NABL 181

# National Accreditation Board for Testing and Calibration Laboratories (NABL)

# SPECIFIC CRITERIA FOR PT PROVIDER ACCREDITATION

**ISSUE NO.: 03**
**ISSUE DATE: 16-May-2020**

**AMENDMENT NO.: --**
**AMENDMENT DATE: --**

# AMENDMENT SHEET

| S. No. | Page No. | Clause No. | Date of Amendment | Amendment made | Reasons | Signature QA Team | Signature CEO |
|---|---|---|---|---|---|---|---|
| 1 | | | | | | | |
| 2 | | | | | | | |
| 3 | | | | | | | |
| 4 | | | | | | | |
| 5 | | | | | | | |
| 6 | | | | | | | |
| 7 | | | | | | | |
| 8 | | | | | | | |
| 9 | | | | | | | |
| 10 | | | | | | | |

# CONTENTS

## 1. INTRODUCTION

All PT Providers are required to comply with all the requirements listed in the International Standard ISO/IEC 17043: 2010 ‘Conformity assessment – General requirement for proficiency testing’ for all types of proficiency testing schemes. ISO 13528 provides support for the implementation of ISO/IEC 17043 particularly, on the requirements for the statistical design, validation of proficiency test items, review of results, and reporting summary statistics.

Proficiency testing is the evaluation of participant’s performance against pre-established criteria by means of interlaboratory comparisons. Proficiency testing schemes vary according to the needs of the sector in which they are used, the nature of the proficiency test items, the methods in use and the number of participants.

The term “proficiency testing” is taken in its widest sense and includes, but is not limited to:

a) quantitative scheme — where the objective is to quantify one or more measurands of the proficiency test item;

b) qualitative scheme — where the objective is to identify or describe one or more characteristics of the proficiency test item;

c) sequential scheme — where one or more proficiency test items are distributed sequentially for testing or measurement and returned to the proficiency testing provider at intervals;

d) simultaneous scheme — where proficiency test items are distributed for concurrent testing or measurement within a defined time period;

e) single occasion exercise — where proficiency test items are provided on a single occasion;

f) continuous scheme — where proficiency test items are provided at regular intervals;

g) sampling — where samples are taken for subsequent analysis; and

h) data transformation and interpretation — where sets of data or other information are furnished and the information is processed to provide an interpretation (or other outcome).

**Note:**

- Some providers of proficiency testing in the medical area use the term “External Quality Assessment (EQA)” for their proficiency testing schemes, or for their broader programmes, or both.
- Actually “quantitative” and “qualitative” are not schemes in pure terms this is only the way statistics of a scheme will be handled. All schemes i.e. sequential, simultaneous, single occasion, continuous, sampling, data transformation and interpretation can be quantitative or qualitative.

## 2. SCOPE

This document specifies additional requirements for provider of proficiency testing to comply along with ISO/IEC 17043: 2010.

## 3. REFERENCES

**3.1** ISO/ IEC 17043: 2010, Conformity assessment – General requirements for proficiency testing,

**3.2** ISO 13528: 2015 Statistical methods for use in proficiency testing by interlaboratory comparisons

## 4. Specific Policy of NABL for PT Providers

**4.1** The PT Provider shall have conducted at least one PT program in same / similar Matrix (PT item) in a group (e.g. building materials, metals and alloys, food and agricultural products etc.) with at least some of the critical parameters in accordance with ISO/IEC 17043:2010 for the applied scope of accreditation.

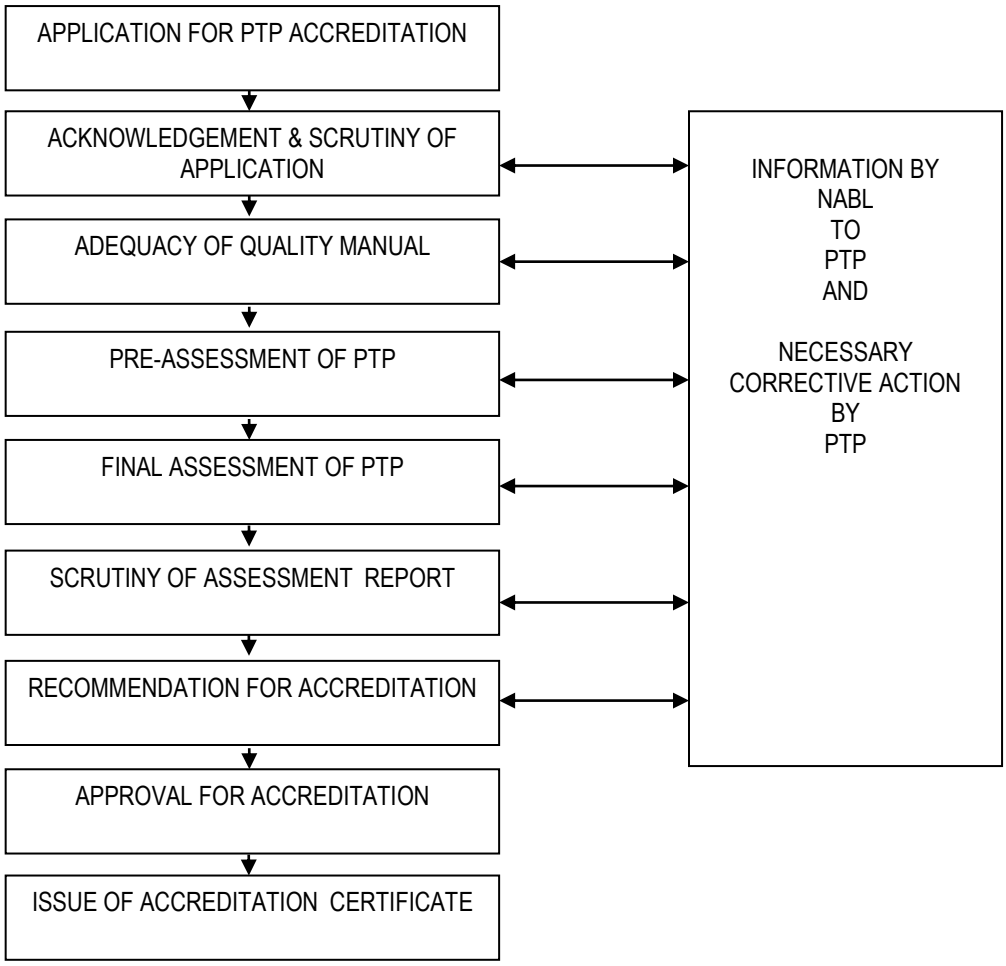
Note – For PT Program not conducted / not completed, at least PT design document (detailed document as how the program will be conducted) shall have been prepared for entire scope of accreditation.

**4.2** The PT Provider shall have completed an Internal Audit and Management review before applying and it shall be conducted after completion of PT scheme.

**4.3** In case of PT programs for calibration laboratory, the Reference standard used for assigning the value of artifact (PT item) shall be calibrated against a standard which is directly traceable to National Metrological Institute (NMI).

Note- NABL documents NABL 100, NABL 131, NABL 132, NABL 133, NABL 134 and NABL 216 are applicable for PT Providers also.

4.2 **PTP Accreditation Process Flow Chart**



## 5. Requirements

### 5.1 Personnel

5.1.1 PT Provider should have personnel who are fully conversant with the product behavior during storage, retention and transport. These personnel should have good knowledge of manufacturing or processing of the product.

5.1.2 PTP should have competent person to review, report and authorize the reports issued by PT provider under NABL accreditation.

5.1.3 For a person to be accepted for this purpose, personal evaluation shall be done during assessment. The Qualification and experience criteria of relevant field (testing / Calibration / Medical) shall apply. In case of medical field, persons are required to be postgraduate in relevant medical field or Ph.D. in relevant scientific fields with work experience of ten years in clinical laboratories out of which 5 years at supervisory level and above. The relevant academic qualifications, experience and demonstration of technical competence to the assessment team shall be the basis for acceptance of person to review, report and authorize the report.

5.1.4 In the event of non availability of personnel within the PT Provider organization, the PT Provider should have access to such expertise (Personnel with comprehensive knowledge of testing / measurement of the product for required parameters in the scope of accreditation) in the Advisory/ Steering Committee formed.

5.1.5 The interview of such personnel (telephonically or directly) during the assessment shall be arranged if required to ensure access of PT Provider to technical experts having knowledge of PT items.

5.1.6 Person having the responsibility of statistical analysis shall be familiar with specific methods used in the evaluation of PT results such as z-score, En ratio or other methods used in the specific PT program and with use of software for the evaluation of PT results.

