

NIST Office of Weights and Measures (OWM) PT Report Supplement

1. Purpose

The purpose of this document is to assist OWM laboratory participants, laboratory management, laboratory recognition and accreditation bodies, and assessors to interpret and analyze OWM PT reports. This supplement is an integral part of each PT report but is not copied and integrated into each report to simplify and minimize extra documentation that is generic and duplicative in each report. Portions of the PT Plan and of the PT Analyses spreadsheets provide the foundation of the PT report. Each unique PT report includes components from the planning, organization, PT artifact identifications and purpose(s), participant identification, operations, as well as the draft and final analyses, along with associated data, charts, and graphs for each unique PT.

2. OWM Policies and Quality System¹

The Office of Weights and Measures (OWM) is not an accredited PT provider. However, the OWM PT Program seeks to comply with well-designed quality systems, laboratory and accreditation body needs, ILAC PT policies, as well as ISO/IEC 17043:2010 and ISO/IEC 13528:2015 (where applicable).

2.1. **NISTIR 7082, Proficiency Test Policy Plan", January, 2018**

This publication provides the policies and plans for the PT Program of the NIST Office of Weights and Measures. This Office of Weights and Measures (OWM) Proficiency Testing (PT) policy and plan has been updated to ensure compliance with the latest applicable documentary standards and policies of the International Laboratory Accreditation Cooperation (ILAC).

The PT program has been in place since the early 1980s as a core part of the support to State weights and measures laboratories through regional measurement assurance programs. Original activities were conducted as “round robins” in support of ongoing measurement assurance activities related to support for State laws with requirements for metrological traceability to national and international standards.

2.2. **NISTIR 7214, Weights and Measures Division Quality Manual For Proficiency Testing and Interlaboratory Comparisons, March, 2005**

This document contains the OWM Quality Manual for Proficiency Testing and Interlaboratory Comparisons. This document provides the quality system to ensure that all Proficiency Testing and Interlaboratory Comparison activities within OWM

¹ All NIST references noted in this section are publicly available and posted on the NIST website at: <https://www.nist.gov/pml/weights-and-measures/laboratory-metrology/proficiency-testing>.

are compliant with ILAC-G13:2000, Guidelines for the Requirements for the Competence of Providers of Proficiency Testing Schemes and ISO/IEC 17043:2010. This document specifies requirements to ensure that OWM and technical advisory groups, PT coordinators, and PT analysts are technically competent to provide specific types of proficiency testing schemes as required by NISTIR 7082, Proficiency Test Policy and Plan (for State Weights and Measures Laboratories), 2018. (NOTE: as written, this document in 2005 was design to comply with ISO Guide 43 and is in the process of being updated to ISO/IEC 17043).

2.3. NIST OWM Standard Operating Procedures and Resources

2.3.1. *SOP for PTs: Standard Operating Procedure for Office of Weights and Measures Proficiency Tests (OWM PT)*

This procedure is used by the OWM Proficiency Testing (PT) Program, which is operated to support the State legal metrology laboratories. This procedure is part of the OWM PT Quality System which includes NISTIR 7214 “Weights and Measures Division Quality Manual for Proficiency Testing and Interlaboratory Comparisons”, NISTIR 7082 “Proficiency Test Policy Plan”, and associated PT Tools. This rigorous procedure describes how to implement a PT in the OWM program from planning through to final reporting. Specific instructions are provided in the SOP for all PT stages to ensure compliance with the OWM Quality System, Policies, and ISO/IEC 17043 and to provide the rigor needed to provide exceptional quality for participants and to meet minimum requirements of accreditation bodies.

