GUIDE 33

Uses of certified reference materials

Second edition 2000

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.ch
Web www.iso.ch

Printed in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

Guides are drafted in accordance with the rules given in the ISO/IEC Directives, Part 3.

Draft Guides adopted by the responsible Committee or Group are circulated to national bodies for voting. Publication as a Guide requires approval by at least 75 % of the national bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this Guide may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO Guide 33 was drawn up by the ISO Committee on reference materials (REMCO) and was approved by ISO member bodies.

This second edition cancels and replaces the first edition (ISO Guide 33:1989), which has been technically revised.

Introduction

Today's world of modern technology requires a large number of certified reference materials (CRMs) in widely diverse fields, and the demand for such materials is expected to increase. The preparation of a CRM is a time-consuming, meticulous and expensive endeavour and consequently it has not always been, and will continue not to be, possible to satisfy the demand for all types and quantities of CRMs. For this reason, CRMs should be used properly, i.e. effectively, efficiently and economically.

Certified reference materials should be used on a regular basis to ensure reliable measurements. However, in doing so, the magnitude of the supply of that CRM, its relative cost, its availability (accessibility) and the measurement technique, be it destructive or non-destructive, should be considered. Also important to the user is the fact that the misuse of a CRM may not provide the intended information.

Misuse of CRMs differs from incorrect use. The user of a CRM is expected to be familiar with all information pertinent to the use of the CRM as specified in its certificate. He should comply with such factors as the period of validity of the CRM, the prescribed conditions for storage of the CRM, instructions for the use of the CRM, and specifications for validity of the certified properties of the CRM. A CRM should not be used for a purpose other than that for which it was intended. Nevertheless, from time to time, when a user must resort to applying a CRM in an incorrect manner because of the unavailability of a suitable CRM, he should be fully cognizant of the potential pitfalls and therefore assess his measurement output accordingly.

There are many measurement processes in which CRMs are in general use but are replaceable by a host of working standards, such as homogeneous materials, previously analysed materials, pure compounds, solutions of pure elements, etc. Some examples are where only a "rough" estimate of the trueness or precision of a method is sought, where "blind" unknown check samples are used routinely in quality control programmes, and where only the variation in trueness or precision of a method with some parameter such as time, analyst, instrument, etc., is being evaluated. The first example illustrates the use of a CRM in which the well-defined certified value and uncertainty of the CRM is under-utilized. The others illustrate the case in which a series of "one-time" trueness and precision assessments are compared with one another. There is no need to base that comparison on a well-defined certified value and uncertainty of a CRM. The advantages in using CRMs are that the user has the means to assess the trueness and precision of his measurement method and establishes metrological traceability for his results.

Whether the use of CRMs in these procedures is in fact "misuse" depends largely on the availability and relative cost of the CRMs. Where CRMs are in short supply or very expensive, their use would indeed be misuse. However, for CRMs in ample supply or where similar CRMs are available from one or more sources, it is strongly recommended that CRMs always be used instead of in-house standards because of the resultant enhanced confidence in the measurement output.

It is important that users remain aware that the preparation of in-house standards for use instead of CRMs has an associated cost based on factors such as material cost, facility usage charges, personnel labour rates, etc., in which the material cost is in general the lowest. For some CRMs such as the complex compositional materials certified for chemical composition, the cost of preparing in-house standards to match the composition of real samples can exceed that of available CRMs. In these cases, the use of CRMs is recommended.

The user should be aware of the potential misuse of CRMs as "blind" unknown check samples in quality control programmes. Where there are only a few CRMs in an area of expertise, they are easily recognized and they may therefore not satisfy the intended purpose. Moreover, the same CRM should never be used for both calibration purposes and as "blind" unknown check sample in a measurement process.

The misuse of CRMs can also occur when the user does not fully take into account the uncertainty in the certified property. The combined standard uncertainty of a certified property of a CRM can have contributions from the inhomogeneity of the material, the within-laboratory uncertainty and, where applicable, the between-laboratory uncertainty. The level of homogeneity defined for a CRM by the producer is dependent on both the statistical design used to evaluate it and the repeatability of the method of measurement. For certain CRMs, the level of

homogeneity is valid for a test portion defined by mass, physical dimension, time of measurement, etc. The user should be aware that the use of a test portion that does not meet or exceed that specification could severely increase the contribution of the inhomogeneity of the CRM to the uncertainty of the certified property, to the point where the statistical parameters of certification are no longer valid.

The variation in the repeatability of different methods has another implication for the user. Since the degree of inhomogeneity of a CRM is dependent on the repeatability of the method of measurement, it is possible that a user, in applying a method capable of better repeatability, could detect inhomogeneity in that CRM. In such cases, the observed inhomogeneity is already accounted for in the statistical parameters for the certified property and therefore the statistical tests presented in this Guide remain valid, but the scientific basis for using that particular CRM to give a true assessment of the user's method should again be questioned.

It is well known that different methods of measurement of a property are not capable of equal repeatability. Accordingly there could arise instances where the user may wish to assess a method that has greater repeatability than that or those used in the certification of the CRM. In such cases, the statistical tests presented in this Guide remain valid but the scientific basis for using that particular CRM to give a true assessment of the precision (and possibly the trueness) normally expected from the user's method should be questioned. It is recommended that the user resorts to a CRM of lesser uncertainty, if available.

For CRMs certified by a primary method, the user should not assume that his method is capable of matching the precision and trueness reported for the CRM. It is unreasonable therefore to apply the statistical procedures in this Guide for assessing the trueness and precision of a method by application to a CRM using the certification parameters for a property reported in the certificate. The user, as a consequence, should either experimentally establish or make estimates based on available information for those parameters that are more appropriate. Similarly, where a user applies a method to a CRM that has been certified by a single different method, the user should not assume that the certification parameters for the certified property are applicable to his method except in cases where the trueness and precision capable by both methods are known to be comparable.

One of the important considerations in selecting a CRM for use, either in assessing the trueness and precision of a method or in the calibration of instruments in a method, is the level of uncertainty required by the end-use of the method. Obviously the user should not apply a CRM of greater uncertainty than permitted by the end-use.

The selection of CRMs should take into account not only the level of uncertainty required for the intended purpose but also their availability, cost, and chemical and physical suitability for the intended purpose. For example, the unavailability or high cost of one CRM could force a user to resort to using another CRM of greater uncertainty than the preferred one. Also, in chemical analysis, a CRM of greater, but still acceptable, uncertainty in the certified property may be preferred over another CRM because of better matching with the composition of real samples. This could result in minimizing "matrix" or chemical effects in the measurement process which are capable of causing errors far greater than the difference between the uncertainties of the CRMs.

In conclusion, CRMs are meant to fulfil many purposes. Accordingly, a CRM used properly for one purpose in one laboratory may be misused for another purpose in another laboratory. It is recommended that the user consider the suitability of a CRM for his intended purpose on a case-by-case basis.

Uses of certified reference materials

1 Scope

This Guide discusses the uses of certified reference materials (CRMs) and their correct applications.

Clause 2 of this Guide presents definitions (with indication of their sources) of terms used, and clause 4 sets out the statistical considerations on which the Guide is based.

Clause 5 discusses the role of CRMs in measurement science and in realization of conventional measurement scales.

Clause 6 presents recommendations for developing criteria for the assessment of the precision and trueness of a measurement procedure by the use of CRMs. It pertains only to CRMs characterized to be homogeneous as described in ISO Guide 35 [4].

NOTE The use of CRMs is essential for assessment of trueness and optional for assessment of precision.

This Guide does not describe the use of certified reference materials as calibrants. This subject is treated in ISO Guide 32 [3].

2 Terms and definitions

For the purposes of this Guide, the following terms and definitions apply.

2.1
measurement process
all the information, equipment and operations relevant to a given measurement

NOTE This concept embraces all aspects relating to the performance and quality of the measurement; it includes, for example, the principle, method, procedure, values of the influence quantities and the measurement standards.

[VIM:1993]

2.2
influence quantity
quantity that is not the measurand but that affects the result of the measurement

EXAMPLE Ambient temperature; frequency of an alternating measured voltage.

[VIM:1993]

2.3
reference material
RM
a material or substance one or more of whose property values are sufficiently homogeneous and well established to be used for the calibration of an apparatus, the assessment of a measurement method, or for assigning values to materials

[ISO Guide 30:1992]

2.4
certified reference material
CRM
a reference material, accompanied by a certificate, one or more of whose property values are certified by a procedure which establishes traceability to an accurate realization of the unit in which the property values are expressed, and for which each certified value is accompanied by an uncertainty at a stated level of confidence

[ISO Guide 30:1992]

2.5
precision
the closeness of agreement between independent test results obtained under prescribed conditions

[ISO 3534-1]

2.6
repeatability
precision under repeatability conditions

[ISO 3534-1]

2.7
repeatability conditions
conditions where independent test results are obtained with the same method on identical test material in the same laboratory by the same operator using the same equipment within short intervals of time

[ISO 3534-1]

2.8
repeatability standard deviation
the standard deviation of test results obtained under repeatability conditions

NOTE It is a measure of the dispersion of the distribution of test results under repeatability conditions.

[ISO 3534-1]

2.9
repeatability limit
r
the value less than or equal to which the absolute difference between two single test results obtained under repeatability conditions is expected to be with a probability of 95 %

[ISO 3534-1]

2.10
reproducibility
precision under reproducibility conditions

[ISO 3534-1]

2.11
reproducibility conditions
conditions where test results are obtained with the same method on identical material in different laboratories by different operators using different equipment

[ISO 3534-1]

2.12
reproducibility limit
R
the value less than or equal to which the absolute difference between two single test results obtained under reproducibility conditions is expected to be with a probability of 95 %

[ISO 3534- 1]

2.13
bias
the difference between the expectation of the test results and an accepted reference value

NOTE Bias is a systematic error as contrasted to random error. There may be one or more systematic error components contributing to the bias. A larger systematic difference from the accepted reference value is reflected by a larger bias value.

[ISO 3534-1]

2.14
accuracy
the closeness of agreement between a test result and the accepted reference value

NOTE The term accuracy, when applied to a set of test results, involves a combination of random components and a common systematic error or bias component.

[ISO 3534-1]

2.15
trueness
the closeness of agreement between the average value obtained from a large series of test results and an accepted reference value

NOTE The measure of trueness is usually expressed in terms of bias.

[ISO 3534-1]

2.16
uncertainty
<of measurement> parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand

[VIM:1993, GUM:1993]

NOTE This definition is that of the *Guide to the Expression of Uncertainty in Measurement* (GUM) in which its rationale is detailed (see, in particular, 2.2.4 and annex D [5].

2.17
estimation
the operation of assigning, from the test result in a sample, numerical values to the parameters of a distribution chosen as the statistical model of the population from which this sample is taken

[ISO 3534-1]

2.18
estimate
the result of an estimation

[ISO 3534-1]

2.19
estimator
a statistic used to estimate a population parameter

[ISO 3534-1]

2.20
null hypothesis
the hypothesis to be rejected or not rejected (accepted) at the outcome of the test

[ISO 3534-1]

3 Symbols and subscripts

3.1 Symbols

a_1; a_2	adjustment values chosen in advance
$E(x)$	expectation of a random variable
G	Grubbs' test statistic
n	number of replicate results
p	number of laboratories participating in interlaboratory measurement programme
r	repeatability limit
R	reproducibility limit
s	estimate of a standard deviation
$V(x)$	variance of a random variable
x	measurement result
$\bar{x}$	arithmetic mean of measurement results
$\bar{\bar{x}}$	overall (grand) mean of measurement results
α	significance level
β	type II error probability
δ	estimated bias of the measurement process
μ	accepted reference value of a property
ν	number of degrees of freedom
σ	true value of a standard deviation
σ_D	uncertainty of the measurement process expressed by its standard deviation

$\chi^2_{(n-1);\,0,95}$ 0,95th quantile of the χ^2 – distribution with (n – 1) degrees of freedom

3.2 Subscripts

c computed value

i identifier for an individual result

L between-laboratory (at CRM certification)

Lm between-laboratory (at the assessed method)

w within-laboratory

wo within-laboratory, required

4 Statistical considerations

4.1 Basic assumptions

All statistical methods used in this Guide are based on the following assumptions.

a) The certified value is the best estimate of the true value of the property of the CRM.

b) All variation, be it associated with the material (i.e. homogeneity) or the measurement process, is random and follows a normal probability distribution. The values of probabilities stated in this Guide assume normality. They may be different if there is deviation from normality.

4.2 Decision errors

The assessment of a measurement process on the basis of precision and trueness is always subject to rendering an incorrect conclusion because of

a) the uncertainty of measurement results and

b) the limited number of replicate results usually performed.

Increasing the number of measurements tends to decrease the chance of an incorrect conclusion but, in many instances, the risk of making a wrong conclusion has to be balanced in economic terms against the cost of increasing the number of measurements. Accordingly, the rigour of the criteria developed for assessing a measurement process must take into account the level of precision and trueness requisite for the end-use.

For the purposes of this Guide the term "null hypothesis" is applied.

In this case the null hypothesis is that the measurement process has bias no greater than the limit chosen by the experimenter and variance no greater than the predetermined value; the alternative hypothesis is the hypothesis which is opposed to the null hypothesis (see also ISO 3534-1[7]).

There are two types of possible error in accepting or rejecting the null hypothesis:

a) **error type I:** The error committed in rejecting the null hypothesis when in reality the null hypothesis is true.

— **type I risk:** The probability of committing error type 1. Its value varies according to the real situation.

— **significance level:** The given value, usually designated by α, which limits the probability of committing error type 1.

b) **error type II:** The error committed in failing to reject the null hypothesis when in reality the alternative hypothesis is true.

— **type II risk:** The probability, usually designated by β, of committing error type II. Its value depends on the real situation and can be calculated only if the alternative hypothesis is adequately specified.

— **power of test:** The probability of not committing error type II, usually designated by $(1 - \beta)$. It is the probability of rejecting the null hypothesis when in reality the alternative hypothesis is true.

The choice of the values of both α and β is usually based on economic considerations dictated by the importance of the consequences of the decision. These values, as well as the alternative hypothesis, should be chosen before the start of the measurement process.

5 The role of certified reference materials in measurement science

5.1 General

Metrology is the field of knowledge concerned with measurement. Metrology or measurement science includes all aspects both theoretical and practical with reference to measurements, whatever their level of accuracy, and in whatever fields of science or technology they occur. This clause describes the role of reference materials in quantitative measurements.

5.2 The role of certified reference materials in the storage and transfer of information on property values

By definition (2.3), a reference material has one or more properties, the values of which are well established by measurement. Once the property value(s) of a particular CRM have been established, they are "stored" by the CRM (up to its expiration date) and are transferred when the CRM itself is conveyed from one place to another. To the extent that the property value of a CRM can be determined with a well-defined uncertainty, that property value can be used as a reference value for intercomparison or transfer purposes. Hence CRMs aid in measurement transfer, in time and space, similar to measuring instruments and material measures.

A CRM must be suitable for the exacting role it performs in storing and transferring information on measured property values. The following technical criteria (legal or commercial criteria may be relevant also) apply to the fitness-for-purpose of CRMs in general:

a) the CRM itself and the property value(s) embodied in it should be stable for an acceptable time-span, under realistic conditions of storage, transport and use;

b) the CRM should be sufficiently homogeneous that the property value(s) measured on one portion of the batch should apply to any other portion of the batch within acceptable limits of uncertainty; in cases of inhomogeneity of a large batch, it may be necessary to certify each unit from the batch separately;

c) the property value(s) of the CRM should have been established with an uncertainty sufficient to the end use(s) of CRM;

d) clear documentation concerning the CRM and its established property value(s) should be available. The property value(s) should have been certified, so the documentation should include a certificate, prepared in accordance with ISO Guide 31.

Whenever possible, the measurement of a given property value should have been made by an acceptable method having negligible uncertainty relative to end-use requirements and by means of measuring instruments or material measures which are traceable to national measurement standards. Subsequent use of a CRM with traceable property values ensures that traceability is propagated to the user. Since most national measurement standards

are themselves harmonized internationally, it follows that measurement standards in one country should be compatible with similar measurements in another country. In many cases, CRMs are appropriate for the intercomparisons of national measurement standards.

5.3 Use of certified reference materials for measurement traceability

A laboratory should control and verify a number of parameters in order to ensure the traceability of the measurements. To do this in all necessary detail is very hard work.

This can be considerably simplified by the use of a certified reference material of established traceability. The reference material shall be sufficiently similar (in matrix) to the actual sample to be analysed in order to include all analytical problems which might cause errors in the measurements. The user should apply to the reference material the same analytical procedure used for his unknown sample.

Hence the role of the certified reference material is comparable to that of the transfer standards used in metrology laboratories in industry, in that it allows working with a specified margin of uncertainty.

The certified reference materials also provide a possible way to determine the uncertainty of a measurement for analytical determinations or technological testing.

5.4 The role of reference materials in the International System of units (SI)

5.4.1 Dependence of the SI base units on substances and materials

The majority of measurements made in the world today are within the framework of the International System of units. In its present form, SI recognizes seven base units, namely the units of length (metre, symbol m), mass (kilogram, kg), time (second, s), electric current (ampere, A), thermodynamic temperature (kelvin, K), amount of substance (mole, mol) and luminous intensity (candela, cd). The definitions of these base units mention the following substances: platinum-iridium (for fabricating the prototype kilogram), caesium-133 (for determining the second), water (for defining the kelvin) and carbon-12 (for defining the mole). Opinions differ as to whether the substances named fall under the definition of reference material (2.3).

Certainly such materials have a special status as defined substances on which the SI is based. The dependency strictly applies to definition of the unit, since realization of the units may involve other substances/materials.

5.4.2 Obtaining of derived SI units with the aid of reference materials

From the seven base units of the SI, a large number of derived units of the SI are obtainable by combining base units as products and/or quotients. For example, a derived unit of mass concentration is defined as $kg \cdot m^{-3}$ and the derived unit of pressure (given the special name pascal, symbol Pa) is defined as $m^{-1} \cdot kg \cdot s^{-2}$. Formally speaking, the derived units ultimately depend on the substances on which the base units themselves depend (see 5.4.1). In practice, the derived units are often obtained not from base units but from CRMs with accepted property values. Thus a variety of substances/materials may be involved in the determination of derived units (examples 1 and 2 below) or even of base units (example 3 below).

EXAMPLE 1 The SI unit of dynamic viscosity, the pascal second ($Pa \cdot s = m^{-1} \cdot kg \cdot s^{-1}$) may be obtained by taking the value for a well-purified sample of water as 0,001 002 Pa·s at 20 °C.

EXAMPLE 2 The SI unit of molar heat capacity, the joule per mole kelvin ($J \cdot mol^{-1} \cdot K^{-1} = kg \cdot m^2 \cdot s^{-2} \cdot mol^{-1} \cdot K^{-1}$) may be obtained by taking the value for purified α-alumina as 79,01 $J \cdot mol^{-1} \cdot K^{-1}$ at 25 °C.

EXAMPLE 3 The SI unit of temperature, the kelvin, may be obtained at any temperature T_1 (273,15 K < T_1 < 903,89 K) from measurements of the resistance of a highly pure platinum wire at the triple point of purified water, at the freezing point of purified tin and at the freezing point of purified zinc, coupled with use of a specified mathematical relation.

5.4.3 Connection of analytical chemistry to the International System of units

It will be noted that purified (often called "pure") chemical substances were cited in each of the examples 1 to 3 (5.4.2). The measurement of degree of purity, or more generally of the chemical composition of materials, is within the realm of analytical chemistry. In addition to the dependence of SI on chemical substances, the dependence of analytical chemistry on SI is worthy of examination. Presently, most analytical chemists employ units within the SI (all base units except the candela and also many derived units) in their measurements. However, compositional analysis depends on an additional concept, namely that pure chemical species exist to which the chemical compositions of other substances and materials are referred, by invoking the laws of chemical change and stoichiometry.

From one or more pure chemical species, considered to be primary measurement standards, it is feasible to construct measurement hierarchies for analytical chemistry similar to those used in physical measurements. Examples of such measurement standards are:

a) the electron, to which other species can be connected by electrochemical analysis;

b) carbon-12, to which other species can in principle be connected by mass spectrometry, Raoult's law measurements, or volumetric measurements with low-density gases, etc.;

c) a highly purified element or compound, to which other species can be connected by electrochemical, gravimetric, titrimetric, spectrometric methods, etc.

The "other species" cited in these examples will in many cases be used as CRMs. Many substances can fill this role of intermediaries between primary and working analytical standards using the diversity of techniques and chemical reactions that an analyst may employ. The concept of traceability applies to analytical chemistry as much as it does to other branches of measurement science. The quality of the result of a chemical analysis will be enhanced if the result's traceability can be clearly stated in terms of the traceability of the instruments, material measures and CRMs employed. In most cases, the traceability will also depend on the numerical values of the relative atomic masses (formerly called "atomic weights") used in the calculations; the source of these should be recorded by the analyst.

5.5 Defining and creating conventional scales

5.5.1 General

Many measurement scales have been used since the earliest civilizations. Originally almost all of them were conventional, independent and inaccurate. Scientific and technical progress as well as international trade have led to both the need and the possibility of a unique, rational, self-consistent international system of units, the SI, which has been officially adopted worldwide. Nevertheless, it is not applicable to certain types of measurements for which it is necessary to create, sustain and use certain conventional units which are not within the scope of SI. In other cases the unit relating to the quantity to be measured lies within the frame of SI, but the reproduction of the unit according to the definition is technically difficult and expensive. It is therefore more convenient to carry out the measurement on a practical scale of reference values assigned to material properties. Though a reference value scale and a pure conventional scale differ theoretically from each other, they are similar with respect to the use of reference materials, and they will therefore be discussed together as conventional scales.

Conventional scales are based on the values assigned to reference materials. The assigned values are stated in standard specifications, international recommendations or other reference documents; therefore a reference material resulting in a fixed point on a conventional scale should have the same quality all over the world. CRMs of this type are certified for property values, i.e. they are measured on standard equipment with reference methods at metrological or other authorized laboratories.

It is evident that the CRMs ensure only the fixed points of a measurement scale. Measurement on a scale requires either a fixed point and a mathematical function passing through it, or two or more fixed points with stated means of interpolation between them.

NOTE Some special discontinuous scales exist, e.g. Mohs' scale for measuring hardness in geological tests. The scale is based on ten minerals to which are assigned ten grades of hardness; each harder mineral scratches the less hard one.

A conventional scale has two fundamental pillars: the certified reference material, providing the fixed point(s), and the standard specification (or similar document), giving the method of measurement. Both of them should be strictly defined to ensure the compatibility of measurements on the conventional scale.

The standard specification provides detailed information necessary to establish and use a scale based on assigned values; or it may provide protocols for the experimental and calculation procedures to be used in measurements which depend on assumptions. It is advisable to specify the requirements of the certified reference material in the same standard specification as that in which the method of measurement is described. By means of the necessary CRMs and relevant standard specifications, the user can create the measurement scale, and with the aid of such a scale can measure his sample or calibrate his instrument.

To estimate the uncertainty of a measurement on the scale, the user should consider the uncertainties in the creation of the scale and the uncertainty associated with the determination of its fixed points by the CRM. Sometimes the users demand a level of uncertainty in the end-use which is lower than the uncertainty of the fixed points defined by the CRM (e.g. in measurement of the pH of blood). They need to realize that the uncertainty of the measurements on the scale is necessarily greater than the uncertainty of the fixed points. The replicated measurement of the CRM and the setting up of a scale (the appropriate selection of the points, the characteristics and repeatability of the interpolating instrument, etc.) also contribute to the overall uncertainty.

The selection of CRMs for determining the fixed points of a scale should be directed by the required level of uncertainty of the end use. To minimize the uncertainty of the measured value on the scale, the user should employ CRMs which have been certified in terms of the units of the scale. Obviously, the user is expected to be familiar with all relevant information about the method for creating the scale and the instructions for the correct use of the CRM.

In certain cases the user can apply pure chemical compounds for determining the fixed points, if CRMs certified in the scale units are unavailable or expensive, or if their use is not necessary at the level of the uncertainty of the measurement. If this method is chosen, the user should be aware of the correlation between the purity of the material and the property on which the scale is based, and the uncertainty of the measurement can be only roughly estimated.

There is a great variety of conventional scales and the methods of application of the CRMs for determining them differ widely. Some examples are given in the referred standards (see Bibliography) to illustrate different features of some conventional scales.

5.5.2 The International Temperature Scale

The unit of thermodynamic temperature, the kelvin, is defined as the fraction 1/273,16 of the thermodynamic temperature of the triple point of water. The direct determination of the kelvin is generally taken to mean the establishment of the thermal equilibrium state of the triple point of water. Thermodynamic temperature being an intensive quantity, the determination of the kelvin through a triple point of water cell is a necessary but not sufficient condition to allow other temperatures to be measured. Some type of primary means is also necessary, e.g. some kind of gas thermometer or total radiation thermometer.

Practical measurements of temperature are not made with primary thermometers; these are difficult to use and cannot match the sensitivity and reproducibility of secondary thermometers such as platinum resistance thermometers or thermocouples. To provide a reproducible, easily accessible means of referring practical temperature measurements to thermodynamic temperature, successive internationally agreed practical temperature scales have been devised.

The current revision of this scale is the International Temperature Scale of 1990 (ITS-90).

The ITS-90 comprises:

— a set of fixed points (usually melting, freezing or triple points of gases or metals);

— interpolating secondary thermometers with specified interpolation equations;

— a set of instructions and recommendations for establishing the scale.

Over most of the range from liquid helium temperature to about 500 °C, the ITS-90 can be determined with an uncertainty of a few millikelvins.

5.5.3 The pH scale

Since absolute single-ion activities cannot be measured experimentally, it is recognized that the pH value is an inexact physical quantity. In order that measured pH be endowed with as much significance as possible, a conventional pH scale has been adopted which is defined by reference solutions with assigned values of pH. These values have been determined by measuring the electromotive force (e.m.f.) of a hydrogen-silver/silver chloride cell without transference and by a given method of calculation, based on a convention.

Various national standard specifications describe the methods of preparing and assigning pH values to the reference solutions. The uncertainty of the certified values of these reference solutions is limited to a few thousandths of a pH unit.

5.5.4 The octane-number scale

The octane-number scale is defined by ASTM-IP joint standard specifications. International Standards[8], [9] as well as a number of national standards refer to these documents. ASTM D 2699-95a/IP 237 and ASTM D 2700-95a/IP 236 describe the test methods for knock characteristics of motor fuels by the research method and by the motor method respectively. In both standards, the octane number of a fuel is determined by comparing its knocking tendency with those for blends of ASTM reference fuels of known octane number under standard operating conditions. The reference materials and blending accessories are given in annexes of both standards.

The ASTM standards refer to NIST SRM No. 1816a (iso-octane, purity 99,987 %) and SRM No. 1815a (*n*-heptane, purity 99,987 %). The principal use of these materials is in certifying the commercially produced ASTM Knock Test Reference Fuels. Specifications for these reference fuels are given in the standard, in which the suppliers are also listed. The responsibility for meeting the specifications for the reference materials rests with the suppliers. ASTM certification is based on the physical properties of the sample. Suppliers are required to test a sample of the reference material to be certified and at the same time test the corresponding SRM to provide traceability of production to an accepted reference material. A certificate is issued by ASTM to the suppliers, authorizing them to guarantee that the material shipped has been tested accordingly, and to quote the results of the tests.

6 Assessment of a measurement process

6.1 The cases to be considered

6.1.1 One laboratory

The test procedure constitutes a check of precision and/or trueness of a measurement method as applied by one particular laboratory. The laboratory uses a CRM to check its measurement process, for any particular reason, at any time.

6.1.2 Interlaboratory programme

In this case the test procedure is performed by a number of laboratories as part of an organized programme, for example as described in the various parts of ISO 5725[10]. The purpose of this programme is to establish the performance characteristics of a measurement method, against which a typical laboratory can compare its own performance.

6.2 Requirements of limits

6.2.1 General

In order to satisfy the requirement, the measurement procedure shall produce results with a precision measure and/or trueness within the predetermined limits when it is applied to a CRM. The limit of precision is usually

expressed in terms of standard deviation, and the trueness requirement is expressed in terms of the bias of the measurement results against the certified value. These limits may originate from various sources.

6.2.2 Legal limits

Legal limits are those limits which are required by statute or regulation; for example, procedures for the analysis of sulphur dioxide in air are required in many countries to have a certain precision and trueness.

6.2.3 Limits in accreditation schemes

In most cases the limits of uncertainty are consensus values agreed upon between the various participants concerned, e.g. producer, consumer and independent. For this reason, in most cases, these limits are derived from some realistic values, e.g. those obtained from the results of the certification campaign of the CRM, international tests of International Standards, etc.

6.2.4 Limits given by the user of the procedure

In this case the laboratory, or the organization of which the laboratory is a part, imposes upon itself the limits of bias and precision, e.g. limits imposed by commercial requirements.

6.2.5 Limits from previous experience

In this case the limits of bias and precision of the measurement process to be tested are based on the values obtained from previously established measurement processes.

6.3 Choice of CRM

6.3.1 Relevance to the measurement procedure

The user of the CRM must decide what properties of the CRM are relevant to his measurement procedure, taking into account the method of certification, the statement on intended use and instructions for the correct use of the CRM on the certificate.

a) **Level.** The CRM should have properties at the level appropriate to the level at which the measurement process is intended to be used, e.g. concentration.

b) **Matrix**. The CRM should have a matrix as close as possible to the matrix of the material to be subjected to the measurement process, e.g. carbon in low-alloy steel or carbon in stainless steel.

c) **Form**. The CRM may be in any physical state and form, e.g. solid, gas, etc. It may be a test piece or a manufactured article or a powder. It may need preparation. It shall be used in the same form as the sample to be measured.

d) **Quantity**. The quantity of the CRM shall be sufficient for the entire experimental programme, including some reserve if it is considered necessary. Avoid having to obtain additional new batches of the CRM later in a given measuring process.

e) **Stability.** Wherever possible the CRM should have stable properties throughout the experiment. Three situations can exist:

   1) the properties are stable and no precaution is necessary;

   2) the certified value of the properties may be influenced by storage conditions, in which case the container should be stored, both before and after its opening, in the way described on the certificate;

   3) the properties (which are changing at a known rate) at specific times are defined in a certificate supplied with the CRM.

f) **Acceptable uncertainty of the certified value.** The uncertainty of the certified value should be compatible with the precision and trueness requirements outlined in 6.2.

6.3.2 Type of certification of CRM

The choice of the type of certification of the CRM is governed by the information required for the experimental programme. Refer to ISO Guide 35[4].

6.4 Experimental procedure

6.4.1 General

The procedure for the measurement shall be fixed, i.e. a written document shall exist laying down all the details. There shall be no changes to the procedure during the course of the experiment.

6.4.2 Check of precision and trueness of a measurement process by one laboratory

6.4.2.1 General

Checking of precision of a measurement procedure as applied by a laboratory involves comparison of the within-laboratory standard deviation under repeatability conditions (or other defined conditions) and the required value of standard deviation.

Checking of trueness of a measurement process as applied by a laboratory involves comparison of the mean of the measurement results and the certified value of the CRM. The between-laboratories component of precision of the measurement procedure should be taken into account when making this comparison.

6.4.2.2 Number of replicate measurements

The number of replicate measurements, n, required depends mainly on the values of α and β and the alternative hypothesis chosen for the assessment of precision.

Table 1 shows the relation between the degrees of freedom ν (where in this case $\nu = n - 1$) and the ratio of the within-laboratory standard deviation of the measurement process, σ_w, and the required value of the within-laboratory standard deviation, σ_{wo}, for various values of β at $\alpha = 0{,}05$. For example, for $n = 10$ the probability that the variance of the measurement results will pass the appropriate χ^2-test at $\alpha = 0{,}05$ is no more than 1 % when the within-laboratory standard deviation, σ_w, of the measurement process is equal to or larger than 2,85 times the required value of σ_{wo}.

Table 1 — Ratio of the standard deviation of the measurement process to the required value for various values of β and degrees of freedom ν at $\alpha = 0,05$

ν	$\alpha = 0,05$			
	$\beta = 0,01$	$\beta = 0,05$	$\beta = 0,1$	$\beta = 0,5$
1	159,5	31,3	15,6	2,73
2	17,3	7,64	5,33	2,08
3	6,25	4,71	3,66	1,82
4	5,65	3,65	2,99	1,68
5	4,47	3,11	2,62	1,59
6	3,80	2,77	2,39	1,53
7	3,37	2,55	2,23	1,49
8	3,07	2,38	2,11	1,45
9	2,85	2,26	2,01	1,42
10	2,67	2,15	1,94	1,40
12	2,43	2,01	1,83	1,36
15	2,19	1,85	1,71	1,32
20	1,95	1,70	1,59	1,27
24	1,83	1,62	1,52	1,25
30	1,71	1,54	1,46	1,22
40	1,59	1,45	1,38	1,19
60	1,45	1,35	1,30	1,15
120	1,30	1,24	1,21	1,11

6.4.2.3 The CRM

The user should confirm the suitability of the CRM with respect to the certified value with its uncertainty, method of characterization, date of certification, statement of intended use, expiration date (particularly for a relatively unstable CRM), packaging and storage conditions and special instructions for correct use given in the certificate and the size of test portion required for the measurement process.

6.4.2.4 Measurement

The user should perform independent replicate measurements. "Independent", in a practical sense, means that a replicate result is not influenced by previous replicate results. To perform replicate measurements means to repeat the whole procedure. For example, in the chemical analyses of a solid material, the procedure should be repeated from the weighing of the test portion to the final reading or calculating of the result. Taking aliquots from the same sample solution is not independent replication.

Independent replicate measurements can be achieved in various ways depending on the nature of the process. In some, however, parallel replication is not recommended because an error committed at any step of the procedure could affect all replicates. For example, in the case of iron ore analyses, replication of the analytical procedure is carried out at different times and includes appropriate calibration.

The measurement results could, if necessary, be scrutinized for possible outliers using the rules described in ISO 5725-2. It should be noted that an excessive number of suspected outliers indicates problems in the measurement process.

6.4.2.5 Assessment of precision

The precision of the measurement process is assessed by comparing the within-laboratory standard deviation under repeatability conditions with the required value of the within-laboratory standard deviation, σ_{wo}.

Compute the average, $\bar{x}$, and standard deviation, s_w:

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} \tag{1}$$

$$s_w = \left[\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1} \right]^{1/2} \tag{2}$$

where

x_i is the individual result;

n is the number of results excluding outliers.

Compute the following ratio:

$$\chi_c^2 = \left(\frac{s_w}{\sigma_{wo}} \right)^2 \tag{3}$$

where σ_{wo} is the required value of the within-laboratory standard deviation.

$$\chi_{table}^2 = \frac{\chi_{(n-1);\,0,95}^2}{n-1}$$

= 0,95th quantile of the χ^2 distribution at degrees of freedom (n – 1) divided by the degrees of freedom (n – 1)

Decision:

$\chi_c^2 \leq \chi_{table}^2$: There is no evidence that the measurement process is not as precise as required.

$\chi_c^2 > \chi_{table}^2$: There is evidence that the measurement process is not as precise as required.

6.4.2.6 Assessment of trueness

The trueness of the measurement process is checked by comparing the average $\bar{x}$ with the certified value, μ.

There are two factors contributing to the difference between the certified value and the measurement results

a) the uncertainty of the certified value;

b) the uncertainty of the results of the measurement process being assessed expressed by its standard deviation σ_D.

For a CRM prepared in accordance with ISO Guide 35[4], the uncertainty of the certified value should be small in comparison with σ_D, the standard deviation associated with the measurement process. The following general condition is used as the criterion for acceptance:

$$-a_2 - 2\sigma_D \leq \bar{x} - \mu \leq a_1 + 2\sigma_D \quad (4)$$

where a_1 and a_2 are adjustment values chosen in advance by the experimenter according to economic or technical limitation or stipulation.

The standard deviation associated with the measurement process, σ_D, arises from the fact that a measurement procedure performed on the same material does not, in general, yield identical results every time it is applied. This fluctuation is attributed to unavoidable random errors inherent in every measurement process because the factors that may influence the outcome of a measurement cannot all be completely controlled. This random fluctuation of the measurement results should be taken into account when assessing the trueness of the procedure. For this purpose, the random fluctuation can be divided into two parts:

a) within-laboratory, or short-term fluctuation, which has a mean of zero and standard deviation of σ_w; an estimate of σ_w is given as s_w in equation (2);

b) between-laboratories fluctuation, which has a mean of zero and standard deviation of σ_{Lm}. This fluctuation is caused by one or a combination of various factors such as operators, equipment, laboratories, time, etc. When the assessment experiment is performed by only one laboratory, σ_{Lm} cannot be determined directly. In many cases it is sufficient to substitute σ_{Lm} by the standard deviation σ_1 obtained under intermediate precision conditions. Otherwise, σ_L supplied by the certificate of the CRM, or from other sources such as an appropriate International Standard, can be used to replace σ_{Lm}.

The value of σ_D^2 is therefore given as the sum of two kinds of random fluctuations:

$$\sigma_D^2 = \sigma_{Lm}^2 + \frac{s_w^2}{n} \quad (5)$$

where n is the number of replicate determinations performed for the assessment of the measurement process by the assessing laboratory.

For many measurement processes, σ_w is small in comparison with σ_{Lm}; consequently for large numbers of replications ($n > 10$), σ_D in equation (5) can be equated with σ_{Lm} or σ_L. Thus, in this case condition (4) can be simplified:

$$-a_2 - 2\sigma_{Lm} \leq \bar{x} - \mu \leq a_1 + 2\sigma_{Lm} \quad (6)$$

6.4.2.7 Example — Analysis for the iron content in iron ores

a) Purpose of investigation:

To check whether a certain analytical method (method A) is sufficiently precise and unbiased by using an iron ore CRM for the case where $a_2 = a_1 = 0$.

b) Certificate information:

The available CRM was certified by an interlaboratory programme for 13 elements, including iron.

μ = 60,73 % Fe

σ_{wo} = 0,09 % Fe

σ_L = 0,20 % Fe

c) Analysis: n = 11 after ordering:

x_i(% Fe) 60,7 60,8 60,8 60,9 60,9 60,9 61,0 61,0 61,1 61,2 61,9

Grubbs test for outliers $x_{(11)}$ is a suspect.

$$G_{(11)} = \frac{x_{(11)} - \bar{x}}{s_w} = \frac{61,9 - 61,018}{0,325} = 2,713$$

The critical value for n = 11 at 5 % is 2,234 and at 1 % is 2,485. Therefore, $x_{(11)}$ is an outlier and should be rejected. The remaining data are to be used for further computation. The new n is 10.

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} = 60,930 \text{ \% Fe}$$

$$s_w = \left[\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1} \right]^{1/2} = 0,149 \text{ \% Fe}$$

$$\chi^2_c = \left(\frac{s_w}{\sigma_{wo}} \right)^2 = \left(\frac{0,149}{0,090} \right)^2 = 2,76$$

$$\chi^2_{table} = \frac{\chi^2_{9;0,95}}{9} = 1,88$$

$$\chi^2_c > \chi^2_{table}$$

In conclusion, there is evidence to suggest that the within-laboratory standard deviation of method A is not as good as required. The method should be investigated chemically and improved.

d) Second assessment: n = 10

The second set of analytical results, after method improvement, is:

x_i = 60,94 60,99 61,04 61,06 61,06 61,09 61,10 61,14 61,21 61,24

Visual observation of the results shows no reason to suspect that there is an outlier; therefore the Grubbs test is not necessary.

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} = 61,087 \text{ \% Fe} \sim 61,09 \text{ \% Fe}$$

$$s_w = \left[\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1} \right]^{1/2} = 0,092 \text{ \% Fe}$$

$$\chi^2_c = \left(\frac{s_w}{\sigma_{wo}} \right)^2 = \left(\frac{0,092}{0,090} \right)^2 = 1,04 < \chi^2_{table}$$

$|\bar{x} - \mu|$ = 61,09 – 60,73

= 0,36 % Fe

$2\sigma_L$ = 0,40 % Fe

$|\bar{x} - \mu| < 2\sigma_L$

Therefore the method is as unbiased as required.

6.4.3 Assessment of a measurement process by an interlaboratory measurement programme

6.4.3.1 General

One of the most important criteria that a measurement process must satisfy in order to receive "widely accepted" or "standard" status is that it is capable of producing results with precision and trueness sufficient for the end-use when applied by a qualified operator. In most instances, the precision and trueness of such a candidate process are assessed by an interlaboratory measurement programme in which the participants are selected so as to provide a representative sample of the laboratories which will ultimately apply that measurement process. The procedure for conducting an interlaboratory measurement programme is described in ISO 5725-1[10].

6.4.3.2 Number of participant laboratories, p, and number of replicate measurements per laboratory, n

Ideally the values of p and n should be selected according to the limit of bias between the certified value of the CRM and the value obtained by the interlaboratory measurement programme, M, the significance level, α, and the type II risk, β. In many cases, the choice of p and n is limited by the availability of participating laboratories. The detailed procedure for computing the ideal values of p and n is described in 6.4.3.7.

6.4.3.3 Experiment

6.4.3.3.1 General

An interlaboratory measurement programme is often conducted as part of an experiment to estimate precision of the method. A detailed procedure for performing such an experiment is described in ISO 5725[10].

6.4.3.3.2 Check and distribution of the CRM

a) The CRM should be checked as described in 6.4.2.3.

b) Where subdivision of the unit of the CRM is unavoidable prior to distribution, it shall be performed with great care to avoid any additional error. Relevant International Standards on sample division should be consulted. If the unit of the CRM has a fixed form, e.g. a metal disc, the units should be selected on a random basis for distribution. If the measurement process is non-destructive, all laboratories in the interlaboratory measurement programme may be given the same unit of the CRM, but this will extend the duration of the programme.

6.4.3.3.3 Measurement

The coordinator of the interlaboratory measurement programme shall specify the number of independent replicate measurements, n, to be performed by each laboratory, and the organizational factors of interlaboratory programmes such as time limit for submission of results, the size of test portion, etc.

Methods for computing the precision measures from the results of an interlaboratory programme are described in the various parts of ISO 5725[10].

6.4.3.4 Assessment of precision

6.4.3.4.1 General

The precision of the measurement process as applied to the CRM is expressed in terms of s_w, the estimate of the within-laboratory standard deviation, and s_{Lm} the estimate of the between-laboratories standard deviation.

6.4.3.4.2 Within-laboratory precision

The estimate of the within-laboratory standard deviation of the interlaboratory comparison, s_w, can be compared with the required value of σ_{wo} in a manner analogous to that described in 6.4.2.5 by:

$$\chi_c^2 = \left(\frac{s_w}{\sigma_{wo}} \right)^2 \tag{3}$$

χ_c^2 is compared with $\chi_{table}^2 = \frac{\chi^2_{p(n-1);0,95}}{p(n-1)}$

Decision:

$\chi_c^2 \le \chi_{table}^2$: There is no evidence that the within-laboratory precision of the measurement process is not as good as required.

$\chi_c^2 > \chi_{table}^2$: There is evidence that the within-laboratory precision of the measurement process is not as good as required.

6.4.3.4.3 Between-laboratories precision

The between-laboratories precision can be assessed indirectly using the following statistic:

$$\chi_c^2 = \frac{s_w^2 + n s_{Lm}^2}{\sigma_{wo}^2 + n\sigma_L^2} \tag{7}$$

The standard deviations in the numerator are usually not estimated separately, an estimate of the numerator is obtained from ANOVA. For many test methods, the within-laboratory standard deviation is equal to or smaller than the between-laboratories standard deviation; in these circumstances equation (7) can be simplified.

$$\chi_c^2 = \frac{n s_{Lm}^2}{n\sigma_L^2} = \frac{s_{Lm}^2}{\sigma_L^2} \tag{8}$$

χ_c^2 is compared with $\chi_{table}^2 = \frac{\chi^2_{(p-1);0,95}}{p-1}$

Decision:

$\chi_c^2 \le \chi_{table}^2$: There is no evidence that the between-laboratories standard deviation of the measurement process is not as good as required.

$\chi_c^2 > \chi_{table}^2$: There is evidence that the between-laboratories standard deviation of the measurement process is not as good as required.