5.1.7 If technical experts / advisory or steering group are used then following information to be available:

5.1.7.1 CV of individuals,

5.1.7.2 Purpose for which the expert is to be used,

5.1.7.3 Contractual agreement and expert’s concurrence to be part of PT Provider,

5.1.7.4 Confidentiality,

5.1.7.5 Minutes of advisory committee and/or records of communications through e-mail or letter

5.1.8 The proficiency testing provider shall authorize personnel for all the key activities and records shall be available with PTP. **Typical format for maintaining records for personnel is as follows:**

| S. No. | Items / Particular | Name of the Person | Remarks/ Comments |
|---|---|---|---|
| 1. | For Selection of appropriate proficiency test items | | |
| 2. | For Planning of proficiency testing schemes | | |
| 3. | To perform particular types of sampling | | |
| 4. | To operate specific equipment | | |
| 5. | To conduct measurements to determine stability and homogeneity, as well as assigned values and associated uncertainties of the measurands of the proficiency test item | | |
| 6. | To prepare, handle and distribute proficiency test items | | |
| 7. | To operate the data processing system | | |
| 8. | For conducting statistical analysis | | |
| 9. | To evaluate the performance of proficiency testing participants | | |
| 10. | To give opinions and interpretations | | |
| 11. | To authorize the issue of proficiency testing reports | | |

## 5.2 Typical format for Design document:

Typical format for design document to include at least the following information:

| S. No. | Title | Particulars | Comments / Remarks |
|---|---|---|---|
| 1. | Unique identity of the PT Scheme | | |
| 2. | Type of Scheme | | |
| 3. | Objective, purpose and basic design of the PT scheme | | |
| 4. | The name of the Proficiency testing provider. | | |
| 5. | Address of the Proficiency testing provider | | |
| 6. | The name, address and affiliation of the coordinator and other personnel involved in the operation of the proficiency testing scheme. | | |
| 7. | The activities to be subcontracted and the names and addresses of subcontractors to be involved in the operation of the proficiency testing scheme. | | |
| 8. | Criteria to be met for participation. | | |
| 9. | The number and type of expected participants in the proficiency testing scheme | | |
| 10. | Selection of the measurand(s) or characteristic(s) of interest, including information on what the participants are to identify, measure or test for in the specific proficiency testing round | | |
| 11. | A description of the range of values or characteristics, or both, to be expected for the proficiency test items. | | |
| 12. | The potential major sources of errors involved in the area of proficiency testing offered | | |
| 13. | Requirements for the production, quality control, storage and distribution of proficiency test items. | | |
| 14. | Reasonable precautions to prevent collusion between participants or falsification of results, and procedures to be employed if collusion or falsification of results is suspected | | |

| S. No. | Title | Particulars | Comments / Remarks |
|---|---|---|---|
| 15. | A description of the information which is to be supplied to participants and the time schedule for the various phases of the proficiency testing scheme. | | |
| 16. | For continuous proficiency testing schemes, the frequency or dates upon which proficiency test items are to be distributed to participants, the deadlines for the return of the results by participants and, where appropriate, the dates on which testing or measurement is to be carried out by participants. | | |
| 17. | Expected initial and target dates or deadlines of the scheme, | | |
| 18. | Any information on methods or procedures which participants need to use to prepare the test material and perform the tests or measurements. | | |
| 19. | Procedures for the test or measurement methods to be used for the homogeneity and stability testing of proficiency test items and, where applicable, to determine their biological viability. | | |
| 20. | Preparation of any standardized reporting formats to be used by participants | | |
| 21. | A detailed description of the statistical analysis to used | | |
| 22. | The origin, metrological traceability and measurement uncertainty of assigned values | | |
| 23. | Criteria for the evaluation of performance of participants | | |
| 24. | A description of the data, interim reports or information to be returned to participants | | |
| 25. | A description of the extent to which participants results, and the conclusions that will be based on the outcome of the proficiency testing scheme, are to be made public | | |
| 26. | Actions to be taken in the case of lost or damaged proficiency test items. | | |

**Note-**

- The participants should be suitably informed that the records of PT scheme participation shall be accessible to NABL, if so desired by it
- Laboratory (subcontractor/ in-house) involved in assessing homogeneity, stability and / or preparation of PT items in PT scheme shall not be participant in that particular PT scheme.

## 5.3 Homogeneity and Stability

*Extract from ISO 13528: 2015 (E) Cl 6.1, 6.1.1, 6.1.2, Annex B, Annex E.2*

5.3.1 The proficiency testing provider shall ensure that batches of proficiency test items are sufficiently homogeneous and stable for the purposes of the proficiency testing scheme. The provider shall assess homogeneity and stability using criteria that ensure that inhomogeneity and instability of proficiency test items do not adversely affect the evaluation of performance. The assessment of homogeneity and stability should use one or more of the following approaches:

5.3.1.1 experimental studies as described in Section III. D or alternative experimental methods that provide equivalent or greater assurance of homogeneity and stability;

5.3.1.2 experience with the behaviour of closely similar proficiency test items in previous rounds of the proficiency testing scheme, verified as necessary for the current round;

5.3.1.3 assessment of participant data in the current round of the proficiency testing scheme for evidence of consistency with previous rounds, for evidence of change with reporting time or production order, or any unexpected dispersion attributable to inhomogeneity or instability.

Note 1- These approaches can be adopted on a case-by-case basis, using appropriate statistical techniques and technical justification. The approach will often change during the lifetime of a proficiency testing scheme, for example as accumulated experience reduces the initial requirement for experimental study.

Note 2- Relying on experience (as in above) is only reasonable so long as:

i. The process for producing batches of the proficiency test item(s) does not change in any way that may impact homogeneity;

ii. The materials used in production of the proficiency test item(s) do not change in any way that may impact homogeneity;

iii. There is not a “failure” in homogeneity identified via either homogeneity testing or participant responses; and,

iv. The homogeneity requirements for the material are reviewed regularly, taking account of the intended use of the material at the time of the review, to ensure that the homogeneity achieved by the production process remains fit for purpose.

EXAMPLE- If previous rounds of a proficiency testing scheme used proficiency test items that were tested and demonstrated to be sufficiently homogeneous and stable, and with the same participants as in previous rounds, then if an interlaboratory standard deviation in the current round is not greater than the standard deviation in previous rounds, there is evidence of sufficient homogeneity and stability in the current round.

5.3.2 For calibration proficiency testing schemes where the same artefact is used by multiple participants, the proficiency testing provider shall assure stability throughout the round, or have procedures to identify and account for instability through the progression of a round of the proficiency testing scheme. This should include consideration of tendencies for particular proficiency test items and measurands, such as drift. Where appropriate, the assurance of stability should consider the effects of multiple shipments of the same artefact.

5.3.3 All measurands (or properties) should normally be checked for homogeneity and stability. However, where the behaviour of a subset of properties can be shown to provide a good indication of stability and/or homogeneity for all properties reported on in a round, the assessment described in section III A may be limited to that subset of properties. The measurands that are checked should be sensitive to sources of inhomogeneity or instability in the processing of the proficiency test item. Some important cases are:

5.3.3.1 When the measurement is a proportion, a characteristic that is a small proportion can be more difficult to homogenize and so be more sensitive in a homogeneity check;

5.3.3.2 If a proficiency test item is heated during processing, then choose a measurand that is sensitive to uneven heating;

5.3.3.3 If a measured property can be affected by settling, precipitation, or other time-dependent effects during the preparation of proficiency test items, then this property should be checked across filling order.

EXAMPLE- In a proficiency testing scheme for the toxic metal content of soils, measured metal content is primarily affected by moisture content. A check for consistent moisture content may then be considered sufficient to ensure adequate stability of toxic metals.

Note- An example of homogeneity and stability checks is provided in Section III E, using statistical methods recommended in Section III. D.

## A. Statistical Method recommended for Homogeneity and Stability Check

### D.1 General Procedure for a homogeneity check

**D.1.1** To conduct an assessment for homogeneity for a bulk preparation of proficiency test items, follow the procedure given below:

Choose a property (or properties) or measurand(s) to assess with the homogeneity check.

Choose a laboratory to carry out the homogeneity check and a measurement method to use. The method should have a sufficiently small repeatability standard deviation ($s_r$) so that any significant inhomogeneity can be detected. The ratio of the repeatability standard deviation for the method to the standard deviation for proficiency assessment should be less than 0,5, as recommended in the IUPAC Harmonized Protocol (or 1/6 of $\delta_E$). It is recognized that this is not always possible, so in that case the proficiency testing provider should use more replicates.

Prepare and package the proficiency test items for a round of the proficiency testing scheme, ensuring that there are sufficient proficiency test items for the participants in the proficiency testing scheme and for the homogeneity check.