Table 1. PT Stages and Resources to be Used.

| PT Stage | Resource to be Used |
|-------------------|--|
| Planning | OWM PT Plan Template (sections P1, P2, P3, and P4) |
| Operation | OWM PT Plan Template (sections O1 and O2) |
| Analysis | OWM PT Analysis Template |
| Reporting | <i>Incorporated into PT Plan and PT Analysis Templates</i> |
| Follow-Up Actions | OWM PT Follow-Up Form (Required by State weights and measures laboratories during annual reviews per NIST Handbook 143, Program Handbook, 2019.) |

2.3.2. *SOP for MiniMAPs:*

This procedure is for the operation of a small interlaboratory comparison, typically only between two or three laboratories where a full proficiency test among a regional group or a national assessment is not readily available to meet the needs of the laboratory or where there is a very small number of laboratories with similar capabilities. Integrated “measurement assurance” assessments are a key part of conducting small proficiency tests, hence the

idea for calling them Mini-Measurement Assurance Programs, or “mini-MAPs”. Given the constraints in the usual small number of participants for a Mini-MAP, additional assessments in addition to the proficiency testing (PT) components are essential to provide additional data validity.

2.4. **ILAC PT Policies**

2.4.1. *ILAC-P9:06/2014, ILAC Policy for Participation in Proficiency Testing Activities*

OWM is not an accreditation body nor an ILAC signatory; however, OWM seeks to comply with the policies described in this ILAC policy document. The following items are paraphrased from the ILAC policy, section 4, and includes Notes regarding OWM applications.

1. Accreditation bodies must verify competency of accredited labs; one way may be through proficiency testing. Note: OWM recognizes laboratories based on demonstrated competency through required training, assigned Laboratory Auditing Program problems, onsite observations, in addition to formal PTs and MiniMAPs.
2. Minimum PT activity related to the laboratory Scope will be a) successful PT participation prior to accreditation where available and appropriate and b) ongoing activities consistent with a PT Plan. Note: OWM requires PTs in all measurement areas prior to Recognition, where *reasonably* available according to NISTIR 7082, Policy and Plan, and requires laboratories to maintain a PT Plan (generally through the Regional Measurement Assurance Program, RMAP groups).
3. Accreditation bodies shall have documented policies and may provide additional resources for laboratories regarding PTs and interlaboratory comparisons used for purposes other than PTs. NOTE: NISTIR 7082 Policy and Plan addresses ILAC policy requirements.
4. Some measurement areas in legal metrology may not readily allow for PTs as demonstration of competence – such as large mass standards above 500 lb, LPG provers, weight carts, railroad test cars, circulation of balances and scales, etc. In those cases, training and on-site observations may be substituted as suitable demonstration of competence.

3. **Technical Analysis**

3.1. **Statistical Concepts and Analyses**

Where possible, artifacts with stable historical reference values are chosen for PTs. During the planning process, clear objectives are chosen and artifacts selected to meet designated PT objectives. Data is evaluated during the course of the PT by the PT Coordinator, PT Analyst, and/or OWM staff as much as possible during the

course of the PT to provide immediate feedback to each laboratory regarding potential failures or need for corrective actions. Interim E_n or P_n values may have been provided to the laboratory, but reference values of individual standards should *not* have been provided. Interim E_n or P_n values may not match with final reports based on the final technical analysis and selection of the assigned reference values that may occasionally change from what was originally planned and intended. A detailed assessment of all data is conducted during the preparation of draft and final reports.

3.1.1. *Official Values Identified for Each Laboratory*

All data is reported and assessed in the final PT report. However, to avoid having a mean value that is unduly influenced by multiple participants from a given laboratory, the statistical evaluation represents the data of only one participant. The data from these designees are referred to as the official values. Official values for one staff member are designated by the laboratory when submitting PT results to the PT Coordinator or Analyst.

