SRI LANKA ACCREDITATION BOARD
for CONFORMITY ASSESSMENT

GUIDELINE FOR

PROFICIENCY TESTING PROGRAMMES

AMENDMENT SHEET

Date	Doc No	Page/ Section	Description of change	Revision No	Approval
2009-07-10	AC-GL(P)-08	Pg 6 4.5	Inclusion of Homogeneity, Stability and traceability of test items	01	DD

CONTENTS

1. INTRODUCTION

Proficiency Testing through Interlaboratory comparisons may be used, for example, to:

a) Determine the performance of individual laboratories for specific tests or measurements and to monitor laboratories' continuing performance.

b) Identify problems in laboratories and initiate remedial actions which may be related to, for example, individual staff performance or calibration of instrumentation.

c) Establish the effectiveness and comparability of new test or measurement methods and similarly to monitor established methods.

d) Provide additional confidence to laboratory clients.

e) Identify inter laboratory differences.

f) Determine the performance characteristics of a method – often known as collaborative trials;

g) Assign values to reference materials (RMs) and assess their suitability for use in specific test or measurement procedures.

Testing and Calibration Laboratories that wish to obtain and maintain SLAB accreditation must comply with the requirements as laid down by the SLAB. Currently accreditation is granted in three fields of testing, biological, chemical, and mechanical and three fields of calibration, electro-technical, mechanical, and thermal & optical measurements) by SLAB. This guideline may be used for organizing proficiency Testing Programmes for Medical / Clinical laboratories.

This publication sets out the criteria SLAB uses for the development, operation and selection of proficiency testing programmes formulated by SLAB for its applicant/accredited testing and calibration laboratories.

This publication would regularly be reviewed and the changes may have to be incorporated because of:

a) Any modification/changes of ISO/IEC Guide – 43-1 & ISO/IEC Guide – 43-2

b) The deliberation and decisions of ILAC

c) The deliberation and decisions of APLAC

d) The feed back from accredited laboratories

e) The feed back from assessors/experts

f) The decisions of the Council and Committees of SLAB

2. SCOPE

2.1 The aim of this publication is to set out the criteria for the operation of proficiency testing programmes for laboratories accredited by the SLAB.

2.2 This guide describes principles and requirements for,

a) Development and operation of proficiency testing programmes.

b) Selection and use of proficiency testing programmes by laboratory accreditation bodies.

2.3 This guide also describes how SLAB should select and operate programmes and assist in harmonizing the use of results of proficiency testing schemes.

2.4 Additional criteria may be specified by SLAB depending upon the specific testing/calibration that is to be evaluated.

2.5 Results from proficiency testing programmes may be used in accreditation decisions, and it is important that both SLAB and participating laboratories have confidence in the design and operation of the programmes. It is also important for participating laboratories and laboratory accreditation assessors to have a clear understanding of SLAB policies for participation in such programmes, the criteria they use for judging successful performance in proficiency testing programmes, and their policies and procedures for following up any unsatisfactory results from a proficiency test.

2.6 It should be recognized, however, that SLAB and its assessors may take into account the suitability of test data produced from other activities apart from proficiency testing programmes. This include results of laboratory's own internal quality control procedures with control samples, comparison with split-sample data from other laboratories, performance of audit tests with certified reference materials etc.

2.7 This document was authorized for adoption and use by the Council of the Sri Lanka Accreditation Board for conformity Assessment (SLAB).

3. TYPES OF PROFICIENCY TESTING PROGRAMMES

Proficiency testing techniques vary depending on the nature of the item or material under test, test method in use and the number of testing laboratories participating. Most possess the common feature of comparison of test results obtained by one testing laboratory with those obtained by one or more other testing laboratories. In some programmes, one of the participating laboratories may be a controlling, coordinating, or reference function. The following are the major types of proficiency testing programmes.

3.1 Inter-laboratory Testing

In this programme participating laboratories are provided with sub-samples from a source of a suitable degree of homogeneity which they are expected to test at comparable levels of competence. Features of such programmes usually as follows,

a) It is essential that the (sub) samples provided to each participant are sufficiently homogenous so that any results later identified as extreme should not be attributed to any significant sample variability.

b) After completion of the testing the results are returned to the coordinating body and compared with the assigned values to give an indication of the performance of the individual laboratories and the group as a whole.

c) This is the type commonly used by accrediting bodies, regulatory bodies and other organizations, when they utilize schemes in the testing fields.

d) Examples of test items used in this type of proficiency testing include food, body fluid, soil, water and other environmental materials. In some cases separate portions of previously established (certified) reference materials are circulated.

e) Occasionally this technique is also used for interlaboratory measurement schemes.