Select a number $g$ of the proficiency test items in their final packaged form using a suitable random selection process, where $g \geq 10$.The number of proficiency test items included in the homogeneity check may be reduced if suitable data are available from previous homogeneity checks on similar proficiency test items prepared by the same procedures.

Prepare $m \geq 2$ test portions from each proficiency test item using techniques appropriate to the proficiency test item to minimize between-test-portion differences.

Taking the $g \times m$ test portions in a random order, obtain a measurement result on each, completing the whole series of measurements under repeatability conditions.

Calculate the general average $\bar{x}$, within-samples standard deviation $s_w$, and between-samples standard deviation $s_s$, as shown in D.3.

**D.1.2** When it is not possible to conduct replicate measurements, for example with destructive tests, then the standard deviation of the results can be used as $s_s$. In this

situation it is important to have a method with a sufficiently low repeatability standard deviation $s_r$.

## D.2 Assessment criteria for a homogeneity check

**D.2.1** The following three checks should be used to assure that the homogeneity test data are valid for analysis:

a) Examine the results for each test portion in order of measurement to look for a trend (or drift) in analysis; if there is an apparent trend, take appropriate corrective action regarding the measurement method, or use caution in the interpretation of the results.

b) Examine the results for proficiency test item averages by production order; if there is a serious trend that causes the proficiency test item to exceed the criterion at D.2.2 or otherwise prevents use of the proficiency test item, then (i) either assign individual values to each proficiency test item; or (ii) discard a subset of proficiency test items significantly affected and retest the remainder for sufficient homogeneity; or (iii) if the trend affects all proficiency test items, follow the provisions at D.2.4.

c) Compare the difference between replicates (or range, if more than 2 replicates) and, if necessary, test for a statistically significant difference between replicates, using Cochran's test (ISO 5725-2). If the difference between replicates is large for any pair, review a technical explanation for the difference and if appropriate, remove the outlying group from the analysis or, if $m$>2 and the high variance is caused by a single outlier, remove the outlying point.

Note- If $m$>2 and a single observation is removed, subsequent calculation of $s_w$ and $s_s$ will need to take the resulting imbalance into account.

**D.2.2** Compare the between-samples standard deviation $s_s$ with the standard deviation for proficiency assessment $\sigma_{pt}$**.** The proficiency test items may be considered to be adequately homogeneous if:

$$s_S \leq 0{,}3\ \sigma_{pt} \qquad \text{(D.1)}$$

Note 1- The justification for the factor of 0,3 is that when this criterion is met the between-samples standard deviation contributes less than 10 % of the variance for evaluation of performance, so the performance evaluation is unlikely to be affected.

Note 2- As an equivalent measure, $s_s$ can be compared to $\delta_E$:

$$s_s \leq 0{,}1\delta_E \qquad \text{(D.2)}$$

**D.2.3** It may be useful to expand the criterion to allow for the actual sampling error and repeatability in the homogeneity check. In these cases, take the following steps:

a) Calculate $\sigma^2_{allow} = (0{,}3\sigma_{pt})^2$

b) Calculate $c = F_1\sigma^2_{allow} + F_2 s^2_w$

Where $s_w$ is the within - sample standard deviation as calculated in section D.3 and $F_1$ and $F_2$ are from standard statistical tables, reproduced in Table D.1, for the number of proficiency test items selected and with each item tested in duplicate

**Table D.1 —Factors $F_1$ and $F_2$ for use in testing for sufficient homogeneity**

| $g$ | 20 | 19 | 18 | 17 | 16 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $F_1$ | 1,59 | 1,60 | 1,62 | 1,64 | 1,67 | 1,69 | 1,72 | 1,75 | 1,79 | 1,83 | 1,88 | 1,94 | 2,01 | 2,10 |
| $F_2$ | 0,57 | 0,59 | 0,62 | 0,64 | 0,68 | 0,71 | 0,75 | 0,80 | 0,86 | 0,93 | 1,01 | 1,11 | 1,25 | 1,43 |

Where $m$>2, $F_2$ in D.2.3 b) and Table D.1 shall be replaced with $F_m = (F_{g-1,\, g(m-1),\, 0{,}95} - 1)/m$ where $F_{g-1,\, g(m-1),\, 0.95-1}$ is the value exceeded with probability 0,05 by a random variable with an $F$-distribution with $g-1$ and $g(m-1)$ degrees of freedom.

Note The two constants in Table D.1 are derived from standard statistical tables as follows:

$F_1 = \chi^2_{0.95(g-1)\,/\,(g-1)}$ where $\chi^2_{0.95(g-1)}$ is the value exceeded with probability 0.05 by a chi-squared random variable with $g-1$ degrees of freedom, and

$F_2 = (F_{0.95(g-1;g)}-1)/2$ where $F_{0.95(g-1;g)}$ is the value exceeded with probability 0.05 by a random variable with an $F$-distribution with $g-1$ and $g$ degrees of freedom.

c) If $s_s > \sqrt{c}$ then there is evidence that the batch of proficiency test items is not sufficiently homogeneous

**D.2.4** When $\sigma_{pt}$ is not known in advance, for example when $\sigma_{pt}$ is the robust standard deviation of participant results, the proficiency testing provider should choose other criteria for determining sufficient homogeneity. Such procedures could include:

a) check for statistically significant differences between proficiency test items using, for example, the Analysis of Variance F test at $\alpha$=0.05;

b) use information from previous rounds of the proficiency testing scheme to estimate $\sigma_{pt}$;

c) use data from a precision experiment (such as a reproducibility standard deviation as described in ISO 5725-2);

d) accept the risk of distributing proficiency test items that are not sufficiently homogeneous, and check the criterion after the consensus $\sigma_{pt}$ has been calculated.

**D.2.5** If the criteria for sufficient homogeneity are not met, the proficiency testing provider shall consider adopting one of the following actions.

a) Include the between-samples standard deviation in the standard deviation for proficiency assessment, by calculating $\sigma'_{pt}$ as in equation (D.3). Note this needs to be described fully to participants.

$$\sigma'_{pt} = \sqrt{\sigma_{pt}^2 + s_s^2} \quad \text{(D.3)}$$

b) Include $s_s$ in the uncertainty of the assigned value and use $z'$ or $\delta_E'$ to assess performance (see Section 9.5 of ISO 13528: 2015);

c) When $\sigma_{pt}$ is the robust standard deviation of participant results, then the inhomogeneity between proficiency test items is included in $\sigma_{pt}$ and so the criterion for acceptability of homogeneity can be relaxed, with caution.

If none of a) to c) apply, discard the proficiency test item and repeat the preparation after correcting the cause of inhomogeneity.

## D.3 Formulae for Homogeneity Check

The estimate of within-sample standard deviation $s_w$ and between-sample variance $s_s$ may be calculated using analysis of variance as shown below. The method shown is for a chosen number $g$ of proficiency test items, measured in replicate $m$ times.

The data from a homogeneity check are represented by $x_{t,k}$ where

$t$ represents the proficiency test item ($t$ = 1,2……..,$g$)

$k$ represents the test portion ($k$ = 1,2….., $m$)

Define the proficiency test item average and variance as:

$$\bar{x_t} = \frac{1}{m}\sum_{k=1}^{m} x_k$$

$$s_t^2 = \frac{1}{m}\sum_{k=1}^{m}(x_k - \bar{x_t})^2 \quad \text{(D.4)}$$

and the estimate of between-test-portion variance as:

$$w_t^2 = \frac{1}{(m-1)}\sum_{k=1}^{m}(x_k - \bar{x_t})^2 \quad \text{(D.5)}$$

Calculate the general average:

$$\bar{\bar{x}} = \frac{1}{g}\sum_{t=1}^{g}\bar{x_t} \quad \text{(D.6)}$$

the estimate of the variance of sample averages:

$$s_x^2 = \frac{1}{(g-1)}\sum_{t=1}^{g}(\bar{x_t} - \bar{\bar{x}})^2 \quad \text{(D.7)}$$

and the within-samples variance:

$$s_w^2 = \frac{1}{g}\sum_{t=1}^{g} s_t^2 \quad \text{(D.8)}$$

Estimate the combined variance of $s_s$ and $s_w$.

$$s_{s,w}^2 = \frac{1}{(g-1)}\sum_{t=1}^{g}(\bar{x_t} - \bar{\bar{x}})^2 + \left(1 - \frac{1}{m}\right)s_w^2 = s_s^2 + s_w^2 \quad \text{(D.9)}$$

Finally, estimate the between-samples variance as

$$s_s^2 = s_{s,w}^2 - s_w^2 = \frac{1}{(g-1)}\sum_{t=1}^{g}(\bar{x_t} - \bar{\bar{x}})^2 - \frac{1}{m}s_w^2 \quad \text{(D.10)}$$

NOTE in the case that ss2< 0, then it is appropriate to use ss=0

For a common design when m is 2, the following formulae can be used.

Define the sample averages as:

$$x_{t,\cdot} = (x_{t,1} + x_{t,2})/2 \quad \text{(D.11)}$$

and the between-test-portion ranges as:

$$w_t = |x_{t,1} - x_{t,2}| \quad \text{(D.12)}$$

Calculate the general average:

$$\bar{x}_{\cdot,\cdot} = \sum \bar{x}_{t,\cdot}/g \quad \text{(D.13)}$$

Estimate the standard deviation of sample averages:

$$s_x = \sqrt{\sum (x_{t,\cdot} - \bar{x}_{\cdot,\cdot})^2/(g-1)} \quad \text{(D.14)}$$

and the within-samples standard deviation:

$$s_W = \sqrt{\sum w_t^2/(2g)} \quad \text{(D.15)}$$

where the summations in formulae D.13, D.14, and D.15 are over samples (t = 1, 2, ..., g).

Finally, estimate the between-samples standard deviation as:

$$s_S = \max\left(0, \sqrt{s_x^2 - (s_W^2/2)}\right) \quad \text{(D.16)}$$

Note 1- The estimate of between-samples variance $s_s^2$ often becomes negative when $s_s$ is relatively smaller than $s_w$. This can be expected when proficiency test items are highly homogeneous. In this case $s_s = 0$.

Note 2- Instead of using ranges, one could use between test portion standard deviations such as

$$s_t = w_t/\sqrt{2}$$

Note 3- An example is provided in Section III. E

## D.4 Procedures for Checking Stability

### D.4.1 General Considerations for checking stability

These clauses give guidance for meeting the stability requirements of section III. The provisions of clause III C with regard to the properties to be studied apply to any experimental check on stability over the duration of the proficiency testing round and on stability during transport.

**D.4.1.1** Where there is reasonable assurance from previous experimental studies, experience, or prior knowledge that instability is unlikely, experimental stability checks may be limited to a check for significant change over the course of the proficiency testing round, carried out during and after the round itself. In other circumstances, studies of transport effects and stability for the typical duration of a proficiency testing round may take the form of planned studies prior to circulation of proficiency test items, either for each round or during early planning and feasibility studies to establish consistent transport and storage conditions. Proficiency testing providers may also check for evidence of instability by checking reported results for a trend with date of measurement.

**D.4.1.2** The following considerations apply to stability checks:

- All properties that are used in the proficiency testing scheme should be checked or otherwise verified for stability. This can be accomplished with previous experience and technical justification based on knowledge of the matrix (or artefact) and measurand.
- More than 2 proficiency test items should be tested if the variability between proficiency test items is large; more proficiency testing items or more replicates should be used if the repeatability is suspect (for example, if $s_w$ or $s_r$>0,5$\sigma_{pt}$ ).

  Note- ISO Guide 35 provides strategies for minimizing the effect on stability studies of long-term variation in the measurement process, such as isochronous studies or the use of stable reference materials.

**D.4.2** Procedure for checking stability during the course of a proficiency testing round.

**D.4.2.1** A convenient model for testing stability in proficiency testing is to test a small sample of proficiency test items at the conclusion of a proficiency testing round and compare these with proficiency test items tested prior to the round, to assure that no change occurred through the time of the round. The check may include a check for any effect of transport conditions by additionally exposing the proficiency test items retained for

the study duration to conditions representing transport conditions. For studies solely intended to check for transport effects, the comparison is between proficiency test items that are shipped with proficiency test items that are retained under controlled conditions.

Note 1- Proficiency testing providers may use the results of homogeneity testing prior to the proficiency testing round instead of selecting and measuring a separate set of proficiency test items.

Note 2- This model applies equally to proficiency testing schemes in testing and in calibration.

**D.4.2.2** If a proficiency testing provider includes shipped proficiency test items in the stability assessment in D.4.2.1, then the effects of transport are included in the assessment of stability. If the effects of transport are checked separately, then the procedure described in section D.6 should be used.

**D.4.2.3** A procedure for a basic stability check using measurements before and after a proficiency testing round is as follows:

a) Select a number $2g$ of the proficiency test items at random, where $g \geq 2$ .

b) Select a single laboratory using a single measurement method with good intermediate precision.

c) Measure $g$ proficiency test items before the planned date of distribution of proficiency test items to participants. Replicated measurements should be made in a fully randomised order.

d) Reserve the remaining $g$ proficiency test items under conditions similar to the expected storage conditions at participants' premises.

e) As soon as reasonably possible after the closing date for return of participant results, measure the remaining $g$ proficiency test items, using the same laboratory, measurement method and number of replicates as at a) above, with all replicates in a randomised order.

f) Calculate the averages $\bar{y}_1$ and $\bar{y}_2$ of the results for the two groups (before and after) respectively.

**D.4.2.4** The following variations to the procedure in D.4.2.3 may be used:

a) The first group of $g$ proficiency test items may be omitted if other measurements on the set of proficiency test items are available from the same laboratory and test method. For example, data from a prior homogeneity check may be used.

b) Conditions likely to accelerate change may be used to provide greater assurance of stability.

c) The second set of proficiency test items may additionally be subjected to conditions expected in shipping, in order to include a test of the effect of shipping.

d) Any other design and conditions that, together with the stability check criterion chosen, provides equal or greater assurance of stability may be used.

**D.5** Assessment criterion for a stability check

**D.5.1** Compare the general average of the measurements obtained in the check prior to distributionwith the general average of the results obtained in the stability check. The proficiency test items may be considered to be adequately stable if:

$$|\bar{y}_1 - \bar{y}_2| \leq 0{,}3\sigma_{pt} \quad \text{or} \quad \leq 0{,}1\delta_E \qquad \text{(D.17)}$$

**D.5.2** If it is likely that the intermediate precision of the measurement method (or the uncertainty of measurement of the item) contributed to the inability to meet the criterion, then one of the following options should be taken:

a) use an isochronous stability study (see ISO Guide 35);

b) increase the uncertainty of the assigned value to account for possible instability;

c) expand the criterion for acceptance by adding the uncertainty of the difference to $\sigma_{pt}$ using the following formula:

$$|\bar{y}_1 - \bar{y}_2| \leq 0{,}3\sigma_{pt} + 2\sqrt{u(\bar{y}_1)^2 + u(\bar{y}_2)^2} \qquad \text{(D.18)}$$

Note- The factor of 2 in equation (B.18) is a coverage factor for the expanded uncertainty of the difference, providing approximately 95% confidence, and the combined uncertainty calculation has intentionally assumed that $\bar{y}_1$ and $\bar{y}_2$ are independent.

**D.5.3** If the criterion in (D.17) or (D.18) is not met, the following options should be considered:

- quantify the effect of instability and take it into account in the evaluation (for example with *z'* scores); or
- examine the proficiency test item preparation and storage procedures to see if improvements are possible; or
- do not evaluate participant performance.

**D.5.4** The criterion at D.5.1 or D.5.2 may be replaced by an appropriate statistical test for a difference between the two sets of data provided that the test takes due account of replication and provides assurance of identifying stability at least equal to that provided by equation (D.18).

Note- A *t*-test for significant difference at the 95% level of confidence, using the means for each proficiency test item, will usually provide similar or better assurance of detecting instability to (D.18) provided that the number of units tested is 3 or more.

**D.6** Stability in transport conditions

**D.6.1** The proficiency testing provider should check the effects of transport on proficiency testing items at least in the early stages of the proficiency testing scheme. Such a check should, where possible, compare proficiency test items retained at the proficiency testing provider's premises with proficiency test items subjected to shipping and return. Studies based on exposure to reasonably foreseeable conditions of transport, for example, may also be used.

**D.6.2** Any known effects of transportation should be considered when evaluating performance. Any significant increase in uncertainty due to transport should be included in the uncertainty of the assigned value.

**D.6.3** Where the transport stability check involves the comparison of results for two groups of proficiency test items, one group being exposed to transport conditions and one group that is not, the criterion for sufficient stability in transport is the same as in section D.5.1 or D.5.2.

Note 1- If the assigned value and standard deviation for proficiency assessment are determined from participant results (e.g., by robust methods), then the average and the standard deviation for proficiency assessment will reflect any bias and increased variability (respectively) caused by transport conditions.

Note 2- An example of a stability check is shown in Section III. E

**B.** An Example for Homogeneity and Stability check

**E.1** Homogeneity and Stability test – Arsenic (As) in chocolate (Example E2 of ISO 13528: 2015)

Proficiency test items are prepared for use in an international proficiency test, and then for subsequent use as reference materials. 1000 vials are manufactured.

Homogeneity check: 10 proficiency test items are selected using a stratified random selection of proficiency test items from different portions of the manufacture process. 2 test portions are extracted from each bottle and tested in a random order, under repeatability conditions. The data are given in Table E.1a below. The procedure in Section III. D.3 is followed, resulting in the summary statistics listed. The fitness-for purpose $\sigma_{pt}$ for As in chocolate is 15%, so the estimate of sample variability is checked against 0,3 times the $\sigma_{pt}$.

**Table E.1a—Homogeneity data for proficiency test items of arsenic in chocolate**

| Bottle ID | Replicate 1 | Replicate 2 |
|---|---|---|
| 3 | 0,185 | 0,194 |
| 111 | 0,187 | 0,189 |
| 201 | 0,182 | 0,186 |
| 330 | 0,188 | 0,196 |
| 405 | 0,191 | 0,181 |
| 481 | 0,188 | 0,180 |
| 599 | 0,187 | 0,196 |
| 704 | 0,177 | 0,186 |
| 766 | 0,179 | 0,187 |
| 858 | 0,188 | 0,196 |

Overall average: 0,18715
SD of averages: 0,00398
$s_w$: 0,00556
$s_s$: 0,00060
$\sigma_{pt}$ =0,18715 x0,15 = 0,02807
Check value: 0,3 x $\sigma_{pt}$ = 0,00842
$s_s$is less than the check value, so homogeneity is sufficient.

Stability check: 2 proficiency test items are randomly selected and stored at an elevated temperature (60°C) for the duration of the round of the proficiency testing scheme (6 weeks). The proficiency test items were tested in duplicate, and the four results are checked against the homogeneity values.

**Table E.2b—Stability data for proficiency test items for arsenic in chocolate**

| Stability sample | Replicate 1 | Replicate 2 |
|---|---|---|
| 164 | 0,191 | 0,198 |
| 732 | 0,190 | 0,196 |

Overall average = 0,19375

Difference from Homogeneity mean: 0,19375 - 0,18715 = 0,00660

Check value: $0{,}3 \times \sigma_{pt}$ = 0,00842

Difference is less than the check value, so stability is sufficient.

## 5.4 Statistical Design

*(Extract from ISO 13528: 2015 (E) Cl 5.2, 5.2.1, 5.2.2, 5.2.3)*

### 5.4.1 Basis of a statistical design

5.4.1.1 According to ISO/IEC 17043, clause 4.4.4.1, the statistical design "shall be developed to meet the objectives of the proficiency testing scheme, based on the nature of the data (quantitative or qualitative including ordinal and categorical), statistical assumptions, the nature of errors, and the expected number of results". Therefore, proficiency testing schemes with different objectives and with different sources of error could have different designs.

Design considerations for common objectives are listed below. Other objectives are possible.

EXAMPLE- 1 For a proficiency testing scheme to compare a participant's result against a pre-determined reference value and within limits that are specified before the round begins, the design will require a method for obtaining an externally defined reference value, a method of setting limits, and a scoring method;

EXAMPLE -2 For a proficiency testing scheme to compare a participant's result with combined results from a group in the same round, and limits that are specified before the round begins, the design will need to consider how the assigned value will be determined from the combined results as well as methods for setting limits and scoring;

EXAMPLE- 3 For a proficiency testing scheme to compare a participant's result with combined results from a group in the same round, and limits determined by the variability of participant results, the design will need to consider the calculation of an assigned value and an appropriate measure of dispersion as well as the method of scoring;

EXAMPLE- 4 For a proficiency testing scheme to compare a participant's result with the assigned value, using the participant's own measurement uncertainty, the design will need to consider how the assigned value and its uncertainty are to be obtained and how participant measurement uncertainties are to be used in scoring.

EXAMPLE -5 For a proficiency testing scheme with an objective to compare the performance of different measurement methods, the design will need to consider the relevant summary statistics and procedures to calculate them.

5.4.1.2 There are various types of data used in proficiency testing, including quantitative, nominal (categorical), and ordinal. Among the quantitative variables, some results might be on an interval scale; or a relative, or ratio scale. For some measurements on a quantitative scale, only a discrete and discontinuous set of values can be realized (for example, sequential dilutions); however, in many cases these results can be treated by techniques that are applicable to continuous quantitative variables.

Note 1- For quantitative values, an interval scale is a scale on which intervals (differences) are meaningful but ratios are not, such as the Celsius temperature scale. A ratio scale is a scale on which intervals and ratios are both meaningful, such as the Kelvin temperature scale, or most common units for length.

Note 2- For qualitative values, a categorical scale has distinct values for which ordering is not meaningful, such as the names of bacterial species. Values on an ordinal scale have a meaningful ordering but differences are not meaningful; for example, a scale such as ‘large, medium, small’ can be ordered but the differences between values are undefined other than in terms of the number of intervening values.

5.4.1.3 Proficiency testing schemes may be used for other purposes in addition to the above, as discussed in introduction of ISO/IEC 17043. The design shall be appropriate for all the stated purposes for the particular proficiency testing scheme.

**5.5 Assigned Value**

PT Provider shall ascertain the validity of assigned value for each round of PT scheme for each parameter.

*(Extract form ISO 13528: 2015 (E) Cl 7.8, 7.8.1, 7.8.2, 7.8.3)*

**A.** Comparison of the assigned value with an independent reference value

**A.1** When the Consensus value from the participants results is used to establish the assigned value ($x_{pt}$), and where a reliable independent estimate (deNoted $x_{ref}$) is available, for example from knowledge of preparation or from a reference value, the consensus value $x_{pt}$ should be compared with $x_{ref}$.

When the Formulation / Certified reference material / Results from one Laboratory / consensus value form expert laboratories methods are used to establish the assigned value, the robust average $x^*$ derived from the results of the round should be compared with the assigned value after each round of a proficiency testing scheme.

The difference is calculated as $x_{diff} = (x_{ref} - x_{pt})$ (or $(x^* - x_{pt})$) and the standard uncertainty of the difference is estimated as:

$$u_{diff} = \sqrt{u^2(x_{ref}) + u^2(x_{pt})}$$

where

$u(x_{ref})$ is the uncertainty of the reference value for comparison; and

$u(x_{pt})$is the uncertainty of the assigned value.

Note- *An example of a comparison of a reference value with a consensus value is included in Section V. B*

**A.2** If the difference is more than twice its standard uncertainty, the reason should be investigated. Possible reasons are:

- bias in the reference measurement method;
- a common bias in the results of the participants;
- failure to appreciate the limitations of the measurement method when using the formulation method for assigning value;
- bias in the results of the “experts” when using Results from one laboratory or consensus value from expert laboratories approaches for assigning value; and
- the comparison value and assigned value are not traceable to the same metrological reference.

**A.3** Depending on the reason for the difference, the proficiency testing provider should decide whether to evaluate results or not, and (for continuous proficiency testing schemes), whether to amend the design for subsequent proficiency testing schemes. Where the difference is sufficiently large to affect performance assessment or to suggest important bias in the measurement methods used by participants, the difference should be Noted in the report for the round. In such cases, the difference should be considered in the design of future proficiency testing schemes.

**B.** An example of a comparison of a reference value with a consensus value

As a demonstration of the procedure to compare a reference value with the robust mean of participant results, following is considered:

In a round of a proficiency testing scheme the robust mean $x^*$ is 0,03161 and the robust standard deviation $s^*$ is 0,0164, calculated with Algorithm A, after removal of blunder and outliers values (n=24). Therefore, the uncertainty of the robust mean is calculated as

$u(x^*) = 1{,}25(s^*/\sqrt{n})$

$u(x^*) = 1{,}25(0{,}0164/\sqrt{24})$ = 0,0042

$x_{ref}$ = 0,044 mg/kg;

$U(x_{ref})$ = 0,0041mg/kg

$u_{diff} = \sqrt{u(x_{ref})^2 + u(x^*)^2} = \sqrt{0{,}0041^2 + 0{,}0042^2}$ = 0,0059

$U_{diff}$ = 2(0,0059) = 0,012

$x_{diff} = x_{ref} - x^*$ = 0,044 – 0,032 = 0,012
so the difference is two times the uncertainty of the difference.

No action is recommended, since the bias in some methods is understood.

## 5.6 Operation of Proficiency Testing Schemes

These instructions could be included along with the one mentioned under section 4.6.1.2 of ISO/IEC 17043: 2010:

i. Test/measurement/sampling/inspection method to be adopted: If the participants are permitted to use a method of their choice, the same should be explicitly mentioned.

ii. Format for recording/reporting of the PT results including unit of measurement and number of significant figures (number of decimals) to be reported.

iii. Last date for the provider to receive the proficiency testing results.

The proficiency test items shall not be stored along with other items which can affect or contaminate them. They should be stored in an appropriately segregated area with controlled access throughout the conduct of the PT Scheme.

Extreme care shall be taken by the PT provider while packing, labeling and transporting the proficiency test items to the participants. The packing materials used such as duralumin boxes, wooden boxes, card board, polystyrene, polyethylene bags, shall be such that the integrity of the proficiency test items is not affected during storage and transport.

Labels containing the unique identity of the individual proficiency test items shall be firmly/securely attached to the packaging. Further, it shall be ensured that the labels are legible and intact till the completion of the PT Scheme.

If the proficiency test item is to be returned to PT Provider or forwarded to another participant, then detailed instructions for packing and transporting the same shall be provided.

Confirmation of the receipt of the proficiency test item is to be obtained promptly from the participants on receipt of the proficiency test items.

## 5.7 Reports

The proficiency test reports shall incorporate results reported by all the participants and their performance in the PT Scheme. However, the identity of the participants shall be codified. In case performance evaluation is reported in terms of En value/Zeta score, PT reports shall also incorporate uncertainty reported by participant.

**5.8 Sampling Program**

PT item (case study) shall be based on laid down national/international stand and or guidelines on sampling or procedures given by regulatory bodies.

**5.9 Qualitative PT Schemes**

*(Extract form ISO 13528: 2015 (E) Cl 11, 11.1, 11.2, 11.3, 11.4, Annexure E.15)*

Design and analysis of qualitative proficiency testing schemes (including nominal and ordinal properties

**A.** Types of qualitative data

A large amount of proficiency testing occurs for properties that are measured or identified on qualitative scales. This includes the following:

- Proficiency testing schemes that require reporting on a categorical scale (sometimes called “nominal”), where the property value has no magnitude (such as a type of substance or organism);

- Proficiency testing schemes for presence or absence of a property, whether determined by subjective criteria or by the magnitude of a signal from a measurement procedure. This can be regarded as a special case of a categorical or ordinal scale, with only two values (also called ‘dichotomous’, or binary);

- Proficiency testing schemes requiring results reported on an ordinal scale, which can be ordered according to magnitude but for which no arithmetic relationships exist among different results. For example, ‘high, medium and low’ form an ordinal scale.

Such proficiency testing schemes require special consideration for the design, value assignment and performance evaluation (scoring) stages because:

- assigned values are very often based on expert opinion; and

- statistical treatment designed for continuous-valued and count data is not applicable to qualitative data. For example, it is not meaningful to take means and standard deviations of ordinal scale results even when they can be placed in a ranking order.

The following paragraphs accordingly provide guidance on design, value assignment and performance evaluation for qualitative proficiency testing schemes.

Note- Guidance for ordinal data does not apply to measurement results that are based on a quantitative scale with discontinuous indications (such as dilutions or titres).

**B.** Statistical design

**B.1** For proficiency testing schemes in which expert opinion is essential either for value assignment or for assessment of participant reports, it will normally be necessary to assemble a panel of appropriately qualified experts and to provide time for debate in order to achieve consensus on appropriate assignment. Where there is a need to rely on individual experts for scoring or value assignment the proficiency testing provider should additionally provide for assessment and control of the consistency of opinion among different experts.

EXAMPLE In a clinical proficiency testing scheme that relies on microscopy for diagnosis, expert opinion is used to assess microscope slides provided to participants and provide an appropriate clinical diagnosis for proficiency test items. The proficiency testing provider may choose to circulate proficiency test items 'blind' to different members of the expert panel to assure consistency of diagnosis, or carry out periodic exercises to evaluate agreement among the panel.

**B.2** For proficiency testing schemes that report simple, single-valued categorical or ordinal results, the proficiency testing provider should consider

- providing two or more proficiency test items per round; or
- requesting the results of a number of replicated observations on each proficiency test item, with the number of replicates specified in advance.

Either of these strategies permits counts of results for each participant that can be used either in reviewing data or in scoring. Provision of two or more proficiency test items may provide additional information on the nature of errors and also allow more sophisticated scoring of proficiency testing performance.

EXAMPLE1 In a proficiency testing scheme intended to report the presence or absence of a contaminant, provision of proficiency test items containing a range of levels of the contaminant allows the proficiency testing provider to examine the number of successful detections at each level as a function of the level of contaminant present. This may be used, for example, to provide information to participants on the detection capability of their chosen test method, or to obtain an average probability of detection which may in turn permit performance scores to be allocated to participants on the basis of estimated probabilities of particular patterns of response.

EXAMPLE 2 Proficiency testing in forensic comparisons often requires matching proficiency test items as to whether they came from the same source or different sources (for example, fingerprints, DNA, bullet shell casings, footprints, etc.). In many cases “indeterminate” is an allowed response. A proficiency testing scheme might include multiple proficiency test items from different sources, and participants are asked to state which are from “same source”, “different source”, or “indeterminate” for every pair. This allows objective scores of number (or %) correct or incorrect, or number (%) correct matches, or correct rejections. Performance criteria can then be determined on fitness for use, or on degree of difficulty of the challenge.

**B.3** Homogeneity should be demonstrated with review of an appropriate sample of proficiency test items, all of which should demonstrate the expected property value. For some qualitative properties, for example presence or absence, it may be possible to verify homogeneity with quantitative measurements; for example a microbiological count or a spectrum absorbance above a threshold. In these situations a conventional test of homogeneity may be appropriate, or a demonstration of all results being above or below a cut-off value.

**C.** Assigned values for qualitative proficiency testing schemes

**C.1** Values may be assigned to proficiency test items:

a) by expert judgement;

b) by use of reference materials as proficiency test items;

c) from knowledge of the origin or preparation of the proficiency test item(s);

d) using the mode or median of participant results (the median is appropriate only for ordinal values).

Any other value assignment method that can be shown to provide reliable results may also be used. The following paragraphs consider each of the above strategies.

Note- It is not usually appropriate to provide quantitative information regarding the uncertainty of the assigned value in qualitative proficiency testing schemes. Each of the paragraphs C.2 to C.5 nonetheless requires the provision of basic information relating to confidence in the assigned value so that participants may judge whether a poor result might reasonably be attributable to an error in value assignment.

**C.2** Values assigned by expert opinion should normally be based on a consensus of a panel of suitably qualified experts. Any significant disagreement among the panel should be recorded in the report for the round. If the panel cannot reach a consensus for a particular proficiency test item, the proficiency testing provider may consider an alternative method of value assignment from those listed in clause C.1. If that is not appropriate the proficiency test item should not be used for performance assessment of participants.

NOTE In some cases it is possible for a single expert to determine the assigned value.

**C.3** Where a reference material is provided to participants as a proficiency test item, the associated reference value, or certified value, should normally be used as the assigned value for the round. Any summary information provided with the reference material that relates to confidence in the assigned value should be available to participants following the round.

**C.4** Where the proficiency test items are prepared from a known source, the assigned value may be determined based on the origin of the material. The proficiency testing provider should retain records of the origin, transport and handling of the material(s) used. Due care must be taken to prevent contamination that might result in incorrect results from participants. Evidence of origin and/or detail of preparation should be available to participants after the round either on request or as part of the report for the proficiency testing round.

EXAMPLE- Proficiency test items of wine circulated for an authenticity proficiency testing scheme may be procured directly from a suitable producer in the designated region of origin, or via a commercial supplier able to provide assurance of authenticity.

**C.4.1** Confirmatory tests or measurements are recommended where possible, especially where contamination may compromise use as a proficiency test item. For example, a proficiency test item identified as an exemplar of a single microbial, plant or animal species should normally be tested for response to tests for other relevant species. Such tests should be as sensitive as possible to ensure that contaminating species are either absent or that the level of contamination is quantified.

**C.4.2** The proficiency testing provider should provide information on any contamination detected or doubts about origin that may compromise use of the proficiency test item.

**C.5** The mode (the most common observation) may be used as the assigned value for results on a categorical or ordinal scale, while the median may be used as the assigned value for results on an ordinal scale. Where these statistics are used, the report for the proficiency testing round should include a statement of the proportion of the results used in value assignment that matched the assigned value. It is never appropriate to calculate means or standard deviations for proficiency testing results for qualitative properties, including ordinal values. This is because there is no arithmetic relationship between different values on each scale.

**C.6** When assigned values are based on measurements (for example, presence or absence), the assigned value can usually be determined definitively; i.e., with low uncertainty. Statistical calculations for uncertainty may be appropriate for levels of measurand in “indeterminate” or “equivocal” levels.

**D.** Performance evaluation and scoring for qualitative proficiency testing schemes

**D.1** Evaluation of participant performance in a qualitative proficiency testing scheme depends in part on the nature of the report required. In some proficiency testing schemes, where a significant amount of evaluation is required of participants and the conclusions require careful consideration and wording, participant reports may be passed to experts for appraisal and may be given an overall mark. At the other extreme, participants may be judged solely on whether their result coincides exactly with the assigned value for the relevant proficiency test item. The following paragraphs accordingly provide guidance on performance assessment and scoring for a range of circumstances.

**D.2** Expert appraisal of participant reports requires one or more individual experts to review each participant report for each proficiency test item and allocate a performance mark or score. In such a proficiency testing scheme, the proficiency testing provider should ensure that:

- the particular participant is not known to the expert. In particular, the report passed to the expert(s) should not include any information that could reasonably identify the participant;
- review, marking and performance assessment follow a set of previously agreed criteria that are as objective as reasonably possible;
- the provisions of paragraph C2 with respect to consistency among experts are met;
- where possible, provision is made for participant appeal against a particular expert opinion and/or for secondary review of opinions close to any important performance threshold.

**D.3** Two systems may be used for scoring a single reported qualitative result based on an assigned value:

I. Each result is marked as acceptable (or scored as a success) if it exactly matches the assigned value and is marked as unacceptable, or given an adverse performance score, otherwise.

EXAMPLE: In a scheme for determining the presence or absence of a contaminant, correct results are scored as 1 and incorrect results as 0.

II. Results that exactly match the assigned value are marked as acceptable and given a corresponding score; results that do not exactly match the assigned value are given a score that depends on the nature of the mismatch. Such scoring designs should assign lower scores to better performance, to be consistent with other types of performance scores (for example, z score, $P_A$ score, $\zeta$, and $E_n$).

EXAMPLE 1- In a clinical pathology proficiency testing scheme, a proficiency testing provider assigns a score of '0' for an exactly correct identification of a microbiological species, '1' point for a result that is incorrect but would not change clinical treatment (for example identification as a different but related microbiological species requiring similar treatment), and 3 points for an identification that is incorrect and would lead to

incorrect treatment of a patient. This scoring scheme will usually require expert judgement on the nature of the mismatch, perhaps obtained prior to scoring.

EXAMPLE 2- In a proficiency testing scheme for which six possible responses ranked on an ordinal scale are possible, a result matching the assigned value is given a score of 0and the score is increased by 2 for each difference in rank until the score increases to a maximum of 6(so a result adjacent to the assigned value would attract a score of 2).

Individual performance scores for each proficiency test item should be provided to participants. Where replicate observations are performed a summary of performance scores for each result may be provided.

**D.4** Where multiple replicates are reported for each proficiency test item or where multiple proficiency test items are provided to each participant, the proficiency testing provider may calculate and use combined performance scores or score summaries in performance assessment. Combined performance scores or summaries may be calculated as, for example:

- the simple sum of performance scores across all proficiency test items;
- the count of each level of performance allocated;
- the proportion of correct results;
- a distance metric based on the differences between results and assigned values.

EXAMPLE- A very general distance metric sometimes used statistics for qualitative data is the Gower coefficient*. This can combine quantitative and qualitative variables based on a combination of scores for similarity. For categorical or binary data the index allocates a score of 1 for exactly matching categories and 0 otherwise; for ordinal scales it allocates a score equal to 1 minus the difference in rank divided by the number of ranks available, and for interval or ratio scale data it allocates a score equal to 1 minus the absolute difference divided by the observed range of all values. These scores, which are all necessarily from 0 to 1, are summed and the sum divided by the number of variables used. A weighted variant may also be used.

Combined performance scores may be associated with a summary performance assessment. For example, particular (usually high) proportion of correct scores may be deemed ‘acceptable’ performance, if that is consistent with the objectives of the proficiency testing scheme.

**D.5** Graphical methods may be used to provide performance information to participants or to provide summary information in a report for a round.

Note- *An example of the analysis of ordinal data is provided in given below in Section E.*

**E.** Qualitative Data Analysis; example of an ordinal quantity: skin reaction to a cosmetic

A proficiency testing scheme involves the analysis of the reaction to a skin care product, when applied to a standard animal subject. Any inflammatory reaction is graded according to the following scale:

1. no reaction
2. moderate redness
3. significant irritation or swelling
4. severe reaction, including suppuration or bleeding

Two proficiency test items consisting of two different products are distributed, labelled product A and product B, and there are 50 participants for each product. The participant results are listed in Table E.15 and shown graphically in Figure E.15. The mode and median are listed for the participant results for each proficiency test item.

**Table E.15—Results for two proficiency test items, skin irritation**

| Reaction | Product A | Product B |
|---|---|---|
| 1 | 20 (40%) # | 8 (16%) |
| 2 | 18 (36%) @ | 12 (24%) |
| 3 | 10 (20%) | 20 (40%) # @ |
| 4 | 2 (4%) | 10 (20%) |

**# - mode,** @ - median

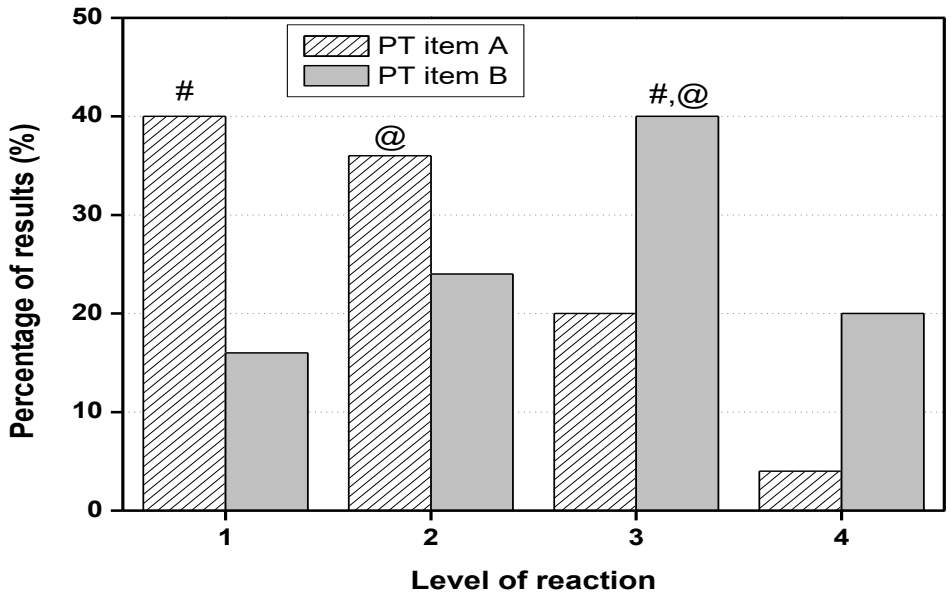



**Figure E.15— Bar chart of percentage responses to two skin irritation proficiency test items**

Note that the median or mode may be used as summary statistics for these proficiency test items, and they suggest that the level of reaction to product B was more severe than the reaction to product A. The proficiency test provider may determine that "action signals" would occur for any result that is more than one ordinal unit away from the median, in which case for product A, action signals occur for the 2 results (4 %) of "4" and for product B, action signals occur for the 8 results (16 %) of "1".

## 5.10 Communication with Participants and Confidentiality

The records of communication with the participants shall be maintained for at least two years (one full accreditation cycle), with controlled access to authorized personnel. PTP to ensure that staff has signed declaration for maintaining confidentiality.

## 5.11 Organization

PT Provider shall establish legal identity mentioning the date of incorporation.
The legal identity of PT Provider may be in either of the following forms:

| Type of Legal Identity | Document(s) to be submitted |
|---|---|
| Proprietorship | Bank passbook/ Account statement and PAN of the laboratory |
| Limited Liability Partnership | Registration certificate under The Limited Liability Partnership Act, 2008 |
| Company | Registration certificate under The Companies Act, 1956 or 2013 |
| Societies/ Trust | Registration under Societies Registration Act, 1860/ Registration under The Indian Trusts Act, 1882 |

| | |
|---|---|
| Government | Gazette or Government Notification or self-Declaration on Letter head by Head of the organization |

If the PT provider is part of an organization which has laboratory/ inspection body, the roles of key personnel shall be clearly defined identifying any potential conflict of interest. The PT provider shall have declaration of independence in case of identified potential conflict of interest. In addition, the organization chart shall clearly define the position and relationship of PT provider with other activities. PTP shall maintain Declaration of Independence in case of identified Potential Conflict of Interest.

The person responsible for quality management system / quality manager should be familiar with and fully aware of the requirements of ISO/IEC 17043:2010, and principles applicable to the organisation's field of accreditation / compliance.

Collusion may take place either amongst the participants or between the PT provider and participants. Few suggested means that will provide guidance to the PT provider and assessor to prevent collusion and falsification are as follows:

a. PT provider should provide adequate justification for the method adopted for assigning value to PT items.

b. The size/amount of PT item sent to participants should not be too large.

c. The design of the PT scheme should ensure that the time gap between receipt of the sample, date of testing and date of reporting should be minimum. Adequate justification should be provided in case of longer time gap.

d. When a PT scheme requires testing on a particular date and time, PT provider should obtain the results from the participants within the specified time by fax/ e-mail initially, followed by courier also.

e. PT provider shall identify, analyze and document risks to impartiality, integrity and objectivity and also collusion and falsification. The threats to the above could arise from following but not limited to:

i. PT provider being part of a testing/ calibration laboratory (common management; same personnel; coercion openly or secretively, etc.)

ii. PT provider may be related to another body which is a laboratory (e.g. under same ownership).

iii. Advisory/ Steering committee members may be from participant laboratories.

iv. Participation from competitor laboratories in PT program.

v. Collusion among participants such as discussion among participants, testing/calibration of PT item in another laboratory (better equipped, specialized).

vi. Request from participants for duplicate PT items with false justification.

f. PT provider shall document and implement a procedure for identifying risks and mitigation measures, where applicable.

g. When a relationship poses unacceptable threat to impartiality, PT services shall not be provided to the related participant.

h. It should be ensured that PT items are prepared, handled and distributed by authorized person / agencies. The access to the records of codification and distribution of PT items shall be controlled.

i. If there is provision for witnessing on line testing, this should be resorted to. In this case the different participant laboratories may be asked to carry out testing on different days at different times to enable the PT Provider to monitor testing activities.

j. Similarly, in case studies as for sampling, PT provider may videograph the conduct of case study if practically possible or get witnessed the case study from independent experts in its advisory group.

k. The case studies in sampling should be well documented, approved by independent experts and objective evaluation should be possible.

**5.12 Subcontracting Services**

Subcontractors and the activity rendered by them shall also be identified and records of assessment of subcontractors' activity should be retrievable. PTP shall have written contract/Agreement with the subcontractor along with the activity being subcontract. Records for verifying the competence (e.g. copy of accreditation certificate, audit reports) shall also be available with PTP.

The assessment shall ensure compliance to the relevant requirements of ISO/IEC 17043 and other related standards like ISO/IEC 17025, ISO 17034, ISO 15189 and ISO/IEC 17020 etc. The subcontractor providing critical activities shall be informed that they may be subjected to an assessment by the accreditation body.

If a PT provider subcontract services pertaining to determination of assigned value, conducting homogeneity and stability studies etc. to a laboratory, it shall ensure that such services are subcontracted to competent laboratories conforming to relevant requirements of ISO/IEC 17043:2010.

*Note 1: ISO/IEC 17025, ISO 15189, ISO 17034 Accredited laboratories with appropriate scope of accreditation could be used as an evidence to establish competence.*

The following records generated by subcontractor shall be readily available with PT Provider:

a. Raw test data generated while performing testing/calibration for the purpose of assigning value or homogeneity or stability testing.

b. Method adopted for assigning values and ensuring homogeneity & stability of PT items.

*Note 2: Subcontracted critical activities may be subjected to assessment at the subcontractor locations.*

**5.13 Complaints & Appeals**

Records of all complaints and appeals shall be maintained along with the corrective actions taken to the satisfaction of the complainant. If any appeal is unresolved for over three months it shall be brought to the notice of NABL. The appeal handling process shall be made known to participants**.**

**5.14 Records**

All records which provide evidence carrying out various PT activities shall be maintained for a minimum period of two years.

**5.15 Internal Audits**

The PTP shall, periodically (minimum once in a year) and in accordance with a predetermined schedule and procedure, conduct internal audits of its activities to verify that its operations continue to comply with the requirements of the management system and the requirements of ISO/IEC 17043.

Every internal audit shall essentially cover various stages in a life cycle of a Proficiency testing program like planning, proficiency test item preparation, homogeneity and stability studies, storage and distribution, evaluation and reporting for the operation of a PT scheme as well as generic management system elements.

*Note: Accredited PT providers who are not engaged in organizing PT programs regularly, shall also comply with the requirements detailed above*

In case any of the activities, as permitted vide clause 5.5 of ISO/IEC 17043, are subcontracted, then these shall be covered under the scope of Internal audit. The audit of the subcontracted activities shall include physical visit to the subcontractor premises, preferably when the activities are being actually undertaken. It should also include witnessing of the subcontracted activities like testing, etc, on a sample basis by an internal auditor competent in the relevant activity being audited.

The audit program should generally include horizontal audit and/or vertical audit or both, so that all the sections/ departments are audited for every aspect/ clause of the management system and ISO / IEC 17043 standard.

The audits shall be carried out by qualified* (relevant qualification) and trained** personnel. The auditor shall understand the technical requirements they are auditing and are trained as per standard ISO/IEC 17043 including auditing techniques/processes. Records in form of Certificate shall be established as evidence of the internal auditor training.

*Relevant qualification for a chemical testing activity means that the personnel should be a chemist. The Qualification requirement may be relaxed, provided a technical expert with relevant qualification, accompanies the trained personnel for conduct of audit. However, in exceptional cases, inter-department personnel can also conduct the internal audit ensuring independency of their activity.

**NABL accepts trained personnel preferably who have undergone a 4 day or 5-day course from reputed organization as per ISO/IEC 17025 and/or ISO 15189 and gained knowledge on ISO/IEC 17043 (either through self study self-evaluation mode or internal training or external training of at least 8 hours accompanied with a certificate). However, the trained personnel shall demonstrate the competence of ISO/IEC 17043 to the assessment team.

In case of any nonconformity observed the PT provider shall conduct detailed root cause analysis and identify and take actions to correct the nonconformity observed and institute appropriate systemic corrective actions to prevent recurrences. These shall be appropriately recorded and demonstrable.

Internal audit shall be independent of the activity which is being audited. Personnel shall not audit their own activities. Internal audit may be done by Internal person or

external person (used for purpose of internal audit) to establish the extent of conformity of the PTP to documented requirements and/ or standard ISO/IEC 17043.

When audit findings cast doubt on the effectiveness of the operations or on the integrity of the PT Provider or on the correctness of their documentation, the PT Provider shall take timely corrective actions and shall notify, in writing, its customers whose activities may have been adversely affected.

Audit findings shall have compliance part and/ or Non- Compliance part, Observations, checklist used as evidence of conduct of internal audit. Audit records may be retained for a minimum period of three years.
*Note: The cycle for internal auditing should normally be completed in one year. First IA shall be done after completion of PT round.*

**5.16 Management Reviews**

In accordance with a predetermined schedule and procedure, the PT Providers top management shall periodically (Minimum once a year and preferably after Internal audit (IA) and / or after closure of Internal audit Non – Conformities) conduct a review of its management system and production processes to ensure their continuing suitability and effectiveness and to introduce any necessary changes or improvements. The review shall take account of:

a) the suitability of policies and procedures;

b) reports from managerial and supervisory personnel;

c) the outcome of recent internal audits

d) corrective and preventive actions;

e) assessments by external bodies;

f) changes in volume and type of work;

g) customer, advisory group or participants feedback;

h) complaints and appeals;

i) recommendation for improvement;

j) other relevant factors such as resources, staff training;

Where PT provider is part of larger organization, it shall be preferable to hold a separate review meeting to cover proficiency testing activities.

The inputs to a management review should generally include the analysis and summary on the above topics, as relevant, instead of just an agenda having the above items listed. Findings from management reviews and the actions that arise from them shall be recorded. The management shall ensure that these actions are discharged within an appropriate and agreed timescale.

The typical format of the Minutes of Meeting of MRM may be:

| S. No. | Agenda Points | Discussion | Decision Taken | Responsibilities & Time Frame |
|---|---|---|---|---|
| 1. | the suitability of policies and procedures | | | |
| 2. | reports from management and supervisory personnel | | | |
| 3. | the outcome of recent internal audits | | | |
| 4. | corrective and preventive actions | | | |
| 5. | assessments by external bodies | | | |
| 6. | changes in the volume and type of work | | | |
| 7. | customer, advisory group or participant feedback | | | |
| 8. | complaints and appeals | | | |
| 9. | recommendations for improvement | | | |
| 10. | other relevant factors, such as resources and staff training | | | |
| 11. | Any other points | | | |

**National Accreditation Board for Testing and Calibration Laboratories**
NABL House
Plot No. 45, Sector- 44,
Gurugram – 122003, Haryana
Tel.: +91-124 4679700
Fax: +91-124 4679799
Website: www.nabl-india.org