3.1.2. *Initial Data Reviews: Outliers, Blunders, Trends (Drift and Shifts), and Corrective Actions During PT and Draft Reviews (Prior to PT Completion)*

The “Initial Statistics” are calculated using all the official values. All data is visually evaluated to look for excessive variability, trends/drift, and/or major changes to the measurement results during the PT and after all measurements are completed. Closing values from the starting laboratory may be necessary when there are questions about the stability of the artifact. Data is reviewed for obvious blunders (such as typographical mistakes), unexplained outliers (values outside of three standard deviations of all participant results), any widely fluctuating results that may represent instability or poor handling of the standards, and potential corrective actions that may be needed which can be completed prior to finalizing the PT (when time is available) and prior to completing the final PT analysis.

3.1.3. *Adjusted Statistics (Trimmed Mean, Trimmed Standard Deviation)*

The “Adjusted Statistics” are determined after official values that fail certain criteria are omitted. Values that are outside of two standard deviations (this must be done in one iteration) are omitted. Then the values that fail the E_n and P_n calculations are omitted. All “omitted” values from the calculations are retained in the report but not used in selecting the assigned reference value(s). The adjusted mean and adjusted standard deviation are used when evaluating and determining the assigned reference value(s).

3.1.4. *PT Standard Deviation*

The standard deviation of the official values, after any adjustments, is used as the PT standard deviation. This value is used in the Z-score calculation and may be used in ongoing analysis of expected PT variability and estimating future expected PT variability during the Planning phase.

3.2. **Determination of the Assigned Reference Value and Its Uncertainty**

3.2.1. *Metrological Traceability Required for Participants and Assigned Reference Value*

The OWM PT policy and plan (NISTIR 7082) requires all OWM PT participants to have demonstrated metrological traceability, either through OWM laboratory Recognition, Accreditation through an Accreditation Body that is an ILAC Signatory, or through an assessed process that is compliant with ISO/IEC 17025:2017 and OWM GMP 13 (NISTIR 6969). Because metrological traceability is a requirement of all participants, any laboratory, group of laboratories, or all official values (one per laboratory) could conceivably be used during the assessment of results when selecting a suitable reference value. Figure 1 provides an example traceability hierarchy that demonstrates the concept of metrological traceability as a characteristic of each participant laboratory, of course keeping in mind that each successive level down usually, though not always, has a larger uncertainty. This Figure is also used in the discussion of selecting the assigned reference value.

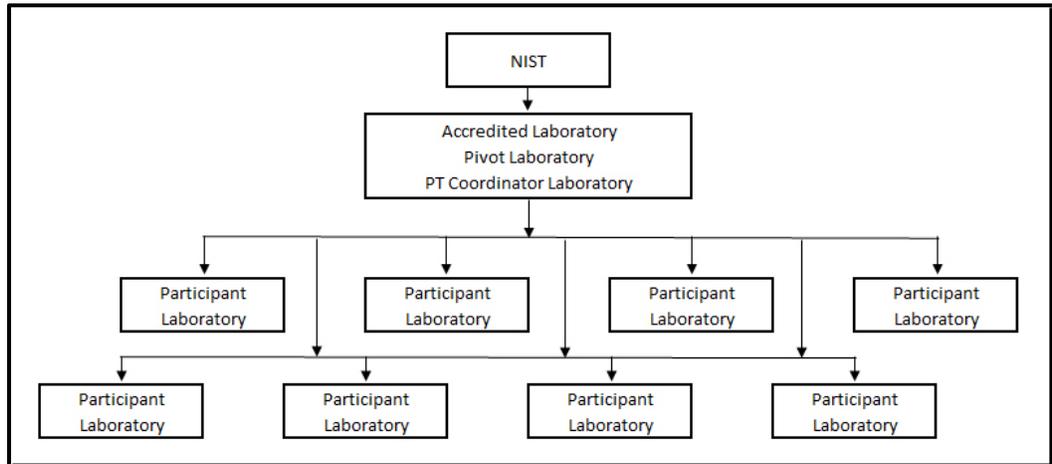


Figure 1. Metrological Traceability Hierarchy.

