NABL 181(A)

National Accreditation Board for Testing and Calibration Laboratories (NABL)

SPECIFIC CRITERIA FOR PT PROVIDER ACCREDITATION (AS PER ISO/IEC 17043:2023)

ISSUE NO.: 01
ISSUE DATE: 13-02-2024
AMENDMENT NO.: --
AMENDMENT DATE: --

AMENDMENT SHEET

S. No.	Amendment no.	Page No.	Clause No.	Date of Amendment	Amendment made	Reasons	Signature QA Team	Signature Competent Authority
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								

CONTENTS

S. No.	Title	Page No.
	Amendment Sheet	1
	Contents	2
1.	Introduction	3
2.	References	4
3.	Specific Policy of NABL for PT Providers	5
4.	General requirements	7
5.	Structural requirements	7
6.	Resource requirements	8
7.	Process requirements	12
8.	Management system requirements	21

1. Introduction

All PT Providers are required to comply with all the requirements listed in the International Standard ISO/IEC 17043: 2023 'Conformity Assessment – General Requirement for the Competence of Proficiency Testing Providers' for all types of proficiency testing schemes. ISO 13528:2022 provides support for the implementation of ISO/IEC 17043:2023 particularly, on the requirements for the statistical design, validation assigned value of proficiency test items, ensuring reasonableness of Standard deviation for proficiency assessment (SDPA), identification of possible blunders (outliers) in case of consensus value approach, review of results, and reporting summary statistics. This document specifies NABL requirements for provider of proficiency testing to comply along with ISO/IEC 17043: 2023. It contains requirements for PT providers to enable them to demonstrate that they operate competently and can generate valid evaluations of participant performance.

PT involves the use of interlaboratory comparisons for the evaluation of laboratory performance. PT can provide evidence of competence and it can be an indicator of an underlying or emerging problem. The need for ongoing confidence in laboratory performance is essential not only for laboratories and their customers but also for other interested parties, such as regulators, accreditation bodies and other organizations that specify requirements for laboratories.

There are many different purposes for interlaboratory comparisons, which can be addressed by PT schemes, including but not limited to:

a) evaluation of the performance of laboratories for specific measurements, tests, calibrations, examinations, inspections or sampling;
b) identification of problems in laboratories that, for example, can be related to measurement or test methods, effectiveness of training and supervision of personnel, or calibration of equipment;
c) establishment of the effectiveness of measurement or test methods and the comparability of measurement and test results;
d) provision of additional confidence to users of measurement and test results;
e) identification of differences in measurement and test results;
f) education of participating laboratories based on the outcomes of such comparisons;
g) validation of measurement uncertainty claims.

For the following types of interlaboratory comparisons, the term PT does not usually apply because laboratory competence must be established in advance, in order to ensure the validity of measurements or tests, as well as the metrological traceability of assigned values:

h) evaluation of the performance characteristics of a measurement or test method (often described as collaborative trials);
i) assignment of values to reference materials;
j) support for statements of the equivalence of measurements of National Metrology Institutes (NMIs), or their Designated Institutes (DIs) through “key and supplementary comparisons”, conducted on behalf of the International Bureau of Weights and Measures (BIPM) and associated Regional Metrology Organizations (RMOs).

It is recognized that interlaboratory comparisons for purposes h), i) and j) can contribute to independent demonstrations of laboratory competence. The requirements of this document can be applied to many of the technical planning and operational activities for these interlaboratory comparisons.

Types of PT schemes – PT schemes vary according to the needs of the sector in which they are used, the nature of the PT items, the measurement and test methods in use and the number of participants. In their simplest form, most PT schemes feature a comparison of results obtained by one laboratory with those obtained by one or more other laboratories. For more details about types of PT schemes refer Annex A – ISO/IEC 17043: 2023.

2. References

2.1. ISO/IEC 17043: 2023, Conformity assessment – General requirements for proficiency testing,

2.2. ISO 13528: 2022 Statistical methods for use in proficiency testing by interlaboratory comparison

3. Specific policy of NABL for PT providers

3.1. The PT provider shall have conducted at least one PT program in same / similar Matrix (PT item) in accordance with ISO/IEC 17043:2023 for the applied scope of accreditation.
Example -In a group (e.g. building materials, metals and alloys, food and agricultural products etc.) with at least some of the critical parameters.

If PT program not conducted or not completed for any part of scope of accreditation, the PT provider shall at least provide PT design document (detailed document as how the program will be conducted) and identify the PT item/artifact and if possible conduct the homogeneity assessment of the PT items prepared.

