, in which measurement is expected to be performed, thus including the object under measurement, the measuring system, the subject who is measuring, and the empirical surrounding environment;
- an *operative* stage, in which – on the basis of the knowledge about the measurement context provided by the modeling stage – experimental activities are performed, so implementing the mapping between the

empirical world and the world of symbols.

In its turn, the operative stage is organized as a two-step process:

– an *experimental step*, in which the measuring instrument interacts with the object under measurement and as a result of the interaction it produces an observable output, i.e., an instrument *indication*, which is (ideally) caused by the measurand;

– a *representational step*, in which from such indication the cause which produced it is reconstructed and symbolized, so properly assigning a value, and generally a related uncertainty, to the measurand; this activity is performed by using the information about the measurement context achieved at the descriptive level and, in particular, the information provided by the calibration process of the measuring system.

These two stages allow implementing, and in any effective measuring system are expected to implement, the requirements of objectivity and inter-subjectivity of measurement results.

In the simplest case, the measuring instrument input is the measurand, so that the two operative steps are sometimes considered as inverse with one another: the experimental step instances the transduction function, which is then inverted in the representational step which, for this reason, is sometimes called *measurand reconstruction*. According to this perspective, the whole operative stage of measurement is said to implement an identity function. This position is in fact wrong because of a wrong ontological superposition: the measuring instrument input is the measurand, i.e., a property, not its value.

This basic case is generalized by admitting the measuring instrument not to be perfectly selective, i.e., the indication it produces depends not only on the measurand but also on other influence properties, so that the mapping from indications to measurand values must also include the information on the effect of such influence properties, and properly correct or compensate it.

This description of measurement as a modeling and operative process is sketched in the diagram of Fig.2: once the model of the measurement context has been obtained, the measurand is acquired in the experimental step, in which the measuring instrument produces an indication that may depend on influence properties. Also the state of the object under measurement – and consequently the measurand – may be affected by influence properties. Finally, on the basis of the information contained in the model of the measurement context, the values measured for the influence properties and the information obtained in calibration, the representation stage handles the transducer indication and the known values of influence properties to provide the measurement result.

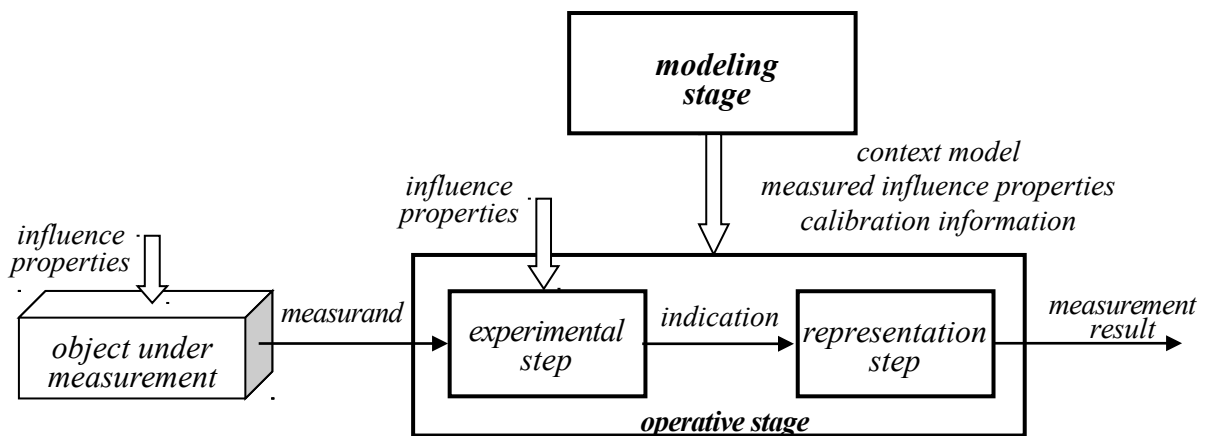


Fig.2. Measurement as a two-stage process

The further, and most significant, generalization introduces the hypothesis that the property subject to measurement and the property intended to be measured (i.e., the measurand) may be different, and therefore that the input property of the transducer carries information on a property which is different from the measurand. This difference can take into account the modification induced on the object under measurement by its interaction with the measuring instrument, and as such it may be corrected by suitably defining the measurand reconstruction function. This is typically the case in which the measurand (e.g., the intensity of electrical current in a given circuit) and the property actually subject to measurement (the intensity of electrical current in the circuit when coupled with a given measuring instrument) are instances of the same general property input of the transducer (intensity of electrical current). On the other hand, such two entities might be instances of different general properties, as when the property subject to measurement is a current

and the measurand is a voltage. In these cases, traditionally called *derived* (or indirect) *measurements*, the concept of measurand reconstruction is complex, since it has to include the relation between the two general properties (e.g., as expressed by a physical law; in the example, the Ohm's law), and thus at least some aspects of a domain theory.

These cases are usual in soft measurement, where the measurand is sometimes a very complex entity, such as the attitude of a given individual to a given task or the performance of a given R&D department, and the information on it is obtained via a set of *indicators*, the properties actually subject to measurement. A substantive theory is then crucial to justify that the searched information on the measurand can be derived from such indicators [15, 16]. Whenever such a theory is not available, not an uncommon case in soft measurement, measurement validation becomes a critical issue: an objective support has to be given to the claim that measurement results actually refer to the stated measurand. Common approaches to validation in this case are based on the use of statistical correlation techniques and cause-effect analysis.

The influence of both the measuring system and the modeling stage on the information provided by the measurement result will be analyzed in the following Sections. When dealing with modeling activities, the specific case of research quality measurement will be considered in order to support the discussion with a hot topic in many academic and research institutions, whose proper application could benefit the knowledge of measurement fundamentals.

5. INFLUENCE OF MEASURING SYSTEM ON MEASUREMENT UNCERTAINTY

The availability of a measuring system, i.e., a coordinated set of one or more measuring instruments, together with a measurement procedure, i.e., a detailed description of how such system has to be operated to obtain measurement results, is the customary condition assumed for performing measurement. Hence from the metrological characteristics of such systems and the knowledge of the measurement procedure, some critical components (and sometimes all components) of the uncertainty budget are derived.

A complete metrological characterization of any specific class of measuring systems requires the specification of multiple features (e.g., sensitivity, selectivity, resolution, dead band, drift, response time, frequency response, measuring interval, operating conditions, energy requirements, lifetime, ...), which can be classified according to different criteria (e.g., features related to dynamic vs. static behavior). On the other hand, we claim that at a very fundamental level the system behavior is characterized by a single general feature: its *stability*, i.e., the "property of a measuring instrument, whereby its metrological properties remain constant in time". This requirement of constancy in time applies in reference to two complementary kinds of measurement conditions (definitions taken again from the VIM3):

- the *repeatability* condition of measurement, i.e., the "condition of measurement, out of a set of conditions that includes the same measurement procedure, same operators, same measuring system, same operating conditions and same location, and replicate measurements on the same or similar objects over a short period of time";
- the *reproducibility* condition of measurement, i.e., the "condition of measurement, out of a set of conditions that includes different locations, operators, measuring systems, and replicate measurements on the same or similar objects".

Under repeatability conditions the short-range stability of the system behavior is assessed. It is a check which can be performed independently of any calibration: with the only assumption that the (even unknown) input property to the transducer and the influence properties are not (significantly) changing during the repetitions, a sample of indications is obtained and the information it conveys can be expressed by means of a scale (or dispersion) statistic [17], such as the sample standard deviation. Hence, such short-range stability is related to measurement *precision*, i.e., the "closeness of agreement between indications or measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions" (VIM3 definition), where the "specified conditions" are repeatability conditions.

In a complementary way, under reproducibility conditions (but including a single measuring system), the long-range stability of the system behavior is assessed. If the check is performed after the transducer has been calibrated, its capability of maintaining the calibration state, as stated in the calibration diagram, is verified: a "reference" indication, as obtained while calibrating the system, is compared with a suitable location (or position) statistic [17], such as the sample mean value, computed on the sample of the repeated indications. Hence, such long-range stability is related to measurement *trueness*, i.e., the "closeness of agreement between the average value obtained from a large series of test results and an accepted reference value" (definition from [18], where "test results" is interpreted as "indications"; the VIM3 definition, "closeness of agreement between the average of an infinite number of replicate measured quantity values and a reference quantity value" is practically unusable since an infinite number of measurements is required).

Accordingly, measurement precision and measurement trueness can be interpreted as the complementary components of an encompassing feature of a measuring system, its *accuracy*, which conveys the basic information required to evaluate the *instrumental measurement uncertainty*, i.e., the “component of measurement uncertainty arising from a measuring instrument or measuring system in use” (VIM3 definition).

6. INFLUENCE OF MODELING ACTIVITIES ON MEASUREMENT UNCERTAINTY: MEASURING RESEARCH QUALITY

As an application example of the effect of modeling activities on the validity of measurement results, let us consider measurement of research quality. This concept is somewhat elusive but it evokes the need of quantifying as much as possible the outcomes of a research process. This is an issue of great social and economic impacts in nations. Deciding about the fundability of research groups or programs or about the *ex-post* effectiveness of governmental funding policies is steadily becoming a mechanism to show accountability in spending public money. At the same time, while scientists know the value and worth of peers, there is an information asymmetry between the scientific world and the general public that can be addressed by using suitable and possibly well-designed measurement procedures. Whichever the model, a well-designed research quality measurement process originates from a consensus-based definition of research quality. This has to be reached at least in the context where the decisions based on measurement outcomes have effects and implies agreement between stakeholders. For example in the research assessment exercise [19], research has a formal definition while the concept of ‘research quality’ is implicitly defined by the indicators chosen as quality determinants. This is similar to what is intended as software quality in the ISO 9126-1 standard [20] that is by using a set of suitably chosen indicators the characteristic is indirectly determined. Moreover, according to what has been suggested above any measurement process must firstly take into account the intended use of the experimentally acquired information. As an example, if research quality is measured for promoting research personnel or for deciding about fundability of research programs, different sets of indicators have to be chosen for properly support the subsequent decision making activities. For wide-range research quality measurements, where the final customer is the general public and the society at large, the indicator set does not contain too specific attributes, but rather generic ones. Among others, a specified requirement in this latter case is often the determination of research impacts within the society. The research environment also plays a role: an environment rich in industrial activities most probably supports scientific and technological research by means of economical resources and creative stimuli. All mature research quality models try to measure also this attribute by introducing indicators related to environment. The measurement procedure designed to achieve the predefined goals results in the definition of a multiattribute model that contains both qualitative and thus judgmental information, and quantitative indicators. This is the case for the many research quality frameworks and assessment procedures proposed in scientific advanced countries. In all proposed models, the measurand is multidimensional and often comprises attributes related to outcomes (e.g., publications, patents), impacts (e.g., amount of social and economical benefits) and environment (e.g., types and level of facilities in the research environment). These main attributes include sub-attributes that are measured by informed experts using an ordinal scale, sometimes treated as a quantized ratio scale, so that weighted summations are calculated at a later stage of data processing. Bibliometric data also come often into play to allow judgments based on indirect sources. Models include the characterization of both enabling factors, such as the amount of available facilities, and results, such as the number of excellent research products, as done for the assessment of organizational quality in the excellence model proposed by the European Foundation for Quality Management [21]. In any case, each indicator is characterized by uncertainty that is often hard or non-economical to estimate and thus is not available or not communicated. Therefore, process validation is required to assure that the intended use of the produced measurement results is achieved. Research quality models are often modified, both to reflect changes in views about the importance of single attributes as expressed by their corresponding weights, and to reduce the consequences of adaptive behaviors in researchers adopting opportunistic strategies.

Does this approach lead to a measurement, as seen from the perspective of the requirements that such a procedure must obey? The representation condition – which requires research of *better quality* to be measured so that symbols/numbers assigned to it preserve this order – might not be properly satisfied. In fact, although experts are trained so to develop discriminative capabilities in evaluating research quality determinants and definitions of research quality attributes are refined to reduce semantic uncertainties, the representation condition could be not fulfilled. Similar considerations apply also to the concepts of objectivity and inter-subjectivity. The level of comprehension in the multidimensional concept of research quality could not assure a proper objectivity, that is the evaluation of indicators measuring attributes and sub-

attributes is not guaranteed to capture the entire essence of the measurand and nothing else. As a counterexample, it is well known that confusing factors may influence prejudicially the evaluation by experts, as in the case of the Matthew effect [22], a phenomenon by which the evaluation of research outcomes or proposals is influenced significantly by the reputation of the evaluated subject. This may lead to increased probability of fundability of high-reputation researchers, thus leading to over-funding already rich research subjects.

Regarding inter-subjectivity, it can be observed that while harmonization sessions may be performed beforehand to guarantee that experts consistently respond in similar ways to similar stimuli, a higher variability in responses can be expected than what is usual in other quantitative disciplines, when different subjects perform the measurement. The concepts of repeatability and reproducibility all have to do with the uncertainty attributed to measurements and with the replication of the experiments under specified conditions. Uncertainty under repeatability conditions may be considered to be low because experts will give equal or similar judgments if asked again over short periods of times. Similar considerations apply to the concept of reproducibility, here to be intended as the precision in measurement obtained when possibly different experts evaluate different subjects performing research with nominally equal quality attributes. The amount of agreement between evaluations will provide information about reproducibility. The difficulty here is associated with the replication of the measurand in different experimental contexts and again with the harmonization of evaluation styles by experts. The concept of calibration is too loose in this context to be applied, also because universal standards or references can hardly be defined in measuring research quality. Although many attributes are applicable only marginally with respect to what happens in hard measurements, measuring research quality using this or similar approaches may provide valid information. This seems to be confirmed indirectly by the fact that some nations are using research evaluation processes recurrently and by means of validation sessions performed using correlations based on citation and peer-review analyses [23]. The issue whether such an evaluation process is a measurement cannot be answered in a distinct and clear way and consensus on this subject is not obtained among stakeholders. In a strict sense, the fact that uncertainty estimates are not available and that the desirable measurements characteristics can hardly be achieved suggest that this procedure is likely a generic evaluation rather than a measurement. However, the needed degree of belief in statements about research quality may be conveyed not only by explicitly stated uncertainty, as suggested in Section 3 but, when this is missing, also through the reputation of the measurer (e.g., as in the case of national agencies) and by the degree of rigor, transparency and discipline exhibited by carrying out the evaluation process.

7. CONCLUSIONS

Discussion about measurement fundamentals is being continuously enriched because of the new challenges offered by advancements in science and technology. Understanding differences in hard and soft measurements characteristics helps in the adoption of the best behaviors when trading-off economic and engineering measurement strategies and adds to progress in science by fostering model recognition and consensus among interested parties. The pragmatic approach described in this paper starts from the recognition that soft measurements ask for additional analysis about models and properties to produce meaningful experimental results. We proceeded accordingly by addressing the general definitions of measurement characteristics adopted when taking hard-measurements and by analyzing their applicability in the realm of soft-measurements. As a by-product of this analysis, an improvement of the body of knowledge about fundamentals of any measurement process has also been achieved.

Interesting reasoning opportunities have arisen to find similarities and interpretative viewpoints to be applied to specific soft multi-attribute measurements. Still, some key points remain open for discussion: validation, uncertainty, calibration and other fundamental measurement concepts hardly apply to soft measurements. Nevertheless, measurements of non-physical properties are becoming widespread and decisions are taken on this basis, having economic impact. Our view is that these may return valid information and that rigorous, disciplined and transparent modeling and operative activities are still the principal mechanisms to guarantee the final user about the reliability of the provided information.

References

- [1] L. Finkelstein, "Widely, strongly and weakly defined measurement", *Measurement*, vol.34, pp.39–48, 2003.
- [2] Bureau International des Poids et Mesures, *Evolving needs for metrology in trade, industry and society and the role of the BIPM* (Kaarls report), <http://www.bipm.org/utis/en/pdf/Kaarls2007.pdf>.
- [3] JCGM 100:2008, *Evaluation of measurement data – Guide to the expression of uncertainty in*

- measurement* (GUM, originally published in 1993), Joint Committee for Guides in Metrology, 2008, <http://www.bipm.org/en/publications/guides/gum.html>.
- [4] JCGM 200:2008, *International Vocabulary of Metrology – Basic and general concepts and associated terms* (VIM), 3rd Edition, Joint Committee for Guides in Metrology, 2008, <http://www.bipm.org/en/publications/guides/vim.html>.
- [5] G. B. Rossi, “Measurability”, *Measurement*, vol.40, pp.545–562, 2007.
- [6] D. Krantz, R. Luce, P. Suppes, A. Tversky, *Foundations of Measurement*, vols. 1–3, Academic Press, 1971, 1989, 1990.
- [7] F. Roberts, *Measurement theory: with applications to decision making, utility, and the social sciences*, Addison-Wesley, 1979.
- [8] L. Mari, “Beyond the representational viewpoint: a new formalization of measurement,” *Measurement*, vol.27, pp.71–84, 2000.
- [9] L. Mari, “Notes towards a qualitative analysis of information in measurement results”, *Measurement*, vol.25, pp.183–192, 1999.
- [10] A. Frigerio, A. Giordani, L. Mari, “Outline of a general model of measurement”, *Synthese*, vol.175, no.2, pp.123–149, 2010.
- [11] V. R. Basili, G. Caldiera, H. D. Rombach, *The Goal Question Metric Approach*, Encyclopedia of Software Engineering (J. J. Marciniak, Ed.), John Wiley & Sons, 1994, Vol. 1, p. 528-532.
- [12] ISO 9000:2005, *Quality management systems – Fundamentals and vocabulary*, International Organization for Standardization, Geneva, Switzerland, 2005.
- [13] S. S. Stevens, “On the theory of scales of measurement”, *Science*, vol.103, pp.677–680, 1946.
- [14] K. R. Popper, *Knowledge and the body–mind problem – In defense of interaction*, Routledge, 1994.
- [15] L. Finkelstein, “Problems of measurement in soft systems”, *Measurement*, vol.38, pp.267–274, 2005.
- [16] L. Mari, V. Lazzarotti, R. Manzini, “Measurement in soft systems: epistemological framework and a case study”, *Measurement*, vol.42, pp.241–253, 2009.
- [17] ISO 3534-1, *Statistics – Vocabulary and symbols – Part 1: General statistical terms and terms used in probability*, International Organization for Standardization, 2006.
- [18] ISO 5725-1, *Accuracy (trueness and precision) of measurement methods and results – Part 1: General principles and definitions*, International Organization for Standardization, 1998.
- [19] REF2014, “Assessment framework and guidance on submissions”, July 2011, http://www.hefce.ac.uk/research/ref/pubs/2011/02_11/02_11.pdf.
- [20] ISO/IEC 9126-1:2001, *Software Engineering - Product quality – Part 1: Quality model*, International Organization for Standardization, 2001.
- [21] *The EFQM Excellence Model*, <http://www.efqm.org>.
- [22] R. K. Merton, “The Matthew Effect in Science”, *Science*, vol.159, pp.3810, Jan. 5, 1968.
- [23] C. Oppenheim, M.A.C. Summers, “Citation counts and the Research Assessment Exercise, part VI: unit of assessment 67 (music)”, *Information Research*, vol.13, no.2, June 2008.