3.2.2. *Technical Analysis Required for Selecting Assigned Reference Value(s)*

After careful review of all PT data and initial (and adjusted) statistics are determined, possible reference values and corresponding uncertainties are considered for each item. The hierarchy of selecting an assigned reference value is

shown below and preferences are prioritized for selecting an assigned reference value, but the technical assessment of the data by the PT Analyst and OWM staff is required when reviewing options for each standard used in the PT. Even when a higher-level (smaller uncertainty or higher on the hierarchy list) reference value is desired or was used to begin the PT, it may not be the best reference value once all data are reviewed. For example, a precision mass standard may have been calibrated by NIST, but once the item has been circulated, its value might not remain constant as compared to the original NIST value, but was stable later on during the course of the PT. In that case, an historical value or consensus may be the technically correct choice to use as the assigned reference value. In the case of a NIST volume calibration, the uncertainty may not be the smallest or most suitable reference for gravimetric calibrations.

The final selection of assigned reference values is always evaluated and approved by OWM prior to release of a final report. Criteria used by PT Analysts and OWM staff have been evaluated by NIST statisticians to ensure appropriate values are selected for standards used in each PT.

3.2.3. *NMI Reference Value and Uncertainty (Externally Derived Criteria)*

An NMI value, such as one from NIST might be considered an ideal reference value to use when there is also evidence of stability and the uncertainties are sufficiently small relative to the participant values. This source is not always an option due to the high cost and the time associated with obtaining this value. Stability of the standard may also make this value less desirable. The uncertainty associated with using NMI measurement results is taken from the calibration certificate.

3.2.4. *Accredited Laboratory, Pivot Laboratory, PT Coordinator Laboratory Initial Reference Value and Uncertainty (Externally Derived Criteria)*

As a part of the PT Plan, OWM and the Technical Advisory Groups may have decided to use an initial reference value and an ending value from an Accredited laboratory, a Pivot Laboratory, or a PT Coordinator laboratory measurement result and uncertainty. This is the next level down from the NIST or NMI value as shown in the hierarchy in Figure 1. Unless measurement results at this level have uncertainties that are sufficiently smaller than other laboratories in the group, exceptional care must be taken to ensure suitable agreement in the final measurement results to avoid conflict among participants and disagreements about the assigned reference value.

This approach may also be suitable in other instances, for example:

- Where more than one level of calibration will be performed in the PT, with some laboratories performing a higher-level procedure

(lower uncertainty) and the remaining laboratories performing a lower level procedure, a mean value from these laboratories may be used to select a best assigned reference value.

- Where the standard to be used in the PT belongs to one of the participants and significant history of calibrations and stability is available the “owner” may be selected to provide initial and closing measurement results and the value from that laboratory used as the assigned reference value.

3.2.5. *Historical Reference Value and Uncertainty (Externally Derived Criteria)*

An historical reference value can be an individual value or a collection of values from a variety of sources including past NMI calibrations, past RMAP calibrations, or past accredited lab calibrations. The uncertainty is often a mean of the uncertainty of the selected values (average uncertainty from contributing values).

3.2.6. *Mean of “Official” Participants and Uncertainty (Consensus Value) (Comparison Derived Criteria)*

When all official values agree with no need for omitting data as part of the analysis, and when the associated uncertainty is acceptable for the assessment needed, the mean value of all participant results may be used. A weighted mean and weighted uncertainty may be used to ensure that laboratories with smaller uncertainties actually contribute a greater proportion of the assigned values. This value is most often used when there is no other good alternative, or when the tolerances are sufficiently large that the use of this value no significant negative impact on the analysis. The uncertainty is from the standard deviation of values used, multiplied by k as a coverage factor.

3.2.7. *Adjusted Mean (Trimmed) Reference Value and Uncertainty (Comparison Derived Criteria)*

After the initial data is reviewed and initial failures are flagged and removed from the analysis used in selecting the assigned reference value, the values and statistics that are remaining are considered the adjusted, trimmed, or Winsorized mean and an associated uncertainty.

3.2.8. *Simulations and Monte Carlo Assessments*

Although not widely used for OWM PT analyses, this tool generates simulated values based on an inputted distribution and variables for your data set. Simulation iterations can run in the tens of thousands, hundreds of thousands, or more depending on the computing capabilities. When this analysis is conducted, the values are often entered as additional participant

data points for reference in reviewing the graphs and the selection of reference values.

3.2.9. *Multiple assigned reference values.*

Selection of different options may be required for each standard within a set of standards circulated for a given PT. The summary data chart for standard/artifact in the PT will designate what value was used and selected as the assigned reference value. All other statistics performed for that standard will be based on the selected reference value.

3.2.10. *Summary of Methods (from PT SOP)*

Table 2. Selection Hierarchy for Reference Values and Uncertainties.

| Item | Source | Value | Uncertainty | Comments |
|-------|--|---|--|---|
| 3.2.3 | NIST or other National Metrology Institute (NMI) Value (demonstrated appropriate through CIPM MRA database review) | From calibration certificate | From calibration certificate | If values are stable and sufficiently small uncertainty |
| 3.2.4 | Accredited Laboratory, Pivot Laboratory, Small Subset of Participants working at higher level | From calibration certificate | From calibration certificate | If values are stable and sufficiently small uncertainty |
| 3.2.5 | Historically Stable Reference Values | Value used in prior group or Mean of values | Uncertainty from value used in prior group or mean uncertainty | E.g., other RMAP regions If values are stable and sufficiently small uncertainty |
| 3.2.6 | Mean/Median Value – Consensus | Mean or Median value | Weighted uncertainty | E.g., one value per lab |
| 3.2.7 | Adjusted Mean/Median Value - Selected Participants | Mean or Median value | Weighted uncertainty | Must be enough remaining data after adjustments to be valid |
| 3.2.8 | Monte Carlo Simulation Values | Special statistics | Special statistics | |

3.3. Statistical Analyses in the PT Report

The final report presents the reported measurement results and associated uncertainties. All participants and official participant results from each laboratory are identified. According to the OWM policy, there is no assurance of confidentiality in OWM PTs.

Items that are included in the PT Analysis include:

- Tabulations of data submitted and the baseline analysis for each standard/artifact that was calibrated in the PT. Tables contain the

laboratory identification, participant initials, date of calibration, measurement results and uncertainties, initial and adjusted statistics, bias (offsets), E_n , P_n , and Z-scores, status of in/out of two standard deviation limits, and selection criteria for values that were not used in selecting the assigned reference values.

- Summary tables of Pass/Fail statistics showing E_n and P_n values with a total number of failed results for each person.
- Graphs showing measurement results and uncertainties with associated reference values for each standard/artifact in the PT.
- Graphs of E_n and P_n values for each standard/artifact in the PT. In OWM reports, the E_n is graphed with the P_n value. Unlike most PT providers, OWM uses an absolute value for the E_n value so that it can be easily graphed with the P_n statistic on the same chart. To determine consistent directionality of measurement offsets from reference values over time when evaluating uncertainties associated with minor biases in laboratory results, the laboratory can review the Z-score values.

3.3.1. Bias (Difference)

The bias or difference of each reported value from the assigned reference value is calculated and reported as part of the PT analysis data. This value is not used as a pass/fail statistic, but is used in the initial assessment of data by the PT analysis and by OWM to review the overall data for obvious blunders and outliers. The laboratory may use this value as a part of its follow-up assessments of laboratory bias, accuracy goals, and plans for recalibration limits. E.g., for precision calibrations, a laboratory might want to set recalibration goals such that whenever the bias/offset exceeds some ratio of its reported uncertainty, a recalibration or interim assessment of metrological traceability is conducted. Historical OWM PT statistics (no longer used) included an assessment of this offset as shