Chapter 6
Quality of Measurements

6.1 Introduction

Measurements are basic tools in any scientific investigation. Many exercises in Science and Technology are aimed at improving the existing state of affairs regarding matter, energy, environment and their interactions—among themselves as also with living organisms. One is reminded of a widely quoted statement made by a twentieth-century German philosopher who runs as follows (Mukherjee)

If I can define it, I can measure it.
If I can measure it, I can analyze it.
If I can analyze it, I can control it.
If I can control it, I can improve it.

Measurements are needed in all scientific investigations to **choose, develop** and **validate models** and used to describe, analyse **predict, control** or **improve** various phenomena. Measurements provide the very basis of all control and improvement actions. Incidentally, one way to differentiate between Science and Technology—if at all one needs to—is to argue that Science is more concerned with *Definition, Measurement* and *Analysis,* while Technology is more engaged in *Control a*nd *Improvement.* However, this differentiation may not be warranted in all cases.

While 'improvement' of any existing 'state' is always our goal, we must remember that 'improvement' comes only at the end of a sequence of actions or process to define the 'state' in an objective manner, to measure the state, to analyse the state in terms of its determinants and correlates and, subsequently, to control the state at a desired level.

In the context of Quality Management, measurements are involved right from quality planning through on-line and off-line quality control and quality assurance to customers and other stakeholders to quality improvement. In a sense, measurements pervade the entire Deming Cycle in terms of Plan-Do-Check-Act operations. And, we have to choose appropriate measures of quality (of incoming materials,

S. P. Mukherjee, *Quality*, India Studies in Business and Economics,
https://doi.org/10.1007/978-981-13-1271-7_6

processes, in-process materials, checks and controls, finished products, etc.) and subsequently carry out measurements on physical or chemical or other features or characteristics of the different entities. In case such features and characteristics are not directly measurable, we have to develop suitable proxy measures.

In both the above situations, we need to speak of Quality of Measures (or measurands which are to be measured) as well as Quality of Measurements which are outcomes of the Measurement Process carried out on units of concrete entities. In fact, measures—of productivity, efficiency, dependability, organizational excellence, people orientation, customer satisfaction and similar other concepts—are all based on and derived from several related measurements. It should also be appreciated that any Measure of Quality of a product or process or service entails measurements on a number of pertinent quality characteristics.

While Quality of Measurements has been discussed a lot in recent times, the priority needs to comprehend Quality of Measures and deploy 'good' quality measures for assessing performance has not been fully addressed. A very important consequence of performance is customer satisfaction and, like other latent variables, customer satisfaction does not admit of a unique definition and, obviously, a variety of constructs, models and methods have been in use in various quarters.

We first take up Quality of Measurements and thereafter Quality of Measures, though the reverse order would have been more logical. This has been partly motivated by the fact that national as also international standards have been developed on Quality of Measurements—not, of course, under this nomenclature—and are being used by many laboratories in industries and research organizations.

Quality of Measurements has been discussed in different contexts by different authors and agencies. A good number of national, regional and international standards have been developed over the years to promote the application of consensus definitions and measures. Even methods to estimate these measures from repeat measurements have been standardized. Mukherjee (1996) presented the concepts, measures and models relating to quality of measurements in his Platinum Jubilee Lecture in the Section of Statistics in the Indian Science Congress Association and the material that follows is based largely on the content of that lecture.

6.2 Measurements in Quality Management

Measurements play an important role in

- identifying opportunities for improvement (e.g. through measurements of quality cost under different heads like appraisal, prevention and failure (internal as also external) costs) and
- comparing performance of different processes against internal standards (as in process control and improvement) as also comparing performance of processes and results thereof against external standards (as in benchmarking).

The Deming Cycle of continuous improvement—Plan, Do, Check, Act, sometimes modified as Plan, Do, Stabilise and Act—clearly requires measurements to drive it, and yet it is a useful design aid for the measurement process itself. In this cycle, the words, Plan, Do, Check and Act, have been explained in somewhat different ways by different users of this approach. A detailed note appears in Chap. 9. Usually, we accept the following elucidations.

Plan—Establish performance objectives and standards for processes, their inputs and outputs.
Do—Measure actual performance in terms of time taken, quality of output, cost incurred, etc.
Check—Compare actual performance with the objective(s) and standards and determine the gap in between.
Act—Take necessary corrective action(s) to close the gap and make necessary improvements.

It has been often said that it is not possible to manage what cannot be measured. To comprehend a quality or a productivity problem fully in terms of its intensity, frequency of occurrence and consequences, we need to collect measurements on the problem. Even after a provisional solution to such a problem has been developed, we need some measurements on trial runs of the proposed solution before we can establish its effectiveness including economic considerations. Relevant measurements have to be collected and analysed to

* rate vendors for their capability to meet our requirements regarding quality, delivery and price and assess performance of vendors selected on the basis of such a rating

- meet customer requirements as mutually agreed upon;
- set sensible objectives and targets and to compliance with them;
- provide standards for establishing comparisons of performance;
- ensure visibility in terms of a scoreboard for people to monitor their own performance levels against corresponding targets;
- identify quality and productivity problems and prioritise those (in terms of time required and gains expected) for corrective and preventive actions;
- work out costs of poor quality, broken down inti pertinent components;
- determine resource needs objectively, including needs for test, inspection and measuring equipments;
- provide feedback for assessing the improvement exercise. And providing directions for desired modification in the same;
- identify appropriate tools and softwares for carrying out necessary quantitative analysis, keeping in mind the necessity of simplicity in use and of economy.

In order to assess and evaluate process performance as also results accurately, appropriate measurement must be designed, developed and systematically documented by people who own the processes concerned. They may find it necessary to measure effectiveness, efficiency, quality, impact and productivity. In these areas, there are many types of measurement, indirect or direct output or input figures, costs of poor quality, economic data, comments and complaints from customers, information from customer or employee surveys about the extent to which they feel satisfied, etc.

6.3 Panorama of Measurements

Measurement is a process that follows a defined sequence of steps/activities involves physical, material and technological resources and results in a numerical value/a set of numerical values that is assigned to an item in respect of a defined property/parameter/characteristic.

Like any other process, measurement process involves both hard and soft inputs, is carried out under some influencing factors as also some controls and checks, and produces some (soft) output. The hard input is the concrete object on which a measurand has to be numerically evaluated, while the prescribed method of measurement along with the conditions under which the process has to be carried out define the soft input. Ambient conditions of temperature, pressure, humidity, wind velocity, vibration, electromagnetic interference, etc., are some of the influencing parameters. The process is controlled by checks carried out on measuring instruments for their stability, sensitivity, etc., and these are calibrated, as and when necessary. The output is a numerical value or a set of such values which can be ascribed to the input object. Measurement is also the output of a (measurement) process—some numerical value(s).

Measurement is a generic term and the panorama of measurements is enthralling. Measurements are derived from a wide spectrum of sources. In the case of direct measurements, the source is a measuring device in contact with the object being measured in respect of a certain parameter or characteristic. However, photographs, images of various sorts, satellite imageries, etc., are also important sources of measurement.

We have very large measurements like those on interstellar distances (in billion light years) to microscopic measurements of intermolecular separations in solids. On the one hand, we talk of a micro-level measurement like the concentration of a suspended particulate matter in the atmosphere over, say, a paddy field while inter-regional disputes arise over total stocks of such matters in a whole region.

Quite often, macro-level measurements like the latter are obtained by multiplying small micro-level measurements taken on much smaller units. It is not difficult to realize that even a minute error in the micro-level measurement gets largely magnified in the macro-measurement.

Sometimes, the reverse procedure is followed to derive the micro-level measurement as the quotient of a large macro-measurement divided by a usually large number of units. This is the case with, say, per capita national income or per capita annual consumption of active substance extracted from nature.

There are many other distinguishing features of measurements. Thus, we have exact measurements as are yielded by some measuring devices against approximations or estimates. The latter not only correspond to rounding off of measurements to a desired order of accuracy, but also relate to situations where exact measurements are ruled out and estimates have to be made on the basis of limited measurements on related entities and/or some assumptions.

For example, we can speak of the exact quantity of coal raised from a pit and can offer only an estimate of the total exploitable reserve of coal in a coalfield.

It may be of some interest to note the recent revision of the Indian Standard Rules for Rounding off of Numbers, requiring rounding off in one direction only in situations where safety or similar other considerations are expressed in terms of one-sided tolerances/permissible limits.

As Mukherjee (1996) pointed out, measurements carry the charisma of objectivity and there has been a growing tendency among investigators to use measurements as bases for arguments—for and against. It should be remembered that subtle, subjective (individualistic) behaviours, attitudes, aspirations, aptitudes, and similar traits studied in social sciences do not strictly admit of unique measurements.

Though uniqueness and objectivity are not synonymous attributes of measurements, they are quite akin to each other. Hence, the use of measurements in unfolding the vectors of the human mind or in related matters should not be downright denounced, but should be taken with due caution.

6.4 Errors in Measurements

Errors in the observed results of a measurement (process) give rise to uncertainty about the true value of the measurand as is obtained (estimated) from those results. Both systematic and random errors affecting the observed results (measurements) contribute to this uncertainty.

Random errors presumably arise from unpredictable and spatial variations of various influence parameters operating on the measurement process, for example:

- the measurement method employed in case it is not a standard method or has not been validated against a standard method;
- the way connections are made or the system configuration is worked out;
- uncontrolled environmental conditions or their influences;
- inherent instability of the measuring equipment;
- personal equation bias of the observer or the operator;
- judgment or discretion used by the observer or operator in securing the value of the measurand from readings on the instrument, etc.

These cannot be eliminated totally but can be reduced by exercising appropriate controls.

Various other kinds of errors, recognized as systematic, are also observed. Some common types of such errors are

- those reported in the calibration certificate of the reference standards/instruments used;
- those due to different influence conditions at the time of measurement compared with those prevalent at the time of calibration of the standard (quite common in length and direct current. measurements), etc.

It should be pointed out that errors which can be recognized as systematic and can be isolated in one case may simply pass off as random in another case.

6.5 Quality of Measurements

Quality of measurements is comprehended in terms of Accuracy and Precision, based on systematic and random errors respectively that get reflected in repeat measurements. A more recent development takes care of both random and systematic errors and results in a measure of uncertainty about the true value. The Indian Standard IS 5420 Part I describes and illustrates the procedures for calculating accuracy, repeatability and reproducibility of test results. These measures which are linked up with errors in measurements have been quoted, illustrated and explained in several other sources. One can refer to the document by Kelkar which has been posted on the Website of the Maharashtra Pollution Control Board www.mpcb.gov.in.

Accuracy is the critical parameter and is not the same as precision. Accuracy is closeness to the true value (of the measurand), while precision implies consistency among repeat measurements (not always available).

Let X be the measurand and $x_1, x_2, \ldots, x_n$ be n repeat measurements (on the same object or on exactly similar objects in case measurement involves a destructive test or determination) carried out in the same laboratory, using the same equipment, by the same operator, in the same environment.

Let $\bar{x} = \frac{1}{n}\sum x_i$ be the mean of the repeat measurements and $s^2 = [1/(n-1)\sum(\bar{x}-x)^2]$ be the variance. Also, let T be the true value of the measurand. Then, $|\bar{x}-T|$ is an inverse measure of accuracy, while the standard deviation provides an inverse measure of precision or internal consistency among repeat measurements. Accuracy has a bearing on the equipment while precision has a bearing on control over repeat measurements

Precision is generally measured and reported in terms of

(1) repeatability and (2) reproducibility

Define a quantity r such that

$$\text{Prob.}\{|x_i - x_j| > r\} < \alpha \quad \text{for any } i = j$$

where α is a pre-assigned small quantity, e.g. 0.05 pr 0.01 Then, r is referred to as the repeatability factor.

In case the repeat measurements were produced in different laboratories (obviously involving associated differences in equipment, operator and environment) and are denoted as y_1, y_2,…y_n, we could define a quantity R such that

$$\text{Prob.}[\,|\,y_i - y_j\,| > R] < a.$$

This R is referred to as the reproducibility factor. Factors r and R can be estimated as $r = k_1 s_x$ and $R = k_2 s_y$ where k_1 and k_2 can be determined from the distribution of x or of y. These measures really characterize a measurement process, though these are also used to interpret variations in measurements. These do not involve the unknown true value (μ) of X and hence do not give an idea about the possible range of true values associated with a single measurement (x_i) or a mean value ($\bar{x}$) of several repeat measurements. Indian Standard IS 5420 Part 1 prescribes a common value 2.77 as the value for k_1 and k_2.

The purpose of a specification (one-sided or two-sided) is to fix a limit for the true value of the property concerned. Given that the true value cannot be established in practice, the results will reveal some scattering due to repeatability or reproducibility, depending on the situation. Accordingly, it will be desirable to evolve and accept specification limits taking due account of repeatability and reproducibility of the test method. Thus, the specification range should be equal to at least 3 R so that the upper (lower) specification limit is more (less) than 1.5 R from the nominal value specified or intended and some inherent variability if measurements are recognized the case of one-sided specification can be similarly dealt with.

6.6 Measurement System Analysis

The term 'measurement system' refers to the collection of instrument/equipment, operations or processes, procedures, people and software (if involved) which affect the outcome of a measurement process or the assignment of a numerical value to a measurable property (measurand). Measurement system analysis (MSA) is concerned with five parameters viz. Bias (inverse to accuracy), Linearity, Stability, repeatability and reproducibility. The first two are linked up with accuracy and the last two with precision. The third one refers to ability of the measurement system to yield the same measure on the same part/sample/unit tested for the same parameter or measurand at different points in time or after several uses. Unless the process is repeated over different and somewhat separated time periods, we do not check stability and take it for granted within the desired calibration interval for the equipment involved.

Bias is defined as the (absolute) difference between the average of repeat measurements on the same part and the (true) reference value. This really is a measure of the controllable, systematic error in the measurement process.

Linearity corresponds to change in bias over the admissible range of the measurement process in terms of the range of values for the give measurand (possessed by different parts or samples.) In fact, if different parts (with corresponding reference values for the same measurand) and the bias is calculated for each part, we can examine the behaviour of bias against the reference value as remaining constant or increasing (decreasing) linearly (nonlinearly). Even a test for linearity can be carried out.

The remaining two parameters correspond to two distinct components of the total variation observed in an experiment where the process is repeated over several parts or samples, with different reference values for the parameter being measured and involving different operators (possibly using different copies of the measuring instrument).

An MSA study also referred to as a Gauge R&R (reproducibility and repeatability) Study is done by either the tabular method or the Analysis of Variance (ANOVA) method. In the tabular method, variances are estimated by using range, as is done on a control chart for variability where we take R/d_2 as the estimate of standard deviation σ where the value of d_2 depends on the sample size. While the tabular method is simpler, the estimates of variance components based on range do not make use of all the observations directly. This is why the ANOVA procedure is usually preferred.

With the usual Analysis of Variance procedure with two factors, viz. parts and operators, these different components of the total observed variation are estimated and we get a valid idea of the inherent capability of the measurement. A variance component model is appropriate in case of random factors viz. parts and operators, as if the parts considered in the experiment constitute a random sample from the population of all possible parts and similarly the operators involved in the experiment define a random sample from a population of operators. A fixed effects mode or even a mixed effects model focusing only on the selected parts and selected operators or taking the levels of only one of these two factors as a random sample has also been tried out.

Measurement System Capability is expressed in terms of the two metrics viz. Signal-to-Noise Ratio (S/N ratio) and Precision-to-Tolerance Ratio (P/T ratio). These ratios are estimated from the results of the MSA experiment and the Measurement System is accepted as capable provided these ratios satisfy some specified range of values.

A linear model to estimate the different components of variance in an experiment involving l parts P_i and m operators O_j each required to measure each part n times can be presented as

$Y_{\text{ijk}} = \mu + \alpha_i + \beta_j + \lambda_{ij} + e_{ijk}$ where y_{ijk} stand for the kth measurement on part I taken by operator j, α_i being the specific effect of part i, β_j the specific effect of operator j, λ_{ij} the interaction (joint) effect of operator j measuring part I and e_{ijk} is the unexplained error component (that corresponds to repeatability). In the usual random effects' model, we denote components of variance due to parts, operators, operator x part interaction and error by symbols σ_{p}^2, σ_{o}^2, σ_{po}^2 and σ_{e}^2, respectively. The total variation in the entire measurement process is often denoted by σ^2. We now have the following relations.

$\sigma^2 = \sigma_{\mathrm{p}}^2 + \sigma_{\mathrm{g}}^2$ where σ_{g}^2 is the component due to gauge variability (or the measurement process). Further, $\sigma_{\mathrm{g}}^2 = (\sigma_{\mathrm{o}}^2 + \sigma_{\mathrm{po}}^2) + \sigma_{\mathrm{e}}^2$. The first part inside parentheses gives reproducibility while repeatability is indicated by the last term.

The two metrics used to assess the capability of a measurement system are defined as

Precision/Tolerance ratio which is taken as 6 σ_{g}/T where T = UTL − LSL is the tolerance range, UTL and LTL being respectively the upper and the lower tolerance limits for the parameter being measured and controlled during production. While this definition has been recommended by Montgomery (1997), the measure of precision recommended by Automotive Industry Action Group (AIAG) is 5.15 σ_{g}/T. While Montgomery suggests that

P/T should preferably be at most 0.1, the AIAG recommends that
if $P/T < 0.1$ the gauge is capable
if $P/T > 0.3$ the gauge is not capable while
if $0.1 \leq P/T \leq 0.3$ the gauge may be capable.

Similarly, Signal/Noise or S/N ratio is defined as $\sigma_{\mathrm{p}}/\sigma_{\mathrm{g}}$. Some recommend that S/N ratio should exceed 5 with at least 90% confidence. Confidence intervals for estimated S/N ratio have been derived.

Coming to the desired number of operators and of parts in a gage R&R study, it has been reported that the lengths of confidence intervals for the variance components diminishes significantly with the number of operators, while the number of parts does not affect these lengths that much (Burdick et al. 2003). In fact, the number of operators should be at least 5 or 6. However, some experimenters prefer to increase the number of parts, rather than the number of operators. Incidentally, assumptions of randomization and of replication—basic principles in the design of an experiment—should also be duly taken into account.

MSA studies have been extended to attribute gauges as also to multiple measurands being simultaneously considered. These studies are, as expected, quite complicated.

6.6.1 An Example

Consider an experiment in which five operators are required to measure the diameters (in mms) of three holes punched on a metallic surface, each operator taking four measurements on each hole. The measurements obtained are reproduced below in Table 6.1.

Table 6.1 Measurements on Hole diameters

Operator	Hole 1	Hole 2	Hole 3
A	56 45 43 46	60 50 45 48	66 57 50 50
B	61 58 55 56	60 59 54 54	59 55 51 52
C	63 53 49 48	65 56 50 50	66 58 52 55
D	65 61 60 63	60 58 56 60	53 53 48 55
E	60 61 50 53	62 68 67 60	73 77 77 65

Source Author (during a visit to an industrial unit)

We present the analysis of variance of these measurements in Table 6.2.

As the F-Ratio for the Part $\times$ Operator Interaction effect is larger than the significance point, this interaction component is worth noting in the context of our analysis.

It would be appropriate to use the random effects model here, since to examine the measurement system we can obviously go beyond the chosen five operators and the three selected holes. In fact, we can think of a population of holes and similarly a population of operators and consider the set of three holes and the group of five operators as random samples from the respective populations. We then get the estimated variance components as

$$\text{Est } \sigma^2_{\text{operators}} = (272.3 - 109.4)/12 = 13.57$$

$$\text{Est } \sigma^2_{\text{Holes}} = (54.6 - 109.4)/20 \text{ to be taken as 0 and}$$

$$\text{Est } \sigma^2_{\text{Interaction}} = (109.4 - 26.0)/4 = 20.85$$

Thus, 13.57 + 20.85 = 34.42 corresponds to reproducibility, while 26.0 corresponds to repeatability. Their total viz. 60.42 represents gage variability, which along with the variability due to parts make up for the total variability in the measurement process.

Repeatability will be estimated as $2.77 \times \sqrt{26} = 14.12$ approximately while estimated reproducibility in this example works out as $2.77 \times \sqrt{34.42} = 16.25$ approximately.

Table 6.2 ANOVA for data in Table 6.1

Source of variation	D. F.	S.S.	M.S.	F-Ratio
Due to holes (Parts)	2	109.2	54.6	2.1
Due to operators	4	1,089.2	272.3	10.5
Due to interaction	8	875.2	109.4	4.2
Error	45	1,170.5	26.0	
Total	59	3244.0		

Confining ourselves, somewhat unimaginatively, to the observed set of holes (parts) and the selected five operators, the mean squares presented in the table above would directly give us measures of reproducibility and repeatability.

Assuming that reference (true) values of the diameter for the three holes were 56, 57 and 58, respectively, the observed mean values (each based on 20 repeat measurements under reproducibility conditions) came out as 55.3, 57.1 and 58.6, respectively, implying biases of −0.7, 0.1 and 0.6. These figures show nonlinearity of the measurement process.

Estimating variability on the basis of range as is done on a control chart for sample range, the same exercise may be simply carried out as follows.

Calculate the average for each operator to yield values 51.33, 56.17, 55.42, 57.67 and 64.42 for operator 1, 2, 3, 4, and 5, respectively. The range of these five averages is $R_0 = 13.1$ and reproducibility S.D. can be estimated as σ (reproducibility) = $R_0/d_{2,}$ the range being based on five (average) values, $d_2 = 2.236$ yielding the Fig. 5.81. To get the repeatability S.D., we get the range of values for each part (hole) based on 20 values and these come out as 22, 23 and 29 with an average R-bar = 24.67 resulting in the estimate of σ (repeatability) = 24.67/3.735 = 6.605. This is seemingly larger than estimated repeatability S.D., some consequence of using a range-based estimate of variability without checking for homogeneity of the ranges. The estimate of σ (parts. Holes) can be obtained by getting the average for each part and getting the range of these three averages which come out as 55.30, 67.10 and 58.55 with a range of 3.25, yielding the s.d. estimate as 3.25/1.693 = 1.920. Components of variance obtained this way will not agree with those given by ANOVA.

6.7 Concept of Uncertainty

It is widely recognized that the true value of a measurand (or a duly specified quantity to be measured) is indeterminate, except when known in terms of theory. What we obtain from the concerned measurement process is at best an estimate of or an approximation to the true value. Even when appropriate corrections for known or suspected components of error have been applied, there still remains an uncertainty, that is, a doubt about how well the result of measurement represents the true value of the quantity being measured.

The true value is indeterminate and unknown, except when given by theory. It is presumed to be the value yielded by the best maintained and used instrument of the desired accuracy class.

Theory of errors is concerned with errors in measurements that can be noted in terms of difference among repeat measurements on the same measurand and that can be explained by a simple model like

$$\text{True value }(X) = \text{Observed Value }(x) + \text{Error }(e).$$

Current interest centers round uncertainty in the true value (as is estimated in terms of a single measurement or a set of repeat measurements). This is understood in terms of the spread of true values where from the observed value(s) could arise. The idea is motivated by the similarity in observed values when different true values of the measurand are considered.

The uncertainty in measurement is a parameter, associated with the result of a measurement, that characterizes the dispersion of the true values which could reasonably be attributed to the measurand. The parameter may be, for example, the standard deviation (or a given multiple of it), or the half width of an interval having a stated level of confidence.

Uncertainty and its evaluation or estimation from repeat measurements and calibration reports using some assumptions have been discussed by many authors and contained in many national standards like NABL 141 in India as also similar standards in other countries required to ensure compliance with the ISO 17025 standard. An important reference could be the document EA-4/02 M rev 01 (2013) published by the European Accreditation Agency. One may also refer to Kelkar which has been referred to by Maharashtra (India) Pradesh Pollution Control Board on their Website. The author dealt with this topic in his Platinum Jubilee Lecture in the Section of Statistics of the Indian Science Congress Association, published in 1996.

6.7.1 Measurement Model

Measurands are particular quantities subject to measurement. One usually deals with only one measurand or output quantity Y that depends upon a number of input quantities X_i (i = 1, 2, …, N) according to the functional relationship.

$$Y = f\,(X_1, X_2, \ldots, X_N) \tag{6.1}$$

The model function f represents the procedure of the measurement and the method of evaluation. It describes how values of the output quantity Y are obtained from values of the input quantities X_i.

In most cases, it will be an analytical expression, but it may also be analytical expressions which include corrections and correction factors for systematic effects, thereby leading to a more complicated relationship that is now written down as one function explicitly. Further, f may be determined experimentally, or exist only as a computer algorithm that must be evaluated numerically, or it may be a combination of all these.

An estimate of the measurand Y (output estimate) denoted by y is obtained from Eq. (6.1) using input estimates x_i for the values of the input quantities X_i.

$$y = f(x_1, x_2, \ldots, x_n) \tag{6.2}$$

It is understood that the input values are best estimates that have been corrected for all effects significant for the model. If not, necessary corrections have been introduced as separate input quantities.

6.7.2 Estimation of Uncertainty

The standard uncertainty in measurement associated with the output estimate y, denoted by $u(y)$, is the standard deviation of the unknown (true) values of the measurand Y corresponding to the output estimate y. It is to be determined from the model Eq. (6.1) using estimates $\mathrm{x_i}$ of the input quantities X_i and their associated standard uncertainties $u(x_i)$.

The set of input X_i may be grouped into two categories according to the way in which the value of the quantity and its associated uncertainty have been determined.

Quantities whose estimate and associated uncertainty are directly determined in the current measurement. These values may be obtained, for example, from a single observation, repeated observations, or judgement based on experience. They may involve the determination of corrections to instrument readings as well as corrections for influence quantities, such as ambient temperature, barometric pressure or humidity.

Quantities whose estimate and associated uncertainty are brought into the measurement from external sources, such as quantities associated with calibrated measurement standards, certified reference materials or reference data obtained from handbooks.

The standard uncertainty in the result of a measurement, when that result is obtained from the values of a number of other quantities, is termed combined standard uncertainty.

An expanded uncertainty is obtained by multiplying the combined standard uncertainty by a coverage factor. This, in essence, yields an interval that is likely to cover the true value of the measurand with a stated high level of confidence.

The standard uncertainty of Y is given by

$$\sigma_y = \left\{ \sum \sum C_i C_j \sigma_{ij} \right\}^{1/2} \tag{6.3}$$

where inputs X_i and X_j have a covariance σ_{ij} and C_i is the sensitivity of Y with respect to variation in $X_{i.}$ The formula simplifies in case the inputs are uncorrelated. The variances can then be easily estimated if repeat measurements are available on an input; otherwise, these are estimated by assuming some distribution of true values (which could be made to correspond to the same observed value), e.g. normal or rectangular or (right) triangular.

The uncertainty analysis of a measurement—sometimes called an uncertainty budget—should include a list of all sources of uncertainty together with the associated standard uncertainties of measurement and the methods for evaluating them. For repeated measurements, the number n of observations also has to be stated. For the sake of clarity, it is recommended to present the data relevant to this analysis in the form of a table. In this table, all quantities to be referenced by a physical symbol X or a short identifier. For each of them at least the estimate of x, the associated standard uncertainty of measurement $U(x)$, the sensitivity coefficient c and the different uncertainty contributions to $u(y)$ should be specified. The dimension of each of the quantities should also be stated with the numerical values given in the table.

6.7.3 An Example

The tensile strength testing machine in a conveyor belt manufacturing unit is calibrated annually. Tensile strength of finished belts is determined using the equipment involving the tensile value disk and the load cell. Ten repeat measurements on tension in kg/cm^2 were available on a particular belt specimen to estimate uncertainty about the true value. The following information about the equipment was also available for the purpose.

Tensile value disk	
Range used for calibration	0–50 kgf
Accuracy	As per manufacturer's data
Resolution	1 div. = 0.1 kgf
Load cell	
Uncertainty (%) from its calibration certificate	0.37 (A1)

Readings on tension are reproduced below

Reading No.	Tension
1.	153.50
2.	159.78
3.	167.04
4.	161.83
5.	156.10
6.	160.39
7.	187.05
8.	156.12
9.	161.39
10.	160.83

Type A Evaluation of Uncertainty

Mean Reading (kg/cm^2) = 160.40 Standard Deviation = 4.20 kg/cm^2
Standard Uncertainty U_r = standard deviation/$\sqrt{10}$ = 1.33 kg/cm^2
Standard Uncertainty (% U_r) = $U_r \times 100$/Mean reading = 0.83%

Type B Evaluation

Uncertainty of load cell received from the corresponding calibration certificate. We assume the underlying distribution to be normal so that the coverage factor at 95% confidence level is approximately 2 Thus, U_l (%) = $A_1/2 = 0.37/2 = 0.185\%$

(A_1 considered as the expanded uncertainty $U_e = 2 \times$ standard uncertainty). Thus, Uncertainty of load cell $U_l = 0.185 \times 160.40 \times 0.01 = 0.297$ kg/cm^2

Since $U_l = U_l$ % Mean Reading/100) Thus, the estimated uncertainty of load cell works out as $0.37 \times 160.40 \times 0.01 = 0.593$ kg/cm^2

Combined standard uncertainty $U_c = \sqrt{[U_r \times Ur + U_l \times U_l]} = 1.46$ kg/cm^2 and % $U_c = 0.91\% = U_c \times 100$/Mean Reading

Expanded combined uncertainty for approximately 95% level of confidence

$U = 2 \times 1.46 = 2.92$ kg/cm^2

And $U\% = 1.8\%$

The uncertainty budget can now be worked out conveniently.

6.8 Improving Quality of Measurements

To improve quality of measurements or to reduce uncertainty (about the true values), we have to reduce chance or random errors (which cannot be completely eliminated) and to remove systematic errors or biases. Broadly speaking, random errors can be reduced by using measuring instruments and maintaining them properly, exercising necessary control on environmental influences on the measurands and training people involved in taking measurements to avoid personal equation biases, etc. Systematic errors are associated with measuring instruments as also with reference measures. This requires instruments to be calibrated regularly against reference standards.

It is generally agreed that calibration of test, inspection and measuring equipments takes care of accuracy, while careful use and maintenance of such equipments lead to improved precision. Both calibration and maintenance are essential to reduce uncertainty about true values. In the following, we provide brief explanations of Calibration and of Good Laboratory Practice that is needed to ensure good quality of measurements.

6.8.1 Calibration—Process and Procedures

The International Standard ISO 10012-1 relating to Quality Assurance Requirements for measuring equipment mentions metrological confirmation system for measuring equipment as ‘the set of operations required to ensure that an item of measuring equipment is in the state of compliance with requirements for its intended use’. This includes calibration, adjustment or repair and subsequent recalibration as well as sealing and labelling. However, many practitioners feel that calibration itself covers these different requirements and hence can provide metrological confirmation for the test, measuring and inspection equipment.

Calibration is the set of operations which establish, under specified conditions, the relationships between values indicated by a measuring instrument of measuring system, or values represented by a material measure or reference material, and the corresponding values of a quantity realized by a reference standard. Such a relationship may be used to adjust or correct an instrument or a system, even values of measures or even reference materials, wherever such adjustments or corrections are feasible and desirable. In other cases, these relations provide bases for corrections in or conversions of measurements. The result of a calibration permits either the assignment of values of the measurand to the indications given by a measuring equipment as they are or the determination of corrections with respect to indications. A calibration may also determine other metrological properties such as the effect of influence quantities. The result of a calibration may be recorded in a document, sometimes called a calibration certificate or a calibration report.

Calibration involves checking the operational integrity of a test or measuring equipment or of a measurement standard of unverified accuracy by comparing its performance with that of a standard of known greater accuracy in order to detect, correlate, report or eliminate (by adjustment) any deviation in accuracy, capability or from any other required performance. Calibration gained importance mainly due to stringent requirements in defence supplies and the MIL standards took a lead in formalizing the calibration philosophy and subsequently boosting the calibration practice. As indicated earlier, measurement implies a process as well as the output of that process. The process of measurement needs control and calibration is an important control exercise.

Calibration can be carried out for three possible purposes viz.

(i) Determining whether or not a particular instrument or standard is within some established tolerance in respect of its deviation from a reference standard.
(ii) Reporting of deviations in measurements from nominal values.
(iii) Repairing/adjusting the instrument or standard to bring it back within the established tolerance.

It is important to note that these three purposes are not mutually exclusive. In fact, all the three may be relevant in a particular situation.

Usually, a hierarchical calibration system is adopted to ensure traceability of measurements given by a measuring instrument to some nationally accepted measurement system through an unbroken chain of comparisons. In the commonest case, we are interested in calibrating a given measuring instrument (which does nor need any material measure to produce a measurement or a value for the measurand of interest) against a certain reference standard.

Calibration procedures vary from one type of measuring equipment to another. For example, in calibrating a micrometer we take sequential measurements of gauge blocks of known size specified by some standard. In the IS, the specified sizes are 2.5, 5.1, 7.7, 10.3, 12.9, 15.0, 17.6, 20.2, 22.8 and 25 mm. Dimensions indicated by the micrometer are noted and deviations from the nominal values recorded. Of course, the accuracy of the gauge blocks themselves has to be ensured or determined as a prerequisite. Alternatively, the length of the gauge blocks can be compared with those of matters of identical nominal lengths.

In the commonest case, we are interested in calibrating a given measuring instrument (which does not need any material measure or reference material to produce a measurement or a value of the measurand) with reference to a certain reference standard. Here also, the same two objectives of calibration—leading to adjustment/correction of the instrument or of the measurements—remain valid, depending on individual situations. We produce n measurements for a measurand (may be n items assessed for the same characteristic) by using both the given and the reference instruments. Let the two series by y, y,…,y and M, M,…, M, respectively. Calibration means establishing a relation—often assumed linear—between the two series of the form $y = \alpha + \beta M$, with numerical values of the parameters α and β determined by the methods of least squares from the two series of measurements.

6.8.2 Calibration System

In a calibration system, the following items shall be defined.

1. Classification of calibration;
2. Standard and levels of standard;
3. Interval of calibration and limit of correction;
4. Procedures of calibration;
5. Action after calibration;
6. Conditions to use measuring instrument;
7. Procedures of measurement.

Table 6.3 gives out relational formulae used in different types of calibration which are used in different contexts.

Table 6.3 Classification of calibration and relational formulae for them

	Type of calibration	Relational formula
a	Calibration with only inspection: Do not correct and take the reading as it is as the measured value	$y = M$
b	Zero point calibration: Conduct the calibration of fixes point by reading of zero point y_0	$y = y_0 + M$
c	Reference point calibration: Conduct the calibration of fixed point by reading y_0 of reference point M_0	$y = y_0 + (M - M_0)$
d	Scale interval calibration: Conduct the calibration of inclination taking the optional point (its reading is y_0) as zero point	$y = y_0 + \beta M$
e	Zero point proportional formula calibration: Suppose the reading of zero point as zero and conduct the calibration of inclination	$y = \beta M$
f	Reference point proportional formula calibration: Conduct calibration of fixed point by the reading y_0 of reference point M_0 and then conduct calibration of inclination	$y = y_0 + \beta(M - M_0)$
g	Linear formula calibration: Conduct simultaneously calibration of fixed point and calibration of inclination with using mean value $\bar{y}$ of reading y and mean value $\bar{M}$ of value of standard M	$y = \bar{y} + \beta(M - \bar{M})$

Source ISO/IEC Standard 17025 on Good laboratory Practice

This table does not deal with calibration by formulae of high degrees of freedom of nonlinear type. In these cases, it is possible to conduct calibration assuming a linear relation within each of several ranges.

In order to take the reading as it is as the measured values, changes of scale or mechanical adjustment may be made. In such cases, correction by change of scale of mechanical adjustment and they are discriminate from no calibration.

6.9 Quality Requirements for Measures

Measures associated with different phenomena or processes and their outcomes should possess some properties in order that we can assess their relevance, appropriateness and dependability for use in making inferences and actions. And these features or properties of a measure really characterise what may be termed as 'quality of the measure'. In this context, we consider these desirable properties for a measure (indicator) of performance. And to be focused on processes which have a bearing on quality, we consider measures of process performance and leave out measures of organizational performance from the scope of the present discussion. In this context, it is worthwhile to mention that 'Process Performance Measurement System' has been discussed by several authors and is regarded as an essential activity to provide inputs for quality improvement.

Kitchenham (1995) and Winchell (1996) have identified the following properties or features as the main requirements for process performance indicators. These were subsequently discussed by Holweg (2000). The list is generic and should not be claimed as exhaustive.

Quantifiability Since a major objective of using measures of performance would be to compare performance across time or units or sections and the like, the measure has to be quantified. If performance indicators are not quantitative by nature, they have to be transformed. For instance, the performance indicator 'customer payment attitude' could be transformed into number of days between 'invoice sent' and 'invoice paid'. This way, qualitative measures may be quantified—though not always uniquely—by using related quantified measures.

Sensitivity Sensitivity expresses how much the performance measure must change before a change in performance can be detected. In fact, a sensitive indicator is able to detect even minor changes in performance. It is well appreciated that improvements in process performance will more often than not be marginal, though continuous. And such marginal improvements will not be detected by a measure that is not sensitive enough. Big changes are obvious and the involvement of a measure is not that critical.

Linearity Linearity indicates the extent to which process performance changes are congruent with the value of a certain indicator. Or, conversely, a small change in the business process performance should lead to a small change in the value of a corresponding performance indicator, whereas an ample performance rise should also lead to strong change in the level of the performance indicator. And this behaviour should be maintained throughout the plausible range of values for the indicator. Otherwise, it will be difficult to interpret the same difference in value of the indicator over different parts of this range.