Note – The intention of the policy is to ensure competency in preparation of all type of PT items covering appropriate discipline/ group/sub-group etc. pertaining to the scope applied.

3.2. The PT provider shall have completed at-least one Internal Audit and one Management review before applying for accreditation and it shall be conducted after completion of PT program(s).

3.3. For calibration PT programs, 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 (say National Physical Laboratory NPL, India).

3.4. For Calibration PT programs, where PT artifact is not owned by PT provider (applicable only for the artifacts where site calibration is done), the following conditions shall be fulfilled:

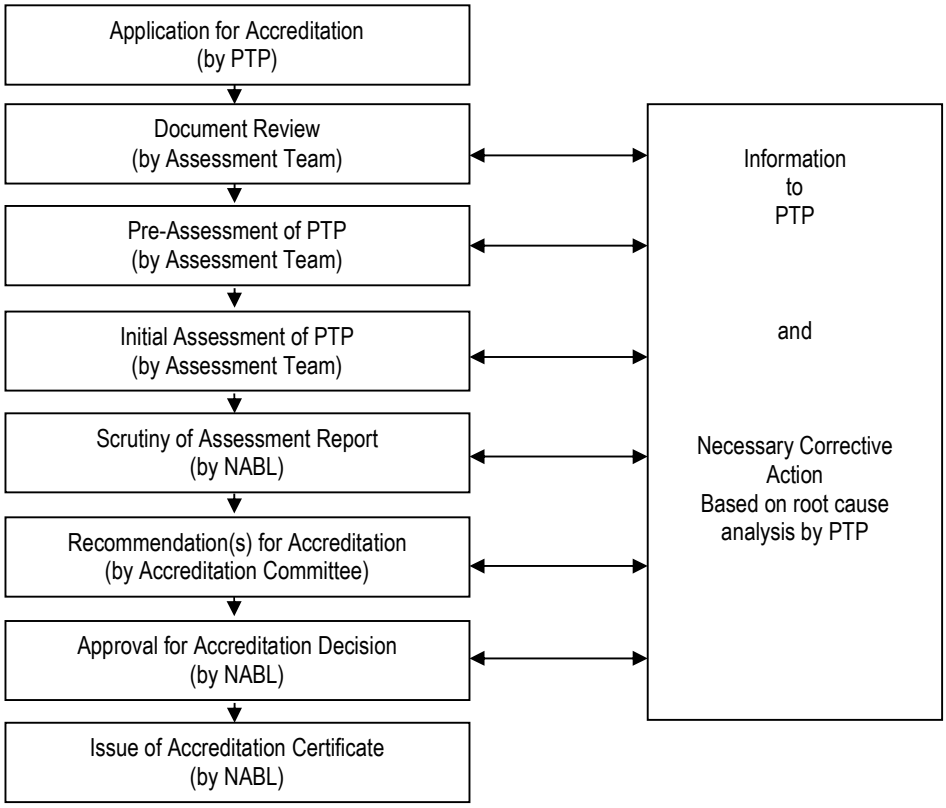
- There shall be a contract between the PT provider and the owner of the artifact. Duration of the contract shall be minimum two years. It shall also describe the conditions for maintenance, servicing, storage, etc., for maintaining integrity of the artifact. It shall also include the condition that during the period of conduct of PT, the artifact shall not be used for any other purpose. The contract shall include among others, requirements for the duration shall not be less than two years.
- The PT provider shall document administrative as well as technical procedure for use of such artifacts, including the system for calibration of the artifact before use/ commencement of PT scheme.
- The PT program shall be completed by PT provider covering all the PT artifacts/ analytes/parameters applied.

3.5. PT provider shall not enroll (or) accept participation from following but not limited to:

- laboratory(ies) used for testing or calibration/ involved in assessing homogeneity and stability of PT item
- organization(s) used for preparation/production of PT item

Note: Checking the competency through this PT is not applicable for these laboratory(ies) / organization(s) as they are already involved / competent in one (or) other activity of PT item.

PTP Accreditation Process Flow Chart



4. General requirements

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 conflict of interest. In addition, the organization chart shall clearly define the position and relationship of PT provider with other activities. PT provider to demonstrate impartiality where conflict of interest is identified and records to be retained.

5. Structural requirements

The applicant/accredited PT provider shall have legal entity as mentioned in NABL 180.

Any changes (e.g. Key Personnel including the members of advisory group/ technical experts, External service providers used for critical PT activities etc.) are to be informed to NABL within the prescribed timeline (refer NABL 131).

