PROFICIENCY TESTING TUNISIA (PTTn)

GUIDE TO PROFICIENCY TESTING TUNISIA

Rés. ZARRAD - Bur. A14 - Rue Des EMERAUDES
Les Jardins du lac - 1053 TUNIS - TUNISIE
Tèl.: 0021671191520
Fax: 0021671191525
E-mail: info@pt-tn.com
Site web: www.pt-tn.com
M.F: 000MA1559157/D
R.C: B0162222018

1. SCOPE

The purpose of this document is to provide participants in Proficiency Testing Tunisia's (**PTTn**) programs with an overview of how the various types of proficiency testing programs are conducted and an explanation of how laboratory performance is evaluated. The document does not attempt to cover each step in the proficiency testing process. These are covered in (**PTTn**)'s internal procedures which are in compliance with the requirements of ISO/IEC 17043. The main body of this document contains general information about (**PTTn**)'s programs and is intended for all users of this document. The appendices contain: a glossary of terms (A); information on the evaluation procedures used for testing programs (B); and details of the evaluation of the results for calibration programs (C).

2. INTRODUCTION

The competence of laboratories is assessed by two complementary techniques. One technique is an on-site evaluation to the requirements of ISO/IEC 17025. The other technique is by proficiency testing which involves the determination of laboratory performance by means of interlaboratory comparisons, whereby the laboratory undergoes practical tests and their results are compared with those of other laboratories. The two techniques each have their own advantages which, when combined, give a high degree of confidence in the integrity and effectiveness of the assessment process. Although proficiency testing schemes may often also provide information for other purposes (e.g. method evaluation), (**PTTn**) uses them specifically for the determination of laboratory performance.

(**PTTn**) programs are divided into two different categories - **testing** interlaboratory comparisons, which involve concurrent testing of samples by two or more laboratories and calculation of consensus values from all participants' results, and **calibration** interlaboratory comparisons in which one test item is distributed sequentially among two or more participating laboratories and each laboratory's results are compared to reference values. A subset of interlaboratory comparisons are one-off practical tests (refer Section 5.8) and measurement audits (refer Section 6.10) where a well characterized test item is distributed to one laboratory and the results are compared to reference values.

Proficiency testing is carried out by (**PTTn**) staff. Technical input for each program is provided by Technical Advisers. The programs are conducted using collaborators for the supply and characterization of the samples and test items. All other activities are undertaken by (**PTTn**).

2.1 CONFIDENTIALITY

All information supplied by a laboratory as part of a proficiency testing program is treated as confidential. There are, however, three exceptions. Information can be disclosed to third parties:

- with the express approval of the client(s);
- when (**PTTn**) has an agreement with or requirement in writing from the State Government which requires the provision of information, and the relevant parties/clients have been informed in writing of such agreement or requirement;

- When (**PTTn**) has any concerns about the conduct of any aspect of the proficiency testing process or in relation to any safety, medical or public health issues identified in the proficiency testing process.

2.2 FUNDING

(**PTTn**) charges a participation fee for each program. This fee varies from program to program and participants are notified accordingly, prior to a program's commencement.

3. REFERENCES

1. ISO/IEC 17043:2010 Conformity assessment: General requirements for proficiency testing
2. ISO/IEC 17025:2017 General requirements for the competence of testing and calibration laboratories
3. ISO/IEC 17011:2017 Conformity assessment: General requirements for accreditation bodies accrediting conformity assessment bodies
4. ISO/IEC Guide 98-3:2008 Uncertainty of measurement – Part 3: Guide to the expression of uncertainty in measurement (GUM)
5. ISO 13528:2015 Statistical methods for use in proficiency testing by interlaboratory comparisons

4. QUALITY MANAGEMENT OF PROFICIENCY TESTING SCHEMES

In accordance with best international practice, (**PTTn**) maintains and documents a quality system for the conduct of its proficiency testing programs. This quality system complies with the requirements specified in ISO/IEC 17043:2010.

5. TESTING INTERLABORATORY COMPARISONS

5.1 INTRODUCTION

(**PTTn**) uses collaborators for the supply, preparation and homogeneity testing of samples. All other activities are undertaken by (**PTTn**) and technical input is provided by program Technical Advisers.