Reliability A reliable performance indicator is free of measurement errors. To illustrate, if a certain business process has to be rated through a given performance indicator by different experts, the results should not depend on the subjective evaluation of an individual. Inter-rater consistency is an important requisite.

Efficiency Since the measurement itself requires human, financial and physical resources, it must be worth the effort from a cost/benefit point of view. The measure has to reflect changes in process performance faithfully with a minimum of effort.

Improvement Orientation Performance indicators should emphasize improvement rather than conformity with instructions. Therefore, measuring billing errors, number of safety violations, data entry errors and the like do not create an atmosphere where feedback sessions are viewed in a positive, constructive light. Indicators should be so defined and scaled that its values speak of aspects of performance which are directly and not inversely linked up with improvement in process performance.

It should be noted that

1. Performance is not absolute.
2. Performance is multi-dimensional.
3. Performance measures are not independent of one another.

Kueng (2002) points out that even if a process performance measure satisfies the above desiderata, it may not be acceptable by the team that is to make use of the same. That way acceptability by the users of a measure is also quite important.

6.10 Concluding Remarks

Quality of measurements has a crucial role to play in Quality Management. Not too unoften, we are told about disputes between the producer/supplier and the customer not agreeing on the value or level of an important quality parameter of the product under transaction. And in a few of such cases, the very fact that there would always remain some small difference between these values or levels obtained by two parties and a genuine problem should correspond to a d difference exceeding, for example, a multiple of the reproducibility factor R or the length of the expanded uncertainty interval. In this context, Measurement System Capability Analysis becomes a must in situations where a high degree of precision is required or very small measurements are involved.

While the concept of 'uncertainty' about the true value as also methods for estimating uncertainty have been documented by national and international regulatory bodies, its use has not yet been that widespread. More than that, the existing method of estimating uncertainty is not above criticism. It makes use of a measurement model which is not completely objective in the identification and incorporation of all possible inputs, refers to a formula for obtaining the standard deviation of the estimated true value which is applicable to large samples and assumes some probability distribution to convert the range of variation in an input parameter into a corresponding standard deviation. One may genuinely object to the use of an uncertain or asymptotic procedure to estimate uncertainty in a set of measurements or a single measurement. However, the attempt to identify different sources of error in measurements and to quantify their contributions to the overall uncertainty in a measurement should be definitely appreciated.

References

American Association for Laboratory Accreditation. (2014). *Guide G104 for Estimation of Measurement Uncertainty in Testing*.

Automobile Institute Action Group. (2002). *Measurement system analysis manual*. Detroit, MI, USA: AIAG.

Bell, S. (1999). *Measurement good practice guide #11 a beginner's guide to uncertainty of measurement*. Technical report. National Physical laboratory.

Burdick, R. K., Borror, C. M., & Montgomery, D. C. (2003). A review of methods for measurement system capability analysis. *Journal of Quality Technology, 35,* 342–354.

Deutler, T. (1991). Grubbs-type estimators of reproducibility variances in an inter-laboratory test study. *Journal of Quality Technology, 23*(4), 324–335.

Dietrich, C. F. (1991). *Uncertainty, calibration and probability*. Bristol, U.K.: Adam Hilger.

EA-4/02M rev.2. (2013). *Evaluation of the uncertainty of measurements in calibration*. European Accreditation Organisation.

Fridman, A. E. (2012). *The quality of measurements: A metrological reference*.USA: Springer.

Grabe, M. (2005). *Measurement uncertainties in science and technology*. USA: Springer.

Holweg, M. (2000). Measuring cost and performance. Cardiff Business School.

IS 5420 Part 1. (1973). *Guide on precision of test methods—Principle and applications*.

ISO Standard 10012-1 *Quality assurance requirements for measuring equipment.*

ISO/ IEC Standard 17025. Good laboratory Practice.

Kelkar, P. S. (2004). *Quality control for sampling and chemical analysis* (Ref mpcb.gov.in/envtdata/QA-QC-PSK.pdf).

Kitchenham, B., et al. (1995). *Case studies for method and tool evaluation*. *IEEE Software*, 52–67.

Kueng, P. (2002). Process performance measurement system. *Total Quality Management, 35*(5), 67–85.

Larsen, G. A. (2002–03). Measurement system analysis: The usual metrics can be non-informative. *Quality Engineering, 15*(2), 293–298.

Lilliken, G. A., & Johnson, D. E. (1984). *Analysis of messy data*, Vol. 1 Designed experiments. New York: Van Nostrand Reinhard.

Mukherjee, S. P. (1996). Platinum Jubilee Lectures, Part I. Indian science Congress Association.

Mukherjee, S. P. (2000). Quality of measurements in role of measurements in science and technology. In S. P. Mukherjee & B. Das (Eds.), Indian Association for Productivity, Quality & Reliability and National Academy of Sciences of India.

Pendrill, L. (2008). *Applications of statistics in measurement and testing*.

Tsai, P. (1988). Variable gage repeatability and reproducibility using the analysis of variance method. *Quality Engineering, 1*(1), 107–115.

Vardeman, S. B., & Van Valkenburg, E. S. (1999). Two way random effects analysis and gauge R&R studies. *Technometrics, 41*(3), 202–211.

White, G. H. (2008). Basics of estimating measurement uncertainty. *Clinical BioChemist Review, 29*(Supplement!), 553–560.

Winchell, W. (1996). *Inspection and measurement: Keys to process panning and improvement*. Society of Manufacturing Engineers.