6. Resource requirements

a. General

Measurements or tests related to PT item characterization or for assessing homogeneity and stability shall be conducted under the responsibility of the PT provider. All the tests shall be conducted in accordance with the relevant requirements of ISO/IEC 17025 or ISO 15189. Following can be used:

- An accredited laboratory whose service is suitable for the intended use (i.e. the scope of accreditation specifically covers the appropriate measurements or tests) and the Accreditation Body is covered by the ILAC arrangements or by Regional Arrangements recognized by ILAC (e.g. NABL).
- National Physical Laboratory (NPL, NMI of India) or any other NMI whose service is suitable for the intended use and is covered by the CIPM MRA.

Where the PT item is a material that meets the definition of “reference material”, it shall be produced by reference material producer accredited in accordance with ISO 17034 or ISO 15194 or by National Physical Laboratory (NPL, NMI of India) or any other NMI whose service is suitable for the intended use and is covered by the CIPM MRA.

- An accredited reference material producer whose service is suitable for the intended use (i.e. the scope of accreditation specifically covers the appropriate reference material) and the Accreditation Body is covered by the ILAC arrangements or by Regional Arrangements recognized by ILAC (e.g. NABL).

b. Personnel

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.

PT provider shall have competent person to review and authorize the PT reports issued by PT provider under NABL accreditation.

The qualification and experience criteria of relevant field (Testing / Calibration / Medical) shall apply. The relevant academic qualifications, experience and demonstration of technical competence to the assessment team shall be the basis for acceptance of person to review and authorize the report.

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.

The availability of personnel performing key activities to be ensured by PT provider during assessment for demonstration of their competence before assessment team.

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

- Biodata/CV of individuals,
- Purpose for which the expert is to be used,
- Contractual agreement and expert's concurrence to be part of PT provider, (contractual agreement to define the scope of activity which PTP avails from technical experts/ advisory or steering group)
- Confidentiality,
- Minutes of advisory committee and/or records of communications through e-mail or letter. In case, where the activity is critical for performance of PT scheme at-least following shall be available with PT provider.
    - Records of activities monitored and controlled as well as inputs given by them (pertaining to PT item preparation, quality control mechanism and/ or homogeneity assessment results).
    - Record of communication when making comments on participants results.
    - Records of communication when making comments or recommendations based on the outcomes of the PT round.

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

S. No.	Items / Particular	Competence requirements	Name of the Person	Remarks/ Comments
1.	For Planning of proficiency testing schemes			
2.	To assess data/information to determine stability and homogeneity, as well as assigned values and associated uncertainties of the properties or characteristics of the PT item			
3.	To evaluate the performance of PT participants			
4.	To give opinions and interpretations as well as advice to the participants			
5.	To review and authorize PT reports			

The person responsible for quality management system / quality manager should be fully aware of the requirements of ISO/IEC 17043:2023, and principles applicable to the organization's field of accreditation / compliance.

Note - The PT provider to ensure availability of documented information demonstrating competence of personnel for authorized PT activities.

c. Facilities and environmental conditions

PT provider to ensure availability of appropriate facilities for the different operations while conducting PT programs including but not limited to the following:

i. Production
ii. Packaging
iii. Handling
iv. Storage

In case of incompatible PT activities, the PT provider shall take measures such as appropriate separation between neighboring areas to prevent cross-contamination or interference. The PT provider shall also control the activities which can adversely influence the PT scheme. Records of such measures and controls shall be maintained by the PT provider.

d. External service provider

The PT provider may use external services for activities permitted by the standard. When using external services, the PT provider must be able to demonstrate to NABL that each external service provider complies with the relevant standard requirements. The external service provider providing critical activities shall be informed that they may be subjected to an assessment by NABL.

The information that shall be made available to NABL about external service provider(s) includes but not limited to:

- Name and address of the external service provider(s);
- Key activities performed by the external service provider(s); and
- Information about how the PTP assesses the competence (e.g. copy of accreditation certificate, audit reports) of each external service provider.
- 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, ISO/IEC 17020 and ISO 15194 etc.

The following information shall be readily available with PT provider for the test or measurement carried out by either in-house or external service provider:

- Test report / calibration certificate along with original observation data generated while performing testing/calibration for the purpose of assigning value and/or homogeneity and/or stability testing.
- Method adopted for assigning values and ensuring homogeneity & stability of PT items.

7. Process requirements

7.1. **Establishing, contracting and communicating the PT scheme objectives**

- The participants should be suitably informed that the records of PT scheme participation shall be accessible to NABL.
- Records of relevant communication shall be maintained and retained by the PT provider for a minimum period of 2 years.