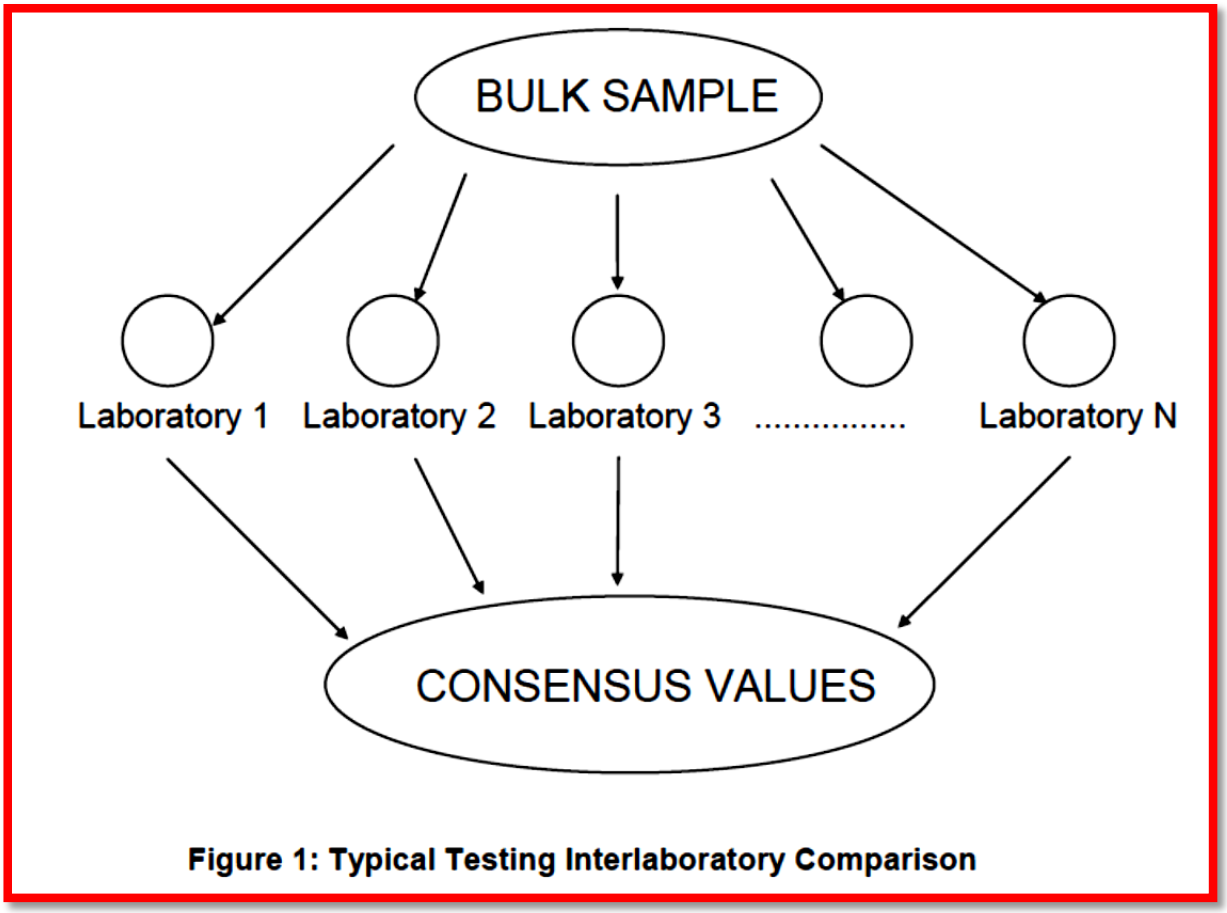


Figure 1: Typical Testing Interlaboratory Comparison

5.2 WORKING GROUP AND PROGRAM DESIGN

Once a program has been selected, a small working group is formed. This group usually comprises one or more Technical Advisers, and the (**PTTn**) Scientific Officer who will act as the Program Coordinator.

It is most important that at least one, but preferably two, technical experts are included in the planning of the program and in the evaluation of the results. Their input is needed in at least the following areas:

- Nomination of tests to be conducted, range of values to be included, test methods to be used and number/design of samples required;
- Preparation of paperwork (instructions and results sheet) particularly with reference to reporting formats, number of decimal places to which results should be reported and correct units for reporting;
- Identification and resolution of any difficulties expected in the preparation and maintenance of homogeneous proficiency test items, or in the provision of a stable assigned value for a proficiency test item;
- Technical commentary in the final report and, in some cases, answer questions from participants.

An appropriate statistical design is essential and therefore must be established during the preliminary stages of the program (see Appendix B for further details).

5.3 SAMPLE SUPPLY AND PREPARATION

The Program Coordinator is responsible for organizing the supply and preparation of the samples. It is often the case that one of the Technical Advisers will also act as the program's sample supplier. In any case, the organization preparing the test items is always one that is considered by (**PTTn**) to have demonstrable competence to do so.

In some cases preparation of sample is subcontracted by external accredited ISO 17025 laboratory.

Sample preparation procedures are designed to ensure that the samples used are as homogeneous and stable as possible, while still being similar to samples routinely tested by laboratories. A number of each type of sample are selected at random and tested, to ensure that they are sufficiently homogeneous for use in the proficiency program. Whenever possible, this is done prior to samples being distributed to participants. The results of this homogeneity testing are analyzed statistically and may be included in the final report.

5.4 DOCUMENTATION

The main documents associated with the initial phase of a proficiency program are:

(a) Letter of Intent
This is sent to prospective participants to advise that the program will be conducted and provides information on the type of samples and tests which will be included, the schedule and participation fees.

(b) Instructions to Participants
These are carefully designed for each individual program and participants are always asked to adhere closely to them.

(c) Results Sheet
For most programs a pro-forma results sheet is supplied to enable consistency in the statistical treatment of results.

Instructions and Results Sheets may be issued with, or prior to, the dispatch of samples.

5.5 PACKAGING AND DISPATCH OF SAMPLES

The packaging and method of transport of the samples are considered carefully to ensure that they are adequate and able to protect the stability and characteristics of the samples. In some cases, samples are packaged and dispatched from the organization supplying them, in other cases they are shipped to (**PTTn**) for this distribution. It is also ensured that certain restrictions on transport such as dangerous goods regulations or customs requirements are complied with.

5.6 RECEIPT OF RESULTS

Results from participating laboratories for (**PTTn**) testing programs are required to be sent to either our office. A 'due date' for return of results is set for each program, usually allowing laboratories two to three weeks to test the samples. If any results are outstanding after the due date, reminders are issued, however, as late results delay the data analysis, these may not be included. Laboratories are requested to submit all results on time.

5.7 ANALYSIS OF DATA AND REPORTING OF RESULTS

Results are usually analyzed together (with necessary distinctions made for method variation) to give consensus values for the entire group. The results received from participating laboratories are entered and analyzed as soon as practicable so that the final report can be issued to participants within six weeks of the due date for results.

The evaluation of the results is by calculation of robust z-scores, which are used to identify any outliers. Summary statistics and charts of the data are also produced, to assist with interpretation of the results. A detailed account of the procedures used to analyze results appears in Appendix B.

Participants are issued with an individual laboratory summary sheet (refer Appendix B) which indicates which, if any, of their results were identified as outlier results. Where appropriate, it also includes other relevant comments (e.g. reporting logistics, method selection).

A final report is produced at the completion of a program and includes data on the distribution of results from all laboratories, together with an indication of each participant's performance. This report typically contains the following information:

(a) introduction;
(b) Features of the program - number of participants, sample description, tests to be carried out;
(c) Results from participants;
(d) Statistical analysis, including graphical displays and data summaries (outlined in Appendix B);
(e) A table summarizing the outlier results;
(f) (**PTTn**) and Technical Adviser's comments (on possible causes of outliers, variation between methods, overall performance etc.);
(g) Sample preparation and homogeneity testing information; and
(h) A copy of the instructions to participants and results sheet.

Note: Outlier results are the results which are judged inconsistent with the consensus values (refer Appendix A for definition).

The final program report is released through email,

5.8 OTHER TYPES OF TESTING PROGRAMS

(**PTTn**) conducts some proficiency testing activities which do not exactly fit the model outlined in Section 5.1. These include known-value programs where samples with well-established reference values are distributed (e.g. slides for asbestos fiber counting).

Further examples are one-off practical tests where material of known composition (e.g. certified reference material) is presented to one laboratory. This type of activity is also extensively used in the calibration area (refer Section 6.10, Measurement Audits). These activities do not, or by their nature cannot, use the usual consensus values as the basis for the evaluation of performance.

Some of (**PTTn**)'s testing interlaboratory comparisons do not produce quantitative results - i.e. qualitative programs where the presence or absence of a particular parameter is to be determined (e.g. pathogens in food). By their nature the results must also be treated differently from the procedures outlined in Appendix B.

6. CALIBRATION INTERLABORATORY COMPARISONS

6.1 INTRODUCTION

(**PTTn**) uses collaborators for the supply and calibration of test items. All other activities are undertaken by (**PTTn**) and technical input is provided by program Technical Advisers. Each calibration laboratory has its capability uniquely expressed both in terms of its ranges of measurements and the least measurement uncertainty (or best accuracy) applicable in each range. Because calibration laboratories are generally working to different levels of accuracy, it is not normally practicable to compare results on a group basis such as in interlaboratory testing programs. For calibration programs, we need to determine each individual laboratory's ability to achieve the level of accuracy for which they have nominated (their least measurement uncertainties).

The assigned (reference) values for a calibration program are not derived from a statistical analysis of the group's results. Instead they are provided by a Reference Laboratory which must have a higher accuracy than that of the participating laboratories. For (**PTTn**) interlaboratory comparisons, the Reference Laboratory is Tunisia's National Measurement Institute (NMI) which maintains Tunisia's primary standards of measurement or accredited calibration laboratory iso 17025,

Another difference between calibration and testing programs is that there is usually only one test item (also known as an artefact) which has to be distributed sequentially around the participating laboratories, making these programs substantially longer to run. Consequently, great care has to be taken to ensure the measurement stability of the test item.

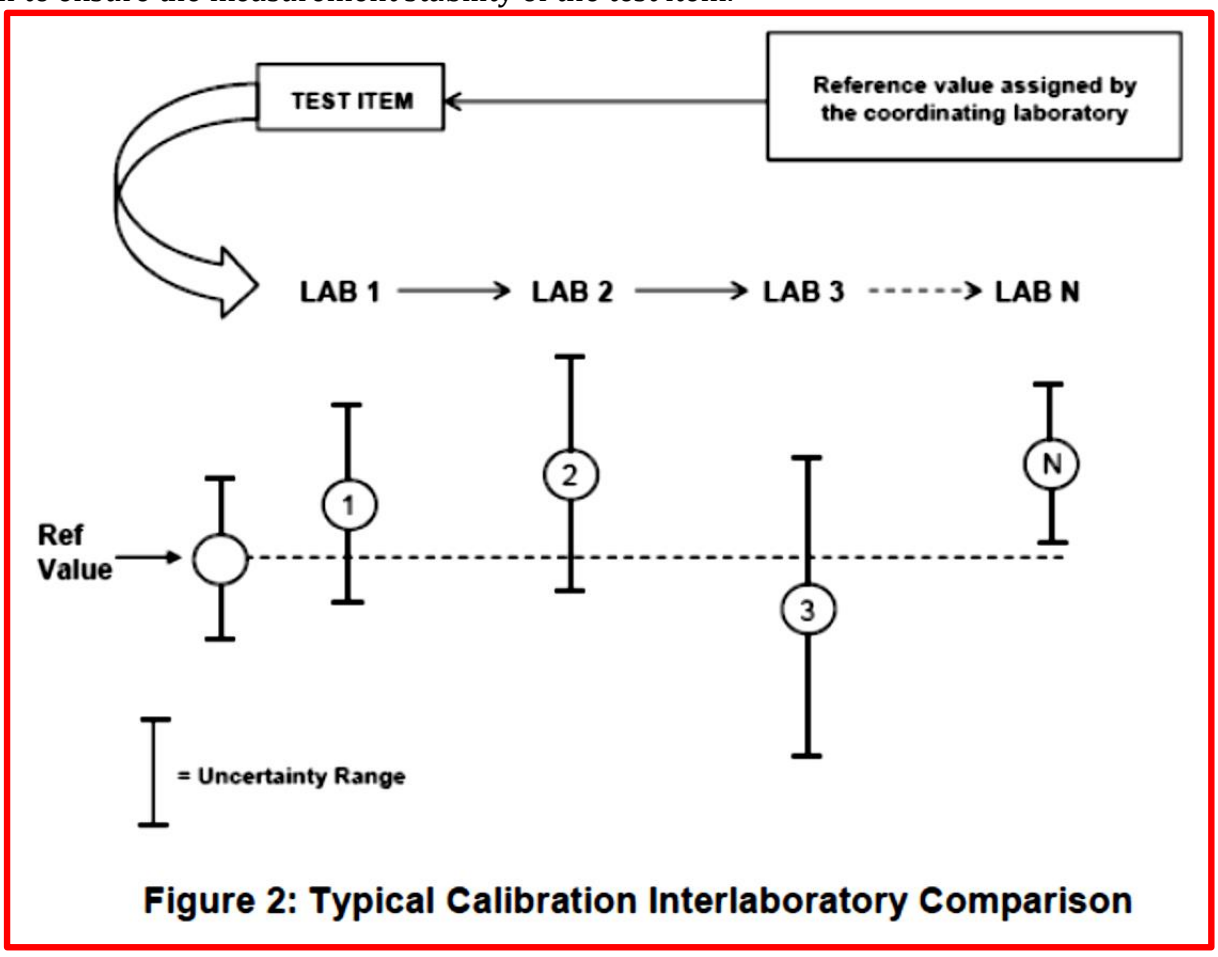


Figure 2: Typical Calibration Interlaboratory Comparison

In Figure 2, LAB 3 has a larger uncertainty range than LAB 1. This means that LAB 1 has the capability to calibrate higher accuracy instruments. This situation, where laboratories are working to different levels of accuracy, is valid provided that each laboratory works within their capabilities and that their nominated level of accuracy (measurement uncertainty) is suitable for the instrument being calibrated.

6.2 PROGRAM DESIGN

Once a program has been selected, a small working group is formed. This group usually comprises one or more Technical Advisers and a (**PTTn**) Scientific Officer who will act as the Program Coordinator. The group decides on the measurements to be conducted, how often the test item will need to be recalibrated and the range of values to be measured. They also formulate instructions and results sheets. (**PTTn**) programs are designed so that it will normally take no more than eight hours for each participant to complete the measurements.

6.3 TEST ITEM SELECTION

Because there can often be a substantial difference in the nominated measurement uncertainties of the participating laboratories, the test item must be carefully chosen. For example, it would be inappropriate to send a 3½ digit Multimeter to a laboratory that had a nominated measurement uncertainty of 5 parts per million (0.0005%) because the resolution, repeatability and stability of such a test item would limit the measurement uncertainty the laboratory could report to no better than 0.05%. What is necessary is a test item with high resolution, good repeatability, good stability and an error that is large enough to be a meaningful test for all participants.

In some intercomparisons (especially international ones), the purpose may not only be to determine how well laboratories can measure specific points but also to highlight differences in methodology and interpretation.

6.4 DOCUMENTATION

A Letter of Intent is sent to all potential participants to advise that the program will be conducted and to provide as much information as possible.

Instructions to Participants are carefully designed for each individual program and it is essential to the success of the program that the participating laboratories adhere closely to them. For most programs a pro-forma Results Sheet is used, to ensure that laboratories supply all the necessary information in a readily accessible format.

6.5 TEST ITEM STABILITY

The test item is distributed sequentially around the participating laboratories. To ensure its stability, it is usually calibrated at least at the start and at the end of the circulation. For test items whose values may drift during the course of the program (e.g. resistors, electronic devices, etc.) more frequent calibrations and checks are necessary.

6.6 EVALUATION OF PERFORMANCE

As stated in Section 6.1, calibration laboratories are generally working to different levels of accuracy. Consequently, their performance is not judged by comparing their results with those of the other laboratories in an interlaboratory comparison. Instead, their results are compared only to the Reference Laboratory's results and their ability to achieve the accuracy for which they have nominated is evaluated by calculating the En number. For further details please refer to Appendix C.

6.7 REFERENCE VALUES

Tunisia's National Measurement provides most of the reference values for (**PTTn**)'s Calibration interlaboratory comparisons. The majority of the participating laboratories' reference equipment is also calibrated by Tunisia's National Measurement.

As stated previously, it is important to select test items with high resolution, good repeatability and good stability. This is to ensure that these factors do not contribute significantly to the reference value uncertainty. Likewise, the Reference Laboratory must have the capability to assign measurement uncertainties that are better than the participating laboratories. Otherwise it will be more difficult to evaluate each laboratory's performance.

Where a test item has exhibited drift, the reference values will usually be derived from the mean of the Reference Laboratory calibrations carried out before and after the measurements made by the participating laboratories. Where a step change is suspected, then the reference values will be derived from the most appropriate Reference Laboratory calibration.

6.8 MEASUREMENT UNCERTAINTY (MU)

To be able to adequately compare laboratories they must report their uncertainties with the same confidence level. A confidence level of 95% is the most commonly used internationally.
Laboratories should also use the same procedures to estimate their uncertainties as given in the ISO Guide.

Laboratories should not report uncertainties smaller than their nominated measurement uncertainty.

6.9 REPORTING

An individual summary sheet is sent to laboratories to give them feedback on their performance.
The summary sheet states the En values for each measurement based on the preliminary reference values and usually does not contain any technical commentary.

A Final Report is issued on the (**PTTn**) website (www.pt-tn.com) at the conclusion of the program. This typically contains more information than is provided in the summary sheet - including all participant's results and uncertainties, final En numbers, technical commentary and graphical displays.

6.10 MEASUREMENT AUDITS

The term measurement audit is used by (**PTTn**) to describe a practical test whereby a well characterized and calibrated test item (or artefact) is sent to a single laboratory and the results are compared with a reference value (supplied by NMI or accredited laboratory ISO 17025).

Procedures are the same as for a normal interlaboratory comparison except that usually only a simple report is generated.

APPENDIX A
GLOSSARY OF TERMS

GLOSSARY OF TERMS

Further details about many of these terms may be found in either Appendix B (testing programs) or Appendix C (calibration programs). A number of these are also defined in ISO/IEC 17043.

Assigned value: value attributed to a particular property of a proficiency test item

Consensus value: an assigned value obtained from the results submitted by participants (e.g. for most testing programs the median is used as the assigned value)

E_n number: stands for error normalized and is the internationally accepted quantitative measure of laboratory performance for calibration programs (see formula in Appendix C)

False negative: failing to report the presence of a parameter (e.g. analyte, organism) which is present in the sample

False positive: erroneously reporting the presence of a parameter (e.g. analyte, organism) which is absent from the sample

Interlaboratory: organization, performance and evaluation of measurements or tests on

Comparison: the same or similar items by two or more laboratories in accordance with predetermined conditions

Measurement: non-negative parameter characterizing the dispersion of the quantity

Uncertainty (MU): values being attributed to a measurand, based on the information used

Outlier: observation in a set of data that appears to be inconsistent with the remainder of that set, e.g. absolute z-score greater than or equal to three (i.e. 3.0) for testing programs

Reference value: an assigned value which is provided by a Reference Laboratory

Robust statistics: statistical method insensitive to small departures from underlying assumptions surrounding an underlying probabilistic model

Z-score (Z): a normalized value which assigns a "score" to the result(s), relative to the other numbers in the group

APPENDIX B
EVALUATION PROCEDURES FOR TESTING PROGRAMS

B.1 INTRODUCTION

This appendix outlines the procedures (**PTTn**) uses to analyze the results of its proficiency testing programs. It is important to note that these procedures are applied only to testing programs, not calibration programs (which are covered in Appendix C). In testing programs the evaluation of results is based on comparison to assigned values which are usually obtained from all participants' results (i.e. consensus values).

The statistical procedures described in this appendix have been chosen so that they can be applied to a wide range of testing programs and, whenever practicable, programs are designed so that these 'standard' procedures can be used to analyze the results. In some cases, however, a program is run where the 'standard' statistical analyses cannot be applied - in these cases other, more appropriate, statistical procedures may be used.

For all programs the statistical analysis is only one part of the evaluation of the results. If a result is identified as an outlier, this means that statistically it is significantly different from the others in the group, however, from the point of view of the specific science involved (e.g. chemistry), there may be nothing "wrong" with this result. This is why the assessment of the results is always a combination of the statistical analysis and input by Technical Advisers (who are experts in the field). In most cases the Technical Adviser's assessment matches the statistical assessment.

B.2 STATISTICAL DESIGN:

In order to assess the testing performance of laboratories in a program, a robust statistical approach, using z-scores, is used. Z-scores give a measure of how far a result is from the assigned value, and give a "score" to each result relative to the other results in the group.
Section B.5 describes the method used by (**PTTn**) for calculating z-scores.

For most testing programs, simple robust z-scores are calculated for each sample.
Occasionally, the samples in a program may be paired and robust z-scores can be calculated for the sample pair. If paired samples are used they may be identical ("blind duplicates") or slightly different (i.e. the properties to be tested are at different levels). The pairs of results which are subsequently obtained fall into two categories: uniform pairs, where the results are expected to be the same (i.e. the samples are identical or the same sample has been tested twice); and split pairs, where the results should be slightly different. The pairing of samples allows the assessment of both between-laboratories and within-laboratory variation in a program.

One of the main statistical considerations made during the planning of a program is that the analysis used is based on the assumption that the results will be approximately normally distributed. This means that the results roughly follow a normal distribution, which is the most common type of statistical distribution (see Figure 3).

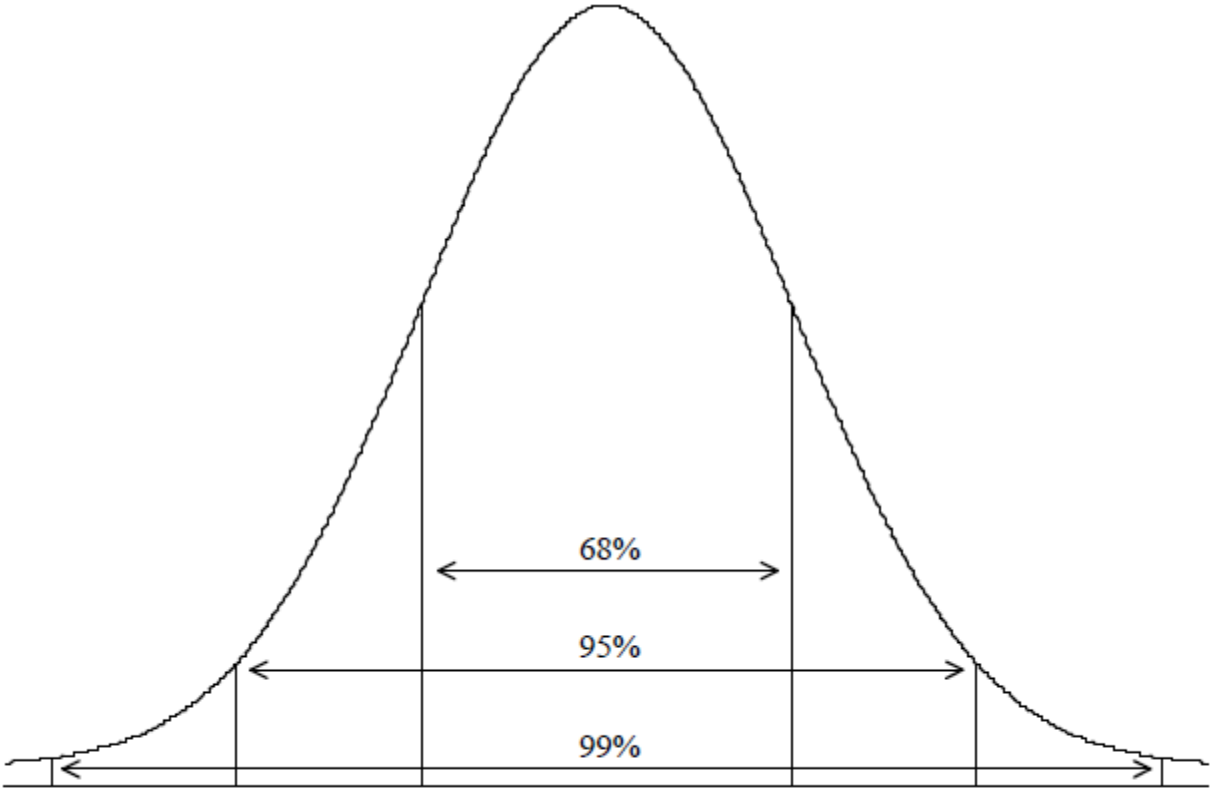


Figure 3: The Normal Distribution

The normal distribution is a "bell-shaped" curve, which is continuous and symmetric, and is defined such that about 68% of the values lie within one standard deviation of the mean, 95% are within two standard deviations and 99% are within three. To ensure that the results for a program will be approximately normal the working group (in particular the Technical Adviser) must think carefully about the results which might be obtained for the samples which are to be used.

For example, for the results to be continuous, careful consideration must be given to the units and number of decimal places requested - otherwise the data may contain a large number of repeated values. Another problem which should be avoided is when the properties to be tested are at very low levels - in this case the results are often not symmetric (i.e. skewed towards zero).

B.3 Data Preparation

Prior to commencing the statistical analysis, a number of steps are undertaken to ensure that the data collected is accurate and appropriate for analysis.

As the results are submitted to (**PTTn**), care is taken to ensure that all of the results are entered correctly. Once all of the results have been received (or the deadline for submission has passed), the entered results are carefully double-checked. It is during this checking phase that gross errors and potential problems with the data in general may be identified.

In some cases the results are then transformed - for example, for microbiological count data the statistical analysis is usually carried out on the log10 of the results, rather than the raw counts. When all of the results have been entered and checked (and transformed if necessary) histograms of the data - which indicate the distribution of the results - are generated to check the assumption of normality.

These histograms are examined to see whether the results are continuous and symmetric. If this is not the case the statistical analysis may not be valid. One problem which may arise is that there are two distinct groups of results on the histogram (i.e. a bi-modal distribution). This is most commonly due to two test methods giving different results, and in this case it may be possible to separate the results for the two methods and then perform the statistical analysis on each group.

B.4 SUMMARY STATISTICS

(**PTTn**) organize a wide range of schemes, which may include qualitative, quantitative, semi-quantitative and interpretive tests. Different approaches to data analysis may therefore be used, the most common approaches being described below. Further information on the statistical approach for specific schemes is also provided in the Scheme Descriptions and Scheme Reports.

The advantages of using a performance score are:
· Results can be expressed in a form that is relatively easy to interpret and understand
· Results can be summarised in graphical or tabular form to depict overall performance
· A performance score allows participants to directly compare their own result with others
· If consistent statistical values are applied, a performance score enables participants to monitor trends in their own performance, over time.

When reviewing results, participants should take into account the methods used to analyse the data and to assess performance, and should review their performance in context, taking into account performance of the whole dataset.

B.4.1 QUALITATIVE SCHEMES

For qualitative tests, participant results will be compared against the intended result, also called the assigned value, based on formulation or expert assessment. A result which is the same as the assigned value is considered satisfactory. This approach is also used for quantitative tests when the target analyte is absent and for semi-quantitative tests where the assigned value may be a range of results. For interpretive schemes where the result is subjective rather than quantifiable, a model answer produced by appropriate experts will be published in the report.

B.4.2 QUANTITATIVE SCHEMES

For quantitative data, participants are assessed on the difference between their result and the assigned value; with this difference being represented by a performance score called a z or z' (z prime) score.

B.4.3 SETTING ASSIGNED VALUES

The assigned value is the value selected as being the best estimate of the 'true value' for the parameter under test. The method used to determine the assigned value may vary depending upon the particular scheme and test parameter, and is detailed in the relevant scheme description, along with details of the traceability in each case.

For quantitative tests, all assigned values are derived in accordance with ISO 13528. Where it is appropriate, practicable and technically feasible the assigned value will be derived through formulation to provide metrological traceability; the associated uncertainty of the value can therefore be estimated.

However, in most cases it will not be possible to use formulation or certified reference materials to set the *assigned value and a consensus value will be the only practicable and technically feasible approach to use*. When the assigned value is determined from the consensus value of participant results, or from expert laboratories, *robust statistical methods* are used for calculation of the consensus value. The uncertainty of the assigned value is then estimated as below described.

B.5 Robust Z-scores and Outliers

$$z\ score = \frac{(x_i - x_{pt})}{\sigma_{pt}}$$ where;

x_i = the result reported by the participant
x_{pt} = the assigned value
σ_{pt} = standard deviation for proficiency assessment

The z score expresses performance in relation to an acceptable variation of the participant result to the assigned value. A z score of 2 represents a result that is 2 x σ_{pt} from the assigned value.

Where alternative scoring methods are used, full details will be given in the Scheme Description and/or report.

B.5.1 STANDARD DEVIATION FOR PROFICIENCY ASSESSMENT (SDPA)

The method used to determine the SDPA may vary depending upon the particular scheme and test parameter. All SDPA's are derived in accordance with ISO 13528. When the SDPA is determined from the dispersion of participant results, robust statistical methods are used for the standard deviation. A fixed, fit for purpose SDPA value is preferable as this enables z scores to be compared from round to round to demonstrate general trends. This fixed value may be absolute or expressed as a percentage of the assigned value.
Where applicable, the value of SDPA is reported in the Scheme Description and/or report.

The assigned value (X_{pt}) has a standard uncertainty ($U(X_{pt})$) that depends upon the method used to derive the assigned value. When the assigned value is determined by the consensus of participants' results, the estimated standard uncertainty of the assigned value can be calculated by;

$U(X_{pt})$ = 1.25 x Robust standard deviation/$\sqrt{n}$ where n = number of results

When the assigned value is determined by formulation, the standard uncertainty is estimated by the combination of uncertainties of all sources of error.

If **$U(X_{pt})$ is ≤ 0.3 x SDPA**, then the uncertainty of the assigned value can be considered negligible and need not be considered in the interpretation of results.

If **$U(X_{pt})$ is > 0.3 x SDPA**, then the uncertainty of the assigned value is not negligible in relation to the SDPA and so z' (z prime) scores, which include the uncertainty of the assigned value in their calculation, will be reported in place of z scores.

z' scores are calculated as follows:

$$z' = \frac{(x_i - x_{pt})}{\sqrt{\sigma_{pt}^2 + u(x_{pt})^2}}$$

Where

x_{pt} = the assigned value
x_i = participant result
σ_{pt} = standard deviation for proficiency assessment
$u(x_{pt})$ = standard uncertainty of the assigned value x_{pt}

The magnitude of z' scores should be interpreted in the same way as z scores.

For small data sets (generally with less than 8 results) there will be increased uncertainty around the assigned value if using consensus values from participants' results

B.5.2 CRITERIA FOR THE EVALUATION OF PERFORMANCE OF PARTICIPANTS
Results will be considered correct, and therefore have satisfactory performance.
For quantitative examinations, the following interpretation is given to z score results.

$|z| \leq 2.00$ satisfactory result
$2.00 < |z| < 3.00$ Questionable result
$|z| \geq 3.00$ unsatisfactory result

Where other performance techniques are used these are described in the Scheme Description and/or report.

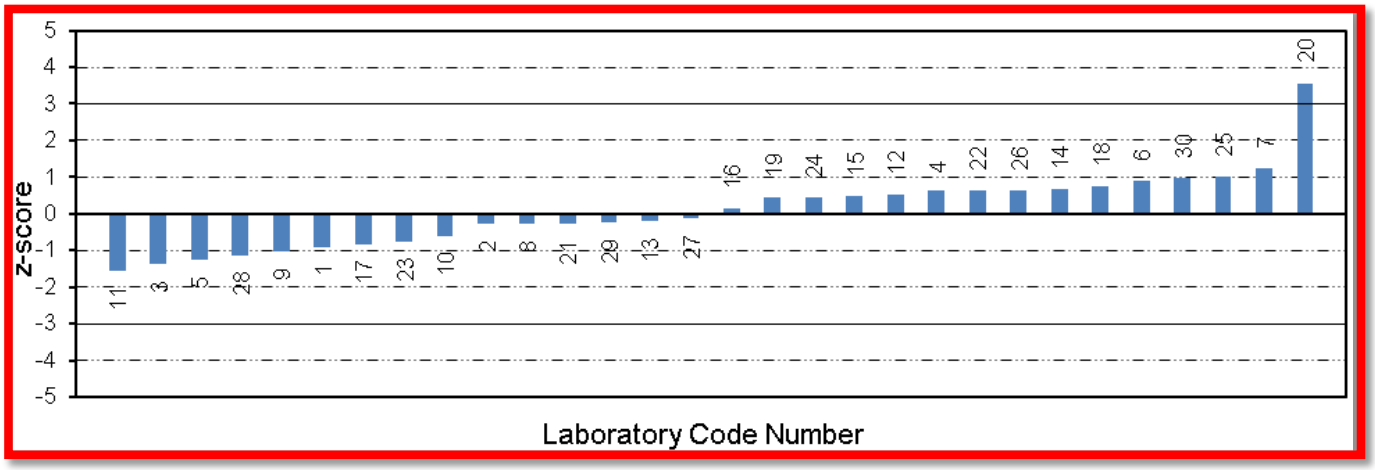
For small data sets (generally with less than 8 results) there will be increased uncertainty around the assigned value if using consensus values from participants' results. For those analyte that use a formulation or reference value as the assigned value and a fixed fit for purpose SDPA z scores will be provided. Where the assigned value and/or SDPA is based on participant results, performance scores will be given for information only. For data sets with very limited results or where the spread of results is large, z scores may not be provided.

B.5.3 REMOVAL OF ERRORS AND BLUNDERS
Although robust estimators are used in order to minimise the influence of outlying results, extreme results or results that are identifiably invalid should not be included in the statistical analysis of the data. For example, these may be results caused by calculation errors or the use of incorrect units. However, such results can be difficult to identify by the PT organiser. For this reason, the robust mean and standard deviation will be calculated as above, but those results that are out of the range of the assigned value **5 x SDPA** will be excluded and the robust mean and standard deviation will then be recalculated. These recalculated values will be used for the statistical analysis. All results, including excluded results, will be given performance scores.

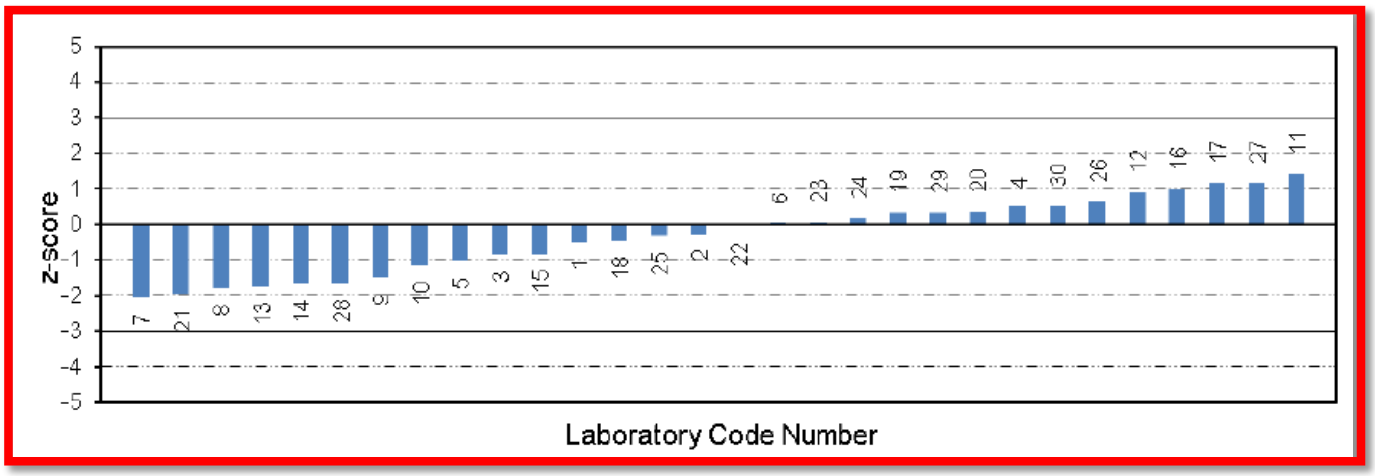
B.5.4 TREND ANALYSIS
A single result simply reflects the performance of the laboratory on the particular day that the test was carried out and can therefore only give limited information. Frequent participation in PT schemes over time can give greater insight into long-term performance and can help identify where an internal bias may be occurring. One of the best methods of summarising z scores over time is graphically, as this gives a clear overview, and is less prone to misinterpretation than numerical methods. Participants are therefore advised to monitor their PT results over time.

B.6 GRAPHICAL DISPLAYS
Examples: Ordered Z-Score Charts

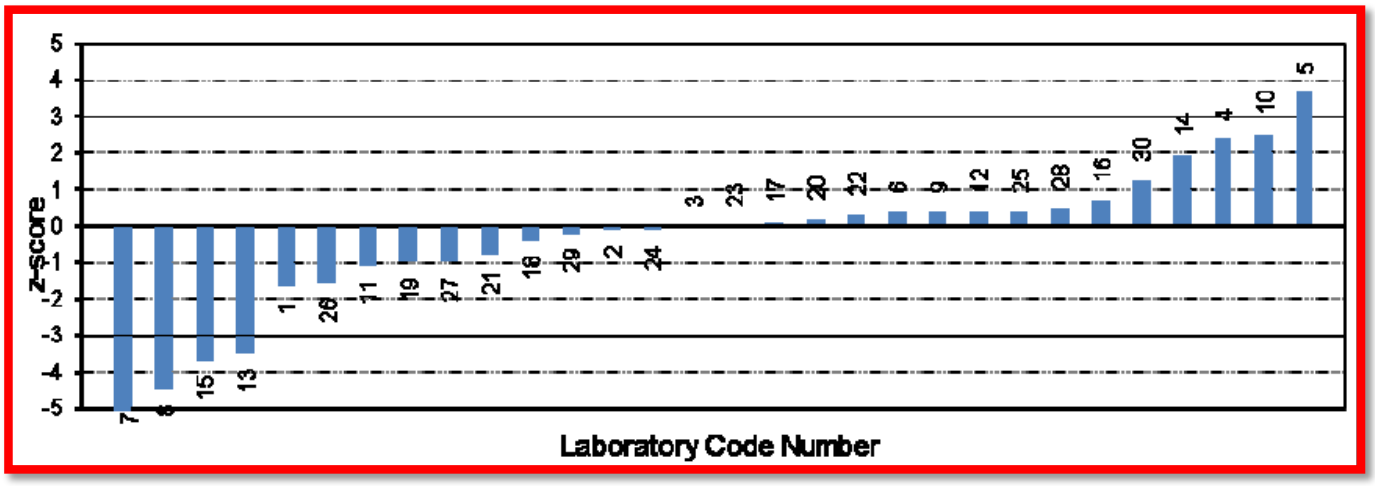
Total Solids - Sample PTT 1 - Robust Z-Scores



Total Suspended Solids - Sample PTT 1 - Robust Z-Score



Total Dissolved Solids - Sample PTT 1 - Robust Z-Scores



B.7 LABORATORY SUMMARY SHEETS

In addition to the final report, which contains complete details of the statistical analysis, an individual summary sheet is prepared for each participant. This laboratory summary sheet contains all of the participant's results, alongside the statistics for that test/sample and the associated z-scores. Comments about the program in general and specific to the laboratory (if necessary) are also included.

APPENDIX C
EVALUATION PROCEDURES FOR CALIBRATION PROGRAMS

C.1 INTRODUCTION

This appendix outlines the procedures (**PTTn**) uses to evaluate the results of its *calibration* programs and *measurement audit programs* (refer to Appendix B for procedures applicable to *testing* programs). The procedures used by (**PTTn**) are consistent with those used for international calibration programs run by the European Cooperation for Accreditation (EA) and Asia Pacific Laboratory Accreditation Cooperation (APLAC).

C.2 CALIBRATION PROGRAM

As stated in Section 6.6, (**PTTn**) uses the En number to evaluate each individual result from a laboratory. En stands for **E**rror **n**ormalized and is defined as:-

$$E_n = \frac{Error_{lab} - \overline{R}}{\sqrt{U_{lab}^2 + U_{\overline{R}}^2}}$$

E_n the normalized error,

$Error_{lab}$: the participant's result,

$\overline{R}$: the reference value,

U_{lab} : the expanded uncertainty of a participant's result,

$U_{\overline{R}}$: the expanded uncertainty of the reference value.

For a result to be acceptable the En number should be between -1.0 and +1.0 i.e. $|En| \leq 1.0$. (The closer to zero the better.)

In *testing* interlaboratory comparisons a laboratory's z-score gives an indication of how close the laboratory's measurement is to the assigned value, however, in *calibration* interlaboratory comparisons the En numbers indicate whether laboratories are within their particular measurement uncertainty of the reference value (assigned value).

The En numbers do not necessarily indicate which laboratory's result is closest to the reference value. Consequently, calibration laboratories reporting small uncertainties may have a similar En number to laboratories working to a much lower level of accuracy (i.e. larger uncertainties).

In a series of similar measurements a normal distribution of En numbers would be expected. So when considering the significance of any results with |En| marginally greater than 1.0, all the results from that laboratory are evaluated to see if there is a systematic bias e.g. consistently positive or consistently negative values of En.

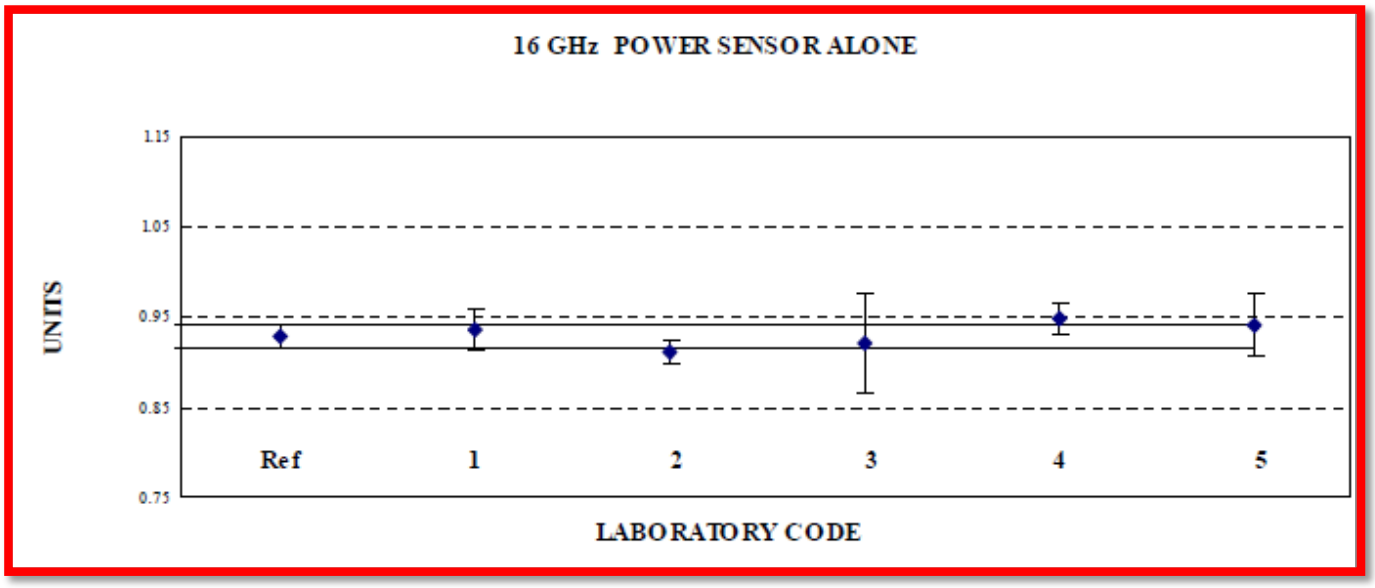
A sample of results from a radio frequency power interlaboratory comparison, their corresponding reported uncertainties and En numbers are tabulated below. The result for laboratory 2 is considered unsatisfactory.

16 GHz Power Sensor Alone

Lab Code	Results	U_{95}	E_n
REF	0.929	0.011	
1	0.936	0.022	0.28
2	0.911	0.012	**-1.09**
3	0.921	0.054	-0.14
4	0.949	0.018	0.94
5	0.942	0.035	0.35

C.3 GRAPHICAL DISPLAYS FOR CALIBRATION PROGRAM

Graphs of reported results and their associated uncertainties are included in final reports for *calibration* programs. The example graph below shows a plot of the results tabulated in Section C.2. Each laboratory's result is represented by a ■ mark. The bars protruding above and below the ■ mark represent that laboratory's reported measurement uncertainty, that is, the region in which the laboratory has statistically calculated (with a 95% confidence level) that the "true value" may lie, or in other words, their estimate of how accurately they can measure.



It is important to note however that the graphs are an illustration of the data only and allow a broad comparison of all participants' results/uncertainties. They do not represent an assessment of results (this is done by the En numbers).

C.4 MEASUREMENT AUDIT PROGRAMS

A sample of results from a pressure transducer *measurement audit*, the laboratory's corresponding reported uncertainties and En numbers are tabulated below. The results for decreasing applied pressures at 9.9999 MPa, 7.5000 MPa and 5.0000 MPa are considered unsatisfactory.

10 MPa Pressure Transducer

APPLIED PRESSURE	REF VALUE MPa	REF U_{95} MPa	LAB MEAN MPa	LAB U_{95} MPa	E_n NO.
5.0000	4.8983	0.0014	4.8982	0.002	-0.03
7.5000	7.3478	0.0014	7.3466	0.002	-0.46
9.9999	9.7973	0.0019	9.7970	0.004	-0.08
9.9999	9.8133	0.0025	9.7972	0.004	**-3.72**
7.5000	7.3605	0.0031	7.3462	0.002	**-3.88**
5.0000	4.9074	0.0025	4.8971	0.002	**-3.51**

Graphs of reported results and their associated uncertainties are provided for *measurement audit* programs when necessary.

C.5 MEASUREMENT UNCERTAINTY (MU)

The measurement uncertainty reported by the laboratory is used in the En number. The test items used in these programs usually have sufficient resolution, repeatability and stability to allow the laboratory to report an uncertainty equal to their claimed "*best measurement capability*".

END OF DOCUMENT