3.2 **Split Sample Testing**

This programme involves samples of a product or a material being divided, into two or more parts with each participating laboratory testing one part of each sample.

a) This programme differs from Inter-laboratory proficiency testing as there is usually limited control of, or preliminary data on, the homogeneity of the sample being divided.

b) This technique is sometimes used by clients of laboratory services, including regulatory authorities.

c) Such programmes often need retention of sufficient material to resolve any perceived differences between the limited numbers of laboratories involved by further analysis by additional laboratories.

d) Similar inter comparisons are regularly conducted in commercial transactions when samples representing a traded commodity are split between a laboratory representing the supplier and another laboratory representing the purchaser.

e) A similar technique of split-sample testing is also used in the monitoring of clinical and environmental laboratories. Typically, these schemes involve in the results from several split samples over a wide concentration interval being compared between an individual laboratory and one or more other laboratories. Under such schemes, one of the laboratories may be considered to operate at a higher metrological level (i.e., lower level of uncertainty) due to the use of reference methodology and more advanced equipment, etc. Its results are considered to be the reference values in such inter comparisons and it may act as an advisory or mentor laboratory to the other laboratories comparing split-sample data with it.

3.3 **Measurement Comparisons**

Measurement comparison schemes involve the test item to be measured or calibrated being circulated successively from one participating laboratory to the next. Features of such schemes usually are as follows,

a) Assigned values for the test item are provided by a Reference Laboratory, which might be a country's highest authority for the measurement concerned. It may be necessary for the test item to be checked at specific stages during the conduct of the proficiency test. This is to ensure that there are no significant changes in the assigned value throughout the course of the proficiency test.

b) Schemes involving sequential participation take time (in some cases years) to complete. This causes a number of difficulties such as: ensuring the stability of the item; the strict monitoring of its circulation and the time allowed for measurement by individual participants; and the need to supply feedback on individual performance to laboratories during the scheme's implementation, rather than waiting until it finishes. In addition, it may be difficult to compare results on a group basis as there may be relatively few laboratories whose measurement capabilities closely match each other.

c) The individual measurement results are compared with the reference value established by the Reference Laboratory. The coordinator should take into account the claimed measurement uncertainty of each participating laboratory.

d) Examples of items (measurement artifacts) used in this type of proficiency testing include reference standards (e.g. resistors, gauges and instruments).

3.4 **Qualitative Schemes**

Evaluation of laboratory testing performance will not always involve interlaboratory comparisons. For example, some schemes are designed to evaluate the capabilities of laboratories to characterize specific entities (e.g. type of asbestos, identity of a specific pathogenic organism etc.)

Such schemes may involve the special preparation of test items with addition of the subject component by the scheme coordinator. As such, the schemes are "qualitative" in nature, and do no need the involvement of multiple laboratories or interlaboratory comparisons to evaluate a laboratory's testing performance.

3.5 **Known-Value Schemes**

Other special types of proficiency testing schemes may involve the preparation of test items with known amounts of the measurand under test. It is then possible to evaluate the capability of an individual laboratory to test the item and provide numerical results for comparison with the assigned value. Once again, such proficiency schemes do not need the involvement of multiple laboratories.

3.6 **Partial – process Schemes**

Special types of proficiency testing involve the evaluation of laboratories abilities to perform parts of the overall testing or measurement process. For example, some existing proficiency schemes evaluate laboratories' abilities to transform and report a given set of data (rather than conduct the actual test or measurement) or to take and prepare samples or specimens in accordance with a specification.

4. ORGANIZATION AND DESIGN

4.1 Framework

Proper design for proficiency testing programme is required to ensure its success and smooth operation. The coordinator in consultation with other experts should develop an appropriate plan for the particular proficiency test. The plan should include the following information.

a) The name and address of the organization conducting the proficiency programme.
b) The name and address of the coordinator and other personnel involved in planning and operation.
c) The nature and the purpose of the programme.
d) A procedure for the manner in which the participants are selected.
e) The name and address of the laboratory or laboratories performing the programme, sampling, sample processing and list of potential participants.
f) The nature of the test item(s) and test(s) selected as well as short description of the consideration underlying these choice.
g) A description of the manner in which the test items are obtained, processed, checked and transported.
h) A description of the information that is supplied to participants in this notification phase and of the time schedule for the various phases of the proficiency testing.
i) The expected initial and target dates or deadlines of the proficiency scheme including the date(s) for the testing to be carried out by the participants.
j) The basis of performance evaluation techniques.
k) Information on methods or procedures which participants may need to use to perform the tests or measurements (commonly their routine procedures).
l) A description of the extent to which the test results, and the conclusions that will be based on the outcome of the proficiency tests, are to be made public.
m) The expected initial and target dates or deadlines of the proficiency scheme including the date(s) for the testing to be carried out by the participants.

4.2 Staff

4.2.1 The staff involved in providing the programme organized by SLAB must have adequate qualifications and experience to design implementation and reporting of interlaboratory tests.

4.2.2 There should be an advisory panel for various disciplines of testing and calibration and functions of this panel may include.

a) The development and review of procedures for the planning, execution, analysis and report of the proficiency testing programme.

b) The identification and evaluation of interlaboratory test comparison organized by bodies.

c) Providing advice to SLAB on the use of proficiency testing as an element of its laboratory evaluation.

4.3 Test and Data-Processing equipments

The equipment and facilities must be adequate to conduct the tests. The storage and security of data file is also to be ensured. The use of computer based system is recommended.

4.4 Statistical design

4.4.1 The statistical model and data analysis techniques to be used should be documented, together with a short description of the background to their selection. Further details of common statistical procedures and treatment of proficiency testing data are discussed in Annex-A.

4.4.2 Appropriate statistical design of a proficiency testing scheme is essential. Careful consideration should be given to the following matters and their interaction.

a) The precision and trueness of the test(s) involved.
b) The smallest differences to be detected between participating laboratories at a desired confidence level.
c) The number of participating laboratories.
d) The number of samples to be tested and the number of repeat tests or measurements to be carried out on each sample.
e) The procedures to be used to estimate the assigned value.
f) Procedures to be used to identify outliers.

4.4.3 The absence of reliable information concerning**,** it may be necessary in some cases to organize a pilot interlaboratory comparison (collaborative trial) to obtain it.

4.5 Test Item Preparation

4.5.1 Preparation of test items may either be contracted out or undertaken by the coordinator. In each case the test items shall be homogenous and stable and the test items obtained through outside sources shall be traceable to a reputed organization which demonstrates traceability requirements to SI units.

4.5.2 The test items or materials to be distributed in the programme should generally be similar in type to those routinely tested by the accredited laboratory.

4.5.3 The assigned value should not be disclosed to the participants until after the results have been collated.

4.6 Test Item Management

4.6.1 Procedures for sampling, randomizing, transporting, sorting and handling of test items or materials to be documented.

4.6.2 The bulk material prepared for the proficiency test must be sufficiently homogenous for each test parameter so that all laboratories will receive test samples that do not differ significantly.

4.6.3 Coordinator should consider any hazards that the test materials might pose and appropriate to advise any party that might be at risk of the potential hazard involved.

4.6.4 Where possible the coordinator should also provide evidence that the test items are sufficiently stable to ensure that they will not undergo any significant change throughout the conduct of the proficiency test. When unstable measured need to be assessed, it may be necessary for the coordinating organization to specify a date by which the testing should be completed and any required special pre-testing procedures.

4.7 Choice of Method/Procedure

4.7.1 Participants will normally be able to use the method of their choice, which is consistent with routine procedures used in their laboratories. However, in certain circumstance, the coordinator may instruct participants to use a specified method. Such methods are usually nationally or internationally accepted standard methods, and will have been validated by an appropriate procedure (e.g. collaborative trial).

4.7.2 Where a calibration procedure is used, the assigned value will often be a reference value obtained from measurements obtained by a high-echelon calibration laboratory (often a National Standards Laboratory) which should use a well defined and accepted procedure.

4.7.3 Where participants are free to use a method of their own choice, coordinators should, where appropriate, request details of the methods used to allow the use of participants' results to compare and comment on the methods.

4.8 Evolution of Proficiency Testing Programmes

To ensure that proficiency testing programmes are able to adapt to technical and scientific developments, they may need to include new types of samples or new methods or procedures. Early conclusions from the results of such programmes on the performance of individual laboratories should be drawn with due care.

5. OPERATION AND REPORTING

5.1 Coordination & Documentation

The day-to-day operation of a programme should be the responsibility of the coordinator of the programme. All practices and procedures must be documented. These may be included in, or supplemented by a Quality Manual.

5.2 Instructions

5.2.1 Director/CEO, SLAB would appoint the coordinator from SLAB or any outside agency competent to conduct the proficiency testing.

5.2.2 Detailed instructions covering all aspects of the programmes which must be adhered to by the participating laboratories should be provided by the coordinator.

5.2.3 These will include details concerning factors, which may influence the testing/ calibration of the supplied proficiency testing items or materials. These factors may include operators, nature of items or materials, equipment status, selection of test/calibration procedure and timing of testing/calibration. The materials are to be prepared by an accredited laboratory with a code number.

5.2.4 SLAB should organize proficiency testing in various disciplines. All laboratories holding SLAB accreditation would be required to participate in those accredited tests. At least one proficiency testing in a major field of testing/calibration would be organized by SLAB annually. These rounds of testing may be organized at various locations of the country.

5.2.5 Proper notice (at least 6 weeks) well in advance would be issued by SLAB/ Coordinating laboratory to all probable participating laboratories informing them regarding the proficiency test, name of the coordinator, his address, probable date of supply of the samples and type of proficiency testing to be conducted.

5.2.6 Participating laboratories would be asked to submit the test report within four weeks from the date of the receipt of the sample to the coordinator of the programme.

5.3 Packaging and Transportation

The packaging has to be adequate and able to protect the stability and characteristics of the test items. There may be certain restrictions on transportation such as dangerous goods, regulations or customs requirements. The laboratories themselves must also take responsibility for the transport of the items, particularly in sequential measurement comparison programmes. SLAB/coordinating laboratory should ensure proper packaging of the samples.

Appropriate customs declaration forms wherever applicable should be completed by the coordinator to ensure that those delays in custom clearance are minimized.

5.4 Testing/Calibration

5.4.1 The coordinator shall ensure that the participating laboratories have used the standard process of testing/calibration for which the programme has been fixed.

5.4.2 The laboratory shall be required to provide a list of equipment used in testing/ calibration.

5.4.3 It is the responsibility of the participants to use calibrated equipment, follow good laboratory upkeep, to be diligent in following the method and to treat test specimens/calibration artifacts in the normal laboratory priority. Specimen/ artifacts identification must be maintained during all work.

5.4.4 The participants are responsible for ensuring that test results are submitted within due date prescribed by SLAB. If a participant suspects that the test will not be completed by the deadline, the laboratory should contact the programme coordinator of SLAB for instructions on how to proceed.

5.5 Data Analysis & Records

5.5.1 The data received from the participating laboratories must be entered and analyzed and then reported back as soon as practicable. It is essential that proper procedures are to be adopted to check the validity of data entries and transfers and subsequent statistical analysis. Retention of data capture sheets, computer back-up files or print-outs, graphs etc, are to be kept for a reasonable period of time i.e., at least for 4 years.

5.5.2 Data analysis should generate summary measures and performance statistics and associated information consistent with the scheme's statistical model and the objectives of the scheme. The influence of extreme results on summary statistics should be minimized by the use of outlier detection tests to identify and then omit them or, preferably, by the use of robust statistics. APPENDIX - A contains some broad suggestions for statistical evaluations.

5.5.3 Programme coordinators should have documented criteria for dealing with test results that may be inappropriate for proficiency testing evaluations. For example, it is recommended that for measured for which the test material has been shown not to be sufficiently homogenous or stable for the purposes of a proficiency test, no grading or scoring should be given.

5.6 Programme Reports

5.6.1 The content of programme reports which will vary depending on the purpose of a particular programme should be clear and comprehensive and include data on the distribution of results from laboratories together with an indication of individual participant's performance.

5.6.2 The following information should normally be included in reports of proficiency testing programmes.

a) Name and address of organization providing the programme.

b) Names and affiliations of the persons involved in the design and conduct of the programme.

c) Date of issue of report

d) Report number and clear identification of programme.

e) Clear description of items or materials used including details of sample preparation, homogeneity testing.

f) Laboratory participation codes and test results.

g) Statistical data and summaries including assigned values and range of accepted results.

h) Procedures used to establish any assigned value.

i) Details of the traceability and uncertainty of any assigned value.

j) Assigned values and summary statistics for test methods/procedures used by other participating laboratories (if different methods are used by different laboratories).

k) Comments on laboratory performances by the coordinator and technical advisers.

l) Procedures used to design and implement the programme (which may include reference to a scheme protocol).

m) Procedures used to statistically analyze the data (see APPENDIX–A).

n) Advice where appropriate, on the interpretation of the statistical analysis.

5.6.3 For schemes operated on a regular basis, it may be sufficient to have simpler reports such that many of the recommended elements as in 7.6.2 could be excluded from routine reports, but inclined in periodic summary reports and on request from participants.

5.6.4 Reports should be made available quickly within specified time-tables. Although, ideally, all original data supplied should be reported to participants, it may not be possible to achieve this in some very extensive schemes. Participants should receive a summary sheet of their result from SLAB/ coordinating laboratory. In some programmes such as long period measurement comparison schemes, interim reports should be issued to individual participants.

6. EVALUATION OF PROFICIENCY TEST RESULTS

6.1 Where an evaluation of performance is needed, the coordinator should be responsible for ensuring that the method of evaluation is appropriate to maintain the credibility of the scheme.

6.2 The coordinator with the assistance of the technical advisors, wherever necessary may provide comments on performance with respect to,

a) Overall performance versus prior expectations (taking uncertainties into account)

b) Variation within and between laboratories

c) Variation between methods or procedures, if applicable

d) Possible sources of error and suggestions for improving performances

e) Any other suggestions, recommendations or general comments

f) Conclusions

6.3 It may be necessary to provide individual sheets for participants after a particular programme and these may include updated summaries of performance of individual laboratories over various rounds of an ongoing programme. Such summaries can be further analyzed and trends highlighted, if required.

6.4 A variety of procedures and treatments (statistical methods) are available for proficiency testing programme. Coordinator may choose the most appropriate one for the ongoing programme.

7. COMMUNICATION WITH PARTICIPANTS

7.1 Participants should be provided with a detailed set of information of joining a proficiency testing programme. Subsequent communication with the participants can be by a report. Participants should be advised immediately of any changes in programme design or operation.

7.2 Participants should be able to refer to the coordinator if they consider that assessment of their performance in a proficiency testing is in error.

7.3 Feedback from laboratories are encouraged so that they can also contribute to the development of a programme.

7.4 Participating laboratories should be advised by SLAB to maintain their own records of performance in proficiency testing including the results of investigation of any unsatisfactory results and subsequent corrective actions taken by them.

7.5 For laboratories reporting unsatisfactory results, SLAB should ask to investigate and comment on its performance within three months

a) Wherever necessary, the laboratory should undertake any subsequent proficiency test to confirm that corrective actions taken by the laboratory are effective.

b) If required, SLAB should depute expert to confirm that corrective actions are effective.

7.6 SLAB should ensure that the records of performance of the proficiency testing are maintained by the participating laboratories and are made available to the Assessment team, whenever necessary.

8. CONFIDENTIALITY AND ETHICAL CONSIDERATIONS

8.1Confidentiality

Normally, it is the policy of most programmes to maintain confidentiality of the identity of individual participants. The identity of participants should only be known to the minimum number of people involved in coordinating a programme, and this should extend to any subsequent remedial advice or action applied to a laboratory exhibiting poor performance. In some circumstances, a coordinating body may be required to report poor performance to a particular authority, but participants should be notified of this possibility when agreeing to participate in the programme.

8.2 Collusion and Falsification of Results

Although proficiency testing programmes are intended primarily to help participants to improve the performance, there may be tendency among some participants to provide a false impression of their capabilities. SLAB should ensure that collusion should not take place between laboratories to submit independent data.

8.3 Proficiency testing programme should be designed to ensure that there is as little collusion and falsification as possible.

8.4 All reasonable measures should be taken by the coordinator to prevent collusion.

8.5 Participating laboratories should also try to avoid collusion.

9. SELECTION OF PROFICIENCY TESTING PROGRAMMES

9.1 To assist in the evaluation of competence of laboratories for laboratory accreditation purposes, during assessment and surveillance SLAB uses proficiency testing programmes complying with the guidelines described above.

9.2 For proficiency testing programme operated by SLAB approved coordinating agencies, SLAB should ask for documentary evidence that the programmes comply with the subdivision described above.

9.3 In selecting a proficiency testing programme, the following factors are considered by SLAB.

a) The tests, measurements or calibrations involved should match the types of tests, measurements or calibrations performed by the applicant or accredited laboratories proposed for participation.

b) With the agreement of their accredited laboratories, SLAB should have access to accredited laboratories results, together with details of the programme's design, procedures for establishment of assigned values, instructions to participants, statistical treatment of data and the final report from each selected proficiency test.

c) The frequency at which the programme is run.

d) The suitability of the organizational logistics for the programme, such as timing, location, sample stability considerations, distribution arrangements etc. relevant to the group of accredited laboratories proposed for the scheme.

e) The availability of acceptance criteria for the participating laboratories (i.e, for judging successful performance in the proficiency test).

f) The costs of the selected programmes.

g) The programmes policy on maintaining participants confidentiality;

h) Confidence in the suitability of test materials, measurement artifacts etc. used in the programme for characteristics such as homogeneity, stability and where appropriate, traceability to national or international standards.

Note – Some proficiency testing programmes may offer tests which are not exactly matching with the tests performed by an accredited laboratory (for example, the use of different national standard for the same determination) but it may still be technically justified to include the laboratories in the programme if the treatment of the data allows for consideration of any significant differences in test methodology or other factors.

9.4 The selection of a specific proficiency testing programme by SLAB should be authorized and supervised by suitably qualified personnel of SLAB.

10. POLICIES ON PARTICIPATION IN PROFICIENCY TESTING PROGRAMMES

The policies on participation for providers of proficiency testing schemes provided here in are made available to laboratories and other interested parties and of which Issues pertaining to proficiency testing programmes would be guided by this document.

11. USE OF RESULTS BY SLAB

11.1 The results of proficiency testing programmes would be utilized by SLAB and participating laboratories as stated in this document.

11.2 The results that fall outside the acceptance criteria for a specific programme will be handled by SLAB as per the documented procedures.

11.3 For laboratories reporting unsatisfactory results, SLAB should have following policies to,

a) Have the laboratory investigate and comment on its performance within an agreed time-frame.

b) Where necessary, have the laboratory undertaken any subsequent proficiency test which may be available, to confirm that any corrective actions taken by the laboratory are effective.

c) Where necessary, have on-site evaluation of the laboratory by appropriate technical assessors/experts to confirm that corrective actions are effective.

11.4 SLAB should advise participating laboratories of the possible outcomes of unsatisfactory performance in proficiency testing programme such as,

a) Continuation of accreditation subject to appropriate corrective action.
b) Withdrawal of accreditation for the relevant tests.

11.5 SLAB should have procedures to ensure that the records of performance of laboratories in proficiency testing programmes are maintained (in accreditation files or records) for the participating laboratories and are made available to technical assessors for on-site assessments.

11.6 SLAB should obtain feedback from accredited laboratories of action taken from results of proficiency testing programmes, particularly for unsatisfactory performance.

12 ACTION AND FEEDBACK BY LABORATORIES

12.1 Accredited laboratories should be required to maintain their own records of performance in proficiency testing, including the outcomes of investigations of any unsatisfactory results and any subsequent corrective or preventative actions.

12.2 The laboratories should draw their own conclusions about their performance. The information that should be taken into consideration includes follows,

a) The origin and character of test samples.
b) The test methods used and, where possible, the assignment of the results to particular methods.

APPENDIX -A

Examples of statistical methods for treatment of proficiency test data

Proficiency test results can appear in many forms, spanning a wide range of data types and underlying statistical distributions. The statistical techniques used to analyze the results need to be appropriate for each situation, and so are too varied to specify.

There are, however, three steps common to all proficiency tests, when participants' results are to be evaluated,

a) Determination of the assigned value.
b) Calculation of performance statistics.
c) Evaluation of performance.

and, in some cases,

d) Preliminary determination of test item homogeneity and stability.

This annex gives general criteria for statistical techniques that can be applied as needed to guide specific applications.

With new interlaboratory comparison schemes, agreement initially is often poor due to new questions, new forms, artificial test items, poor agreement of methods, or variable laboratory procedures. Coordinators may have to use robust measures of relative performance (such as percentiles) until agreement is improved. Statistical techniques may need to be refined once interlaboratory agreement has improved and proficiency testing is well established.

This annex does not consider statistical techniques for analytical studies other than for treatment of proficiency test data. Different techniques may be needed to implement the other uses of interlaboratory comparison data listed in the Introduction.

NOTE – ISO/TC 69 is currently preparing a document providing detailed information on statistical methods contained in this annex.

A.1 Determination of the assigned value and its uncertainty

A.1.1 There are various procedures available for the establishment of assigned values. The most common procedures are listed below in an order that, in most cases, will result in increasing uncertainty for the assigned value. These procedures involve use of followings,

a) **Known values** – with results determined by specific test item formulation

(E.g. manufacture or dilution).

b) **Certified reference values** – as determined by definitive methods (for quantitative tests).

c) **Reference values** – as determined by analysis, measurement or comparison of the test item alongside a reference material or standard, traceable to a national or international standard.

d) **Consensus values from expert laboratories**

Expert laboratories should have demonstrable competence in the determination of the measured(s) under test, using validated methods known to be highly precise and accurate, and comparable to methods in general use. The laboratories may, in some situations, be Reference Laboratories.

e) **Consensus values from participant laboratories**

Using statistics described in A.1.3 with consideration of the effects of extreme values.

A.1.2 Assigned values should be determined to evaluate participants fairly, yet to encourage interlaboratory and inter method agreement. This is accomplished through selection of common comparison groups, wherever possible, and the use of common assigned values.

A.1.3 The following statistics may be appropriate when assigned values are determined by consensus techniques,

a) **Qualitative value** – consensus of a predetermined majority percentage (usually expressed on a nominal or ordinal scale).

b) **Quantitative value** – "average" for an appropriate comparison group such as

i) Mean, which may be weighted or transformed (e.g. trimmed or geometric mean)

ii) Median, mode or other robust measure.

A.1.4 Where appropriate, the uncertainty of assigned values should be determined using procedures described in *Guide to the Expression of Uncertainty in Measurement*.

A.1.5 Extreme results are treated as follows

a) When participants' results are used to determine assigned values, techniques should be in place to minimize the influence of extreme results. This can be accomplished with robust statistical methods or by removing outliers prior to calculation (Refer: ISO 5725-2). In larger or routine schemes, it may be possible to have automated outlier's screens.

b) If results are removed as outliers, they should be removed only for calculation of summary statistics. These results should still be evaluated within the proficiency scheme and be given the appropriate performance rating.

A.1.6 Other considerations are as follows.

a) Ideally, if assigned values are determined by reference or participant consensus, the coordinator should have a procedure to establish the trueness of the assigned values and for reviewing the distribution of the data.

b) The coordinator should have criteria for the acceptability of an assigned value in terms of its uncertainty.

A.2 Calculation of performance statistics

A.2.1 Performance on single test items

A.2.1.1 Proficiency test results often need to be transformed into a performance statistic, to aid interpretation and to allow comparison with defined goals. The objective is to measure the deviation from the assigned value in a manner that allows comparison with performance criteria. Techniques may range from no processing required to complex statistical transformations.

A.2.1.2 Performance measures should be meaningful to scheme participants. Therefore, measures should relate to the application needs for the test and be well understood or traditional within a particular field.

A.2.1.3 Variability measures are often used for calculation of performance statistics and in summary reports of proficiency testing schemes. Common examples of such variability measures for an appropriate comparison group include,

a) Standard deviation (SD)

b) Coefficient of variation (CV) or relative standard deviation (RSD)

c) Percentiles, median absolute deviation or other robust measures

A.2.1.4 For qualitative results, no calculation is usually necessary.

Commonly used statistics for quantitative results are listed below in order of increasing degree of transformation of participants' results

a) Difference (x – X), where "x" is the participant's result and "X" is the assigned value.

b) Percent difference

$$\frac{(x - X)}{X} \times 100$$

c) Percentile or rank

d) z scores, where

$$z = \frac{(x - X)}{s}$$

And s is an appropriate estimate/measure of variability which is selected to meet the requirements of the scheme. This model can be used both in the situation where X and s are derived from participants' results or when X and s are not derived from (all) the participant results. [For example, when assigned values and variability are specified; refer to 4.2 of *International Harmonized Protocol for Proficiency Testing of (Chemical) Analytical Laboratories*.]

e) E_n numbers (typically used in measurement comparison schemes), where

$$E_n = \frac{(x - X)}{\sqrt{U^2_{lab} + U^2_{ref}}}$$

and U_{lab} is the uncertainty of a participant's result and U_{ref} is the uncertainty of the reference laboratory's assigned value.

A.2.1.5 Considerations are as follows,

a) The simple difference between the participant's result and the assigned value may be adequate to determine performance, and is most easily understood by participants. The quantity (x – X) is called the "estimate of laboratory bias" in *ISO 5725-4.*

b) The percent difference adjusts for concentration, and is well understood by participants.

c) Percentiles or ranks are useful for highly disperse or skewed, results, ordinal responses, or when there are a limited number of different responses. This technique should be used with caution.

d) Transformed results may be preferred, or necessary, depending on the nature of the test. For example, dilution-based results are a form of geometric scale, transformable by logarithms.

e) If statistical criteria are used (e.g. z scores), the estimates of variability should be reliable; that is; based on enough observations to reduce the influence of extreme results and achieve low uncertainty.

A.2.2 Combined performance scores

A.2.2.1 Performance may be evaluated on the basis of more than one result in a single proficiency test round. This occurs when there is more than one test item for a particular measurand, or a family of related measurands. This would be done to provide a more comprehensive evaluation of performance.

Graphical Methods such as the Youden Plot or a plot showing Mandel's *h*-statistics are effective techniques for interpreting performance (*Refer: ISO 5725-2).*

Examples are as follows.

1) Composite score for the same measurand.

   a. Number of satisfactory results.

   b. Average or summed z score.

   c. Average absolute difference (in units or percent).

   d. Summed absolute difference (or square difference).

2) Composite score for different measurands:

   a. Number (or percent) of satisfactory results.

   b. Average absolute z scores.

   c. Average absolute difference relative to the evaluation limits.

A.2.2.2 Considerations are as follows

a) Scores may be transformed (if necessary) so that they all follow the same assumed distribution (e.g. Gaussian for z scores or chi square for squared differences).

b) There should be a check for extreme values that could heavily influence a quantitative composite score.

A.3 Evaluation of performance

A.3.1 Initial performance

Criteria for performance evaluation should be established after taking into account whether the performance measure involves certain features.

A.3.1.1 These features are the following.

a) **Expert consensus**: where the advisory group, or other qualified experts, directly determine whether reported results are fit for the purpose. Expert consensus is the typical way to assess results for qualitative tests.

b) **Fitness for purpose**: considering, for example, method performance specifications and participants' recognized level of operation.

c) **Statistical determination for scores**: where criteria should be appropriate for each scores. Common examples of application of scores are:

i) for z scores:

$|z| \leq 2$ = satisfactory

$2 < |z| < 3$ = questionable

$|z| \geq 3$ = unsatisfactory

ii) for E_n numbers:

$|E_n| \leq 1$ = satisfactory

$|E_n| > 1$ = unsatisfactory

d) **Consensus of participants**: the range of scores or results used by some percentage of participants, or from a reference group, such as:

- Central percentage (80%, 90% or 95%) satisfactory, or
- One-sided percentage (lowest 90%) satisfactory.

A.3.1.2 For split-sample designs, an objective may be to identify inadequate calibration and/or large random fluctuation in results. In these cases, evaluations should be based on an adequate number of results and across a wide range of concentrations. Graphical techniques are useful for identifying and describing these problems, in particular, graphs showing the differences between the laboratories plotted against the corresponding average values. Results can be compared using regression analysis and analysis of residuals with appropriate parametric or non-parametric techniques.

A.3.1.3 Graphs should be used whenever possible to show performance (e.g. histograms, error bar charts, ordered z score charts). These charts can be used to show:

a) Distributions of participant values;

b) Relationship between results on multiple test items; and

c) Comparative distributions for different methods.

A.3.2 Monitoring performance over time

A.3.2.1 A proficiency test scheme can include techniques to monitor performance over time. The statistical techniques should allow participants to see the variability in their performance; whether there are general trends or inconsistencies, and where the performance varies randomly.

A.3.2.2 Graphical methods should be used to facilitate interpretation by a wider variety of readers. Traditional "Shewhart" control charts are useful, particularly for self-improvement purposes. Data listings and summary statistics allow more detailed review. Statistics used to evaluate performance should be used for these graphs and tables.

A. 4 Preliminary determination of test item homogeneity

Appropriate statistical techniques should be used for the evaluation of data from homogeneity testing of test items. One suitable approach is described in *The International Harmonized Protocol for the Proficiency testing of (Chemical) Analytical Laboratories*. See Appendix II: A Recommended Procedure for Testing Material for Sufficient Homogeneity.

APPENDIX - B

TERMS AND DEFINISIONS

4.1 **Test**

Technical operation that consists of the determination of one or more characteristics of a given product, process or service according to specified procedure.

4.2 **Calibration**

The set of operations which establish, under specified conditions, the relationship between values indicated by a measuring instrument or system or values of a measure.

4.3 **Test Method**

Specified technical procedure for performing a test.

4.4 **Test result**

The value of a characteristic, obtained by completely carrying out a specified measurement method.

4.5 **Test item**

Material or artifact presented to the participating laboratory for the purpose of proficiency testing.

4.6 **Testing/Calibration Laboratory**

A party seeking or holding SLAB accreditation for testing and calibration. This party may be an individual, an organization or a part of an organization.

4.7 **Laboratory accreditation**

A formal recognition that a testing/calibration laboratory is competent to carry out specific test or specific types of tests or measurements.

4.8 **Laboratory assessment**

Examination of a testing/calibration to evaluate its compliance with specific laboratory accreditation criteria.

4.9 **Accreditation body**

A Governmental or non-governmental body which conducts and administers a laboratory accreditation system and grants accreditation.

4.10 **Accredited laboratory** – A laboratory to which accreditation has been granted.

4.11 **Accreditation criteria**

A set of requirements used by an accreditation body which a testing/calibration laboratory must meet to be accredited.

4.12 **Assessor**

An expert who carries out some or all functions related to laboratory assessment.

4.13 **Accredited laboratory test report**

A test report which includes a statement by the testing laboratory that it is accredited for the test reported and that the test has been performed in accordance with the conditions prescribed by the accreditation body.

4.14 **Reference material (RM)**

A material or substance one or more properties of which are sufficiently well established to be used for the calibration of an apparatus, the assessment of the measurement method, or for assigning the values to materials.

4.15 **Certified Reference Material (CRM)**

Reference material (RM), accompanied by a certificate, one or more of which property values are certified by a procedure which establishes its 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.

4.16 **Interlaboratory comparison**

Organization, performance and evaluation of testing/calibration (competence) on the same or similar items by two or more laboratories in accordance with pre-determined conditions.

4.17 **Proficiency Testing**

methods of checking laboratory testing/calibration performances by means of interlaboratory comparison.

4.18 **Reference Laboratory**

Laboratory that provides reference values on a test item with a known uncertainty (usually a National Calibration Laboratory).

4.19 **Accepted Reference Value**

a) A theoretical or established value based on scientific principles

b) An assigned value, based on experimental work of some national or international organization.

c) A consensus value, based on collaborative experimental work under the auspices of a scientific or engineering group.

4.20 **Assigned Value**

Estimate of true value used in the assessment of proficiency.

4.21 **Coordinator**

Person or body which coordinates all the activities associated with a proficiency testing programme.

4.22 **Accuracy**

The closeness of agreement between test/calibration results and the accepted reference value.

4.23 **Trueness**

The closeness of agreement between the average value obtained from large series of test/calibration results and an accepted reference value.

4.24 **Bias**

The difference between the test results and an accepted reference value.

4.25 **Precision**

The closeness of agreement between independent test/calibration results obtained under prescribed conditions.

4.26 **Traceability**

Property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties.

4.27 **Nodal laboratory/body**

The approved laboratory/organization by SLAB which may be responsible for organizing specific type of proficiency testing programmes on testing/calibration.

4.28 **Laboratory Coordinator**

The individual who will be responsible to organize the proficiency testing programme from the approved nodal laboratory/ organization.

4.29 **Outlier**

Member of a set of values which is inconsistent with the other members of that set.

4.30 **Extreme results**

Outliers and other values which are grossly inconsistent with other members of the data set.

4.31 **Robust statistical techniques**

Techniques to minimize the influence that extreme results can have on estimates of the mean and standard deviation.

4.32 **Uncertainty of measurement**

Parameter associated with the results of a measurement that characterizes the dispersion of the values that could reasonably be attributed to the measurand.

APPENDIX - C

References

a) ISO/IEC Guide 43–1: Proficiency testing by inter-laboratory comparisons-Part1: Development and operation of Proficiency testing schemes.

b) ISO/IEC Guide 43 – 2: Proficiency testing by inter-laboratory comparisons- Part 1: Selection and use of Proficiency testing schemes by laboratory accreditation bodies.

c) ILAC Guide 13: Guidelines for the requirements for the competence of providers of proficiency testing schemes.

APPENDIX - D

COMPOSITION OF THE TECHNICAL ADVISORY COMMITTEE

1	Mr.E.G.Somapala Former Government Analyst	Chairman
2	Dr.A.M.Mubarak Director Industrial Technology Institute (ITI)	Member
3	Dr.S.A.Fernando Senior Lecture University of Colombo	Member
4	Prof.W.D.W.Jayathilake Professor of Chemistry University of Sri Jayawardenepura	Member
5	Mr.R.M.G.B.Rajanayake City Analyst Colombo Municipal Council	Member
6	Mr. B S P Mendis Director/ CEO Sri Lanka Accreditation Board for Conformity Assessment (SLAB)	Member
7	Mr. T Wickremasinghe Deputy Director Sri Lanka Accreditation Board for Conformity Assessment (SLAB)	Member