7.2. **Design and Planning of a PT scheme**

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

S. No.	Title	Particulars	Comments / Remarks
a)	the personnel involved in the design and operation of the PT scheme		
b)	the activities to be undertaken by external providers of products and services and their contact details		
c)	criteria to be met for participation in the PT scheme		
d)	the number and type of expected participants in the PT scheme		
e)	description of activities to be performed and results to be reported by participants		
f)	a description of the range of values or characteristics, or both, to be expected for the PT items		
g)	the potential major sources of errors involved in the area of PT offered		
h)	requirements for the production, quality control, storage and distribution of PT items		
i)	arrangements to prevent collusion between participants or falsification of results and procedures to be employed if collusion or falsification of results is suspected		
j)	a description of the information which will be supplied to participants and the time schedule for the various phases of the PT scheme		

S. No.	Title	Particulars	Comments / Remarks
k)	for continuous PT schemes, the frequency or dates upon which PT items will be distributed to participants, the deadlines for the return of results by participants and, where appropriate, the dates on which measurements or tests will be carried out by participants		
l)	any information on methods or procedures which participants must use to store, handle, prepare, ship or dispose of the PT item and perform the measurements or tests		
m)	procedures for the measurement or test methods to be used for the homogeneity and stability testing of PT items and, where applicable, to determine their biological viability		
n)	preparation of any standardized reporting formats to be used by participants		
o)	a detailed description of the statistical analysis to be used		
p)	the origin, metrological traceability and uncertainty of any assigned values; *NOTE: Assigned values can have uncertainty contributions from sources in addition to the uncertainty of measurement results used for characterization, such as inhomogeneity and instability, and interlaboratory differences if more than one laboratory is used for characterization*		
q)	the treatment of results from different measurement or test methods, where permitted by the PT scheme		
r)	criteria for the evaluation of the performance of participants		
s)	a description of the data, interim reports or information to be returned to participants		
t)	a description of the extent to which participant results, and the conclusions that will be based on the outcome of the PT scheme, will be made public or shared		

S. No.	Title	Particulars	Comments / Remarks
u)	actions to be taken in the case of lost, delayed or damaged PT items		

Note:

- *It is expected that PT provider incorporates sufficient information in PT plan document for all the requirements given in clause 7.2.1.3 (a to u) of ISO/IEC 17043: 2023.*
- *When assigned value (AV) is determined independent of participants, the uncertainty of AV should be calculated considering, uncertainty due to inhomogeneity and instability of the PT items. While following consensus value approach uncertainty need not include components due to inhomogeneity and instability but should include uncertainty due to interlaboratory differences.*
- *Uncertainty due to interlaboratory differences to be included when AV determined from consensus approach from expert laboratories.*

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:

- PT provider should provide adequate justification for the method adopted for assigning value to PT items.
- The size/amount of PT item sent to participants should not be too large. 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.
- Distributing PT items to participants for a particular property or characteristics with slightly varying property values instead of sending same PT items to all the participants, without prior intimation to participants.
- 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.

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:

- PT provider being part/related to another body such as testing/ calibration laboratory (under same ownership, common management, same personnel, coercion openly or secretively, etc.)
- Participation from competitor laboratories in PT program.
- Collusion among participants such as discussion among participants, testing/calibration of PT item in another laboratory (better equipped, specialized).
- Request from participants for duplicate PT items with false justification.

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

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

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.

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.

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

The PT provider should have provision to get participants original observation data of test or measurement conducted on PT item, when collusion or falsification of results are suspected.

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

Statistical design – Refer Annexure – A

ISO 13528 provides following statistical methods for the determination of the assigned value.

a) Formulation
b) Certified reference material
c) Results from one laboratory
d) Consensus value from expert laboratories
e) Consensus value from participant results

Similarly, ISO 13528 provides following statistical methods for the determination of SDPA.

a) By perception of experts
b) By experience from previous rounds of a proficiency testing scheme
c) By use of a general model (Horwitz Formula)
d) Using the repeatability and reproducibility standard deviations from a previous collaborative study of precision of a measurement method
e) From data obtained in the same round of a proficiency testing scheme
f) However, In order to ensure the reliability of PT scheme it is always preferable to determine both AV and SDPA independent of the participants results.

In case of determination of AV and/or SDPA from participants results PT provider to ensure:

- validation of AV and reasonableness of the SDPA.
- robust statistical methods such as Algorithm A, nIQR etc for the determination of AV and SDPA can only be used when the number of effective participation (number of participants after removing the odd values/blunders) is ≥12.

When the number of effective participations is <12, SDPA shall be fixed independent of the participant result.

Measurement Uncertainty by participants in PT programs on testing

PT programs on testing should request participants to report Measurement Uncertainty also. Wherever PT provider asks the participants to report the measurement uncertainty, PT provider should consider the measurement uncertainty while evaluation the performance of the participants and same to be incorporated in the PT report. (refer clause 9.6 and 9.8 (including exercise E 4 of ISO 13528:2022)

Ensuring reasonableness of SDPA

When SDPA is determined from the participants results ensure the reasonableness of the same. The SDPA determined by this approach can be either too low or too high due to various reasons. In such cases the PT provider shall fix Lower limit as well as Upper limit of SDPA (refer clause 8.6.2.1 and 8.6.2.2 of ISO 13528:2022).

Note – The limits may be fixed by perception e.g. Lower limit may be the repeatability standard deviation and Upper limit may be fixed as 1.5 times reproducibility standard deviation defined in the measurement method.

Assigned value - Refer Annexure - A

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

In sequential PT scheme including calibration PT program, where a reference value is assigned prior to the commencement of a round, and the reference value is subsequently checked using the same measuring system, the difference between the values shall be less than two times the uncertainty of that difference (that is, the results shall be metrologically compatible). In such cases the proficiency testing provider may choose to use an average of the measurements as the assigned value, with the appropriate uncertainty. If the results are not metrologically compatible, the proficiency testing provider should investigate the reason for the difference, and take appropriate steps, including use of alternative methods to determine the assigned value and its uncertainty or abandonment of the round of the proficiency testing scheme.

7.3. **Production and distribution of PT items**

i. The PT provider shall have procedure for production of homogeneous and stable PT items which are fit for the PT scheme purpose. PT item production procedures shall be designed in such a way to ensure that the items are as homogeneous and stable as possible. PT items shall be similar to samples routinely tested by laboratories.

ii. The records of selection, acquisition, collection, identification, preparation, handling, storage and, where required, disposal of all PT items to be maintained by PT provider.

iii. Records to demonstrate controls during PT item preparation/ production which directly affects the validity of PT items/scheme shall be retained.

iv. Individual samples purchased from market having traceability to ensure from the same batch production by the manufacturer to be established during acquisition. Merely purchasing of PT

item from the market and statistically establishing homogeneity and stability through testing is not the intent of the standard.

v. In case of calibration PT programs, stability of PT items/artifacts shall be done before commencement / distribution for a period commensurate with the duration of PT scheme. Where the error is introduced in the PT artifact there shall be sufficient control and documentation/records to prove that PT provider has checked that the repeatability and other characteristics of the PT item are not compromised.

Note-

- *The PT provider to inform the participant that the distributed PT item/material is fit only for the evaluation purpose.*
- *Assessment team to have confidence in all PT schemes before recommending the scope. If there is any doubt on the competency then those schemes may not be recommended in accredited scope.*

vi. Homogeneity and stability assessment of PT item

For assessment of homogeneity PT provider to select a number g of the proficiency test items in their final packaged form using a suitable random selection process, where g ≥ 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.

As stated under NOTE 2 in clause 7.2.2 of ISO 13528:2022, “when σpt (SDPA) is calculated as the standard deviation of participant results, the uncertainty components due to in-homogeneity, transport, and instability are in large part reflected in the variability of participant results. In this case the uncertainty of characterization, as described in clauses 7.3 to 7.7 of ISO 13528:2022 is sufficient”.

As stated in clause B.1.2 of ISO 13528:2022, 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_w.

As stated in clause B.2.5 (c) of ISO 13528:2022, “when σpt (SDPA) is the robust standard deviation of participant results, then the in-homogeneity between proficiency test items is included in σpt (SDPA) and so the criterion for acceptability of homogeneity can be relaxed, with caution”.

When evaluating homogeneity and stability in a microbiological PT program the requirement of ISO/TS 22117 shall also be considered where applicable.

7.3.1. **Handling and storage of PT items**

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.

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

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

7.3.2. **Instructions for participants**

PT provider to ensure all the instructions as stipulated in clause 7.3.5 of ISO/IEC 17043: 2023 are duly incorporated. Further, the following points are also to be included:

- Participants shall carryout testing or calibration of PT item/artifact by themselves only and the participants are not allowed to utilize external service providers.
- Participants shall send original observation data of test or measurement conducted on PT item, when requested by PT provider.
- For quantitative PT schemes, participants shall not provide censored results (such as <10 ppb) and the result obtained by the participants on conduct of measurement or test shall be reported quantitively even if the result obtained by the participant is below their limit of detection (LOQ/LOD).
- Format for recording/reporting of the PT results including unit of measurement and number of significant figures (number of decimals) to be reported.

7.4. **Evaluation and reporting of PT scheme results**

Identification of potential outliers (odd values / gross errors/ blunders)

Different methods may be used for identification of potential outliers:

- Visual examination of the plot of the data
- Use of perception
- Grubbs test etc., However Grubbs test alone may not be adequate.

PT provider may examine the performance scores to validate the identification of potential outliers e.g. identified potential outliers score “unsatisfactory” performance in the PT round. If not, identification of potential outlier needs to be reviewed.

Approach for small number of participants

When the number of participants (after removal of blunders) is less than 4 and either assigned value or SDPA has to be determined from the participants’ results, don’t evaluate the performance of the participants, as the same may not be reliable.

PT 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 participants.

In sequential PT schemes, e.g. where the turn-around time can be very long, and in PT schemes involving perishable materials, preliminary or anticipated results may be provided by PT provider before final results are disclosed. Preliminary or anticipated results allow the participant(s) for early investigation of possible errors. While providing preliminary or anticipated results PT provider to ensure that it should not amount to early disclosure of the assigned value.

7.5. **Controls of the PT scheme process**

All technical records shall be maintained and retained for at-least 2 years.

Surveillance of the processes

The PT provider shall have documented plan and procedure to ensure the validity of the PT scheme. The surveillance activity shall be carried out at-least once a year in between the internal audits.

The surveillance activity may include verification of following:

- evaluation of externally provided products and services;
- reference materials or other control items;
- for continuous schemes, comparisons against previous PT rounds.
- Quality of starting material used for production of PT items
- Adequacy of PT Plan prepared
- Adequacy of production or preparation of PT items, their storage and dispatch
- Adequacy of container of PT items and packing material used
- Checking the statistical calculations including homogeneity and stability assessment, assignment of property values, SDPA and performance evaluation of participants
- Checking transmission of results from participants and the validity of data entry and data transfer
- Checking of PT reports

7.6. **Handling of Complaints**

Records of all complaints shall be maintained along with the corrective actions.

7.7. **Handling of Appeals**

Records of all appeals shall be maintained along with the actions taken. 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.

8. Management system requirements

8.1. **General requirements**

Requirements given in clause 8.1 of ISO/IEC 17043: 2023 to be complied.

8.2. **Management system documentation**

Requirements given in clause 8.2 of ISO/IEC 17043: 2023 to be complied.

8.3. **Control of management system documents**

Requirements given in clause 8.3 of ISO/IEC 17043: 2023 to be complied.

8.4. **Control of records**

Requirements given in clause 8.4 of ISO/IEC 17043: 2023 to be complied.

8.5. **Action to address risks and opportunities**

Following may be considered but not limited to as possible risks:

- Collusion either amongst the participants or between the PT provider and participants. (refer section 7.2)
- Adequacy of container, packing material, mode of distribution of PT items
- Non-identification of outliers / blunders for determining AV and SDPA
- Insufficient number of participants for the PT statistical design

Following may be considered but not limited to as possible opportunities:

- Increase in number of participants in a PT scheme
- Expanding the scope of PT activities
- Making the PT scheme more cost effective and reducing the time of PT item preparation

8.6. **Improvement**

Requirements given in clause 8.6 of ISO/IEC 17043: 2023 to be complied.

8.7. **Corrective actions**

Requirements given in clause 8.7 of ISO/IEC 17043: 2023 to be complied.

8.8. **Internal Audits**

The PT provider shall, periodically (minimum once in a year) and in accordance with a predetermined schedule and procedure, conduct internal audits (IA) 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: 2023.

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, wherein external service provider(s) are used, then these shall be covered under the scope of Internal audit. The audit of the external service provider activities shall include physical visit to the external service providers premises, preferably when the activities are being actually undertaken. It should also include witnessing of the externally provided 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: 2023 standard.

The audits shall be carried out by qualified* (relevant qualification) and trained** personnel. The auditor shall understand the requirements they are auditing and are trained as per standard ISO/IEC 17043: 2023 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 who have undergone training as per ISO/IEC 17025 and/or ISO 15189 and gained knowledge on ISO/IEC 17043: 2023 (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: 2023 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 PT provider to documented requirements and/ or standard ISO/IEC 17043: 2023.

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 inter audit shall be done after completion of PT round.

8.9. **Management Reviews**

In accordance with a predetermined schedule and procedure, the PT provider's 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) changes in internal and external issues that are relevant to the PT provider;
b) fulfilment of objectives;
c) suitability of policies and procedures;
d) status of actions from previous management reviews;
e) outcome of recent internal audits;
f) corrective actions;
g) assessments by external bodies;
h) changes in the volume and type of the work or in the range of PT activities;
i) customer, participant and personnel feedback;
j) complaints and appeals;
k) effectiveness of any implemented improvements;
l) adequacy of resources;
m) results of risk identification;
n) outcomes of the surveillance of the processes;
o) other relevant factors, such as 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 managemnet review is given below:

S. No.	Agenda Points	Discussion	Decision Taken	Responsibilities & Time Frame
1.	changes in internal and external issues that are relevant to the PT provider			
2.	fulfilment of objectives			
3.	suitability of policies and procedures			
4.	status of actions from previous management reviews			
5.	outcome of recent internal audits			
6.	corrective actions			
7.	assessments by external bodies			
8.	changes in the volume and type of the work or in the range of PT activities			
9.	customer, participant and personnel feedback			
10.	complaints and appeals			
11.	effectiveness of any implemented improvements			
12.	adequacy of resources			
13.	results of risk identification			
14.	outcomes of the surveillance of the processes			
15.	other relevant factors, such as training			

Annexure

A. Statistical design

(Extract from ISO 13528: 2022 (E) Cl 5.2, 5.2.1, 5.2.2, 5.2.3)

A.1 **Basis of a statistical design**

- According to ISO/IEC 17043, 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.

- 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.

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.

A.2 **Homogeneity and stability**

Extract from ISO 13528: 2022 (E) Cl 6.1, 6.1.1, 6.1.2, 6.1.3 Annex B, Annex E.2

- 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:
    - experimental studies as described in Section B or alternative experimental methods that provide equivalent or greater assurance of homogeneity and stability;
    - experience with the behavior of closely similar proficiency test items in previous rounds of the proficiency testing scheme, verified as necessary for the current round;
    - 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.

- 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 artifact.

- All measurands (or properties) should normally be checked for homogeneity and stability. However, where the behavior 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 A.2 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:

    - 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;
    - If a proficiency test item is heated during processing, then choose a measurand that is sensitive to uneven heating;
    - 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 Table a, Table b, using statistical methods recommended.

B. Statistical Method recommended for Homogeneity and Stability Check

B.1 General Procedure for a homogeneity check

B.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 δ_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 ≥ 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 ≥ 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 x 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 $\bar{x}$ average, within-samples standard deviation s_w, and between-samples standard deviation s_s, as shown in B.3.

B.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.

B.2 Assessment criteria for a homogeneity check

B.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 B.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 B.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.

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.

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

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

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 δ_E:

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

B.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 B.3 and F_1 and F_2 are from standard statistical tables, reproduced in Table B.1, for the number of proficiency test items selected and with each item tested in duplicate

Table B.1 —Factors F_1 and F_2 for use in testing for sufficient homogeneity

gm	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 B.2.3 b) and Table B.1 shall be replaced with $F_m = (F_{g-1,\, g(m\text{-}1),\, 0{,}95-1})/m$ where $F_{g-1,\, g(m\text{-}1),\, 0.95-1}$ 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 B.1 are derived from standard statistical tables as follows:
$F_1 = \chi^2_{0.95(g-1)} / (g\text{-}1)$ 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
$F_2 = (F_{0.95(g-1;g)}\text{-}1)/2$ where $F_{0.95(g\text{-}1;g)}$ 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

B.2.4 When σ_{pt} is not known in advance, for example when σ_{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 α=0.05;
b) use information from previous rounds of the proficiency testing scheme to estimate σ_{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 σ_{pt} has been calculated.

B.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 σ'_{pt} as in equation (B.3). Note this needs to be described fully to participants.

$$\sigma'_{pt} = \sqrt{\sigma_{pt}^2 + s_S^2} \qquad \text{(B.3)}$$

b) Include s_s in the uncertainty of the assigned value and use *z'*or δ_E' to assess performance (see Section 9.5 of ISO 13528: 2022);

c) When σ_{pt} is the robust standard deviation of participant results, then the inhomogeneity between proficiency test items is included in σ_{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.

B.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:

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

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

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

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

Calculate the general average:

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

the estimate of the variance of sample averages:

$$s_x^2 = \frac{1}{(g-1)}\sum_{t=1}^{g}(\bar{x}_t - \bar{\bar{x}})^2 \qquad \text{(B.7)}$$

and the within-samples variance:

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

Estimate the combined variance of s_s and s_w.

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

Finally, estimate the between-samples variance as

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

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:

$$x_{t,\cdot} = \left(x_{t,1} + x_{t,2}\right)/2 \qquad \text{(B.11)}$$

and the between-test-portion ranges as:

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

Calculate the general average:

$$\bar{x}_{..} = \sum \bar{x}_{t,.}/g \qquad \text{(B.13)}$$

Estimate the standard deviation of sample averages:

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

and the within-samples standard deviation:

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

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

Finally, estimate the between-samples standard deviation as:

$$s_S = \max\left(0, \sqrt{s_x^2 - (s_w^2/2)}\right) \qquad \text{(B.16)}$$

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 Table a.

B.4 **Procedures for Checking Stability**

B.4.1 **General Considerations for checking stability**

B.4.1.1 These clauses give guidance for meeting the stability requirements of A.2. The provisions of section A.2 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.

B.4.1.2 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.

B.4.1.3 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>0,5σ_{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.

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

B.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.

B.4.2.2 If a proficiency testing provider includes shipped proficiency test items in the stability assessment in B.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 B.6 should be used.

B.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.

B.4.2.4 The following variations to the procedure in B.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.

NOTE Procedures using observations at regular intervals between the beginning and end of a proficiency testing scheme round can also be used and can be advantageous if measurement system variation over time is large enough to compromise the assessment described at B.5.

B.5 Assessment criterion for a stability check

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

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

B.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 σ_{pt} using the following formula:

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

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.

B.5.3 If the criterion in (B.17) or (B.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.

B.5.4 The criterion at B.5.1 or B.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 (B.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.

B.6 **Stability in transport conditions**

B.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.

B.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.

B.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 B.5.1 or B.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 Table b

An Example for Homogeneity and Stability check

Homogeneity and Stability test – Arsenic (As) in chocolate (Example E2 of ISO 13528:2022)
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-a below. The procedure in B.3 is followed, resulting in the summary statistics listed. The fitness-for purpose σ_{pt} for As in chocolate is 15%, so the estimate of sample variability is checked against 0,3 times the σ_{pt}.

Table a—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

σ_{pt} =0,18715 x0,15 = 0,02807

Check value: 0,3 x σ_{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 b—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: 0,3xσ_{pt} = 0,00842

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

C. Assigned value

(Extract form ISO 13528: 2022 (E) Cl 7.8, 7.8.1, 7.8.2, 7.8.3)

Comparison of the assigned value with an independent reference value

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 given below

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.

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.

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.

D. Qualitative PT Schemes

(Extract form ISO 13528: 2022 (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

D.1 **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).

D.2 **Statistical design**

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.

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.

EXAMPLE 1 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.

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.

D.3 Assigned values for qualitative proficiency testing schemes

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.

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. 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.

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.

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.

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

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.

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.4 Performance evaluation and scoring for qualitative proficiency testing schemes

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.

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.

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, ζ, 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 of6(so a result adjacent to the assigned value would attract a score of 2).

III. 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.

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 performanceassessment. 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.

Graphical methods may be used to provide performance information to participants or toprovide 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.*

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

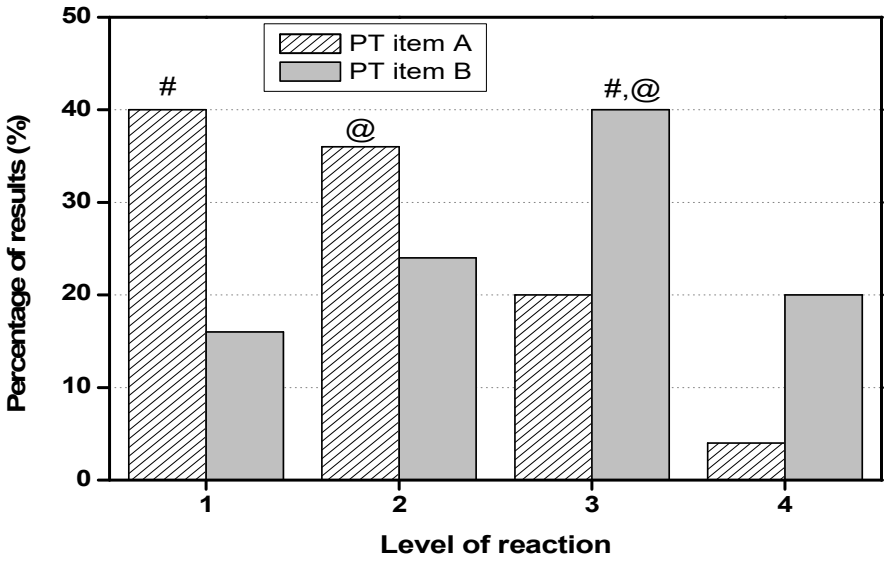


Figure E.15— Bar chart of percentage responses to two skin irritation proficiency test items

Note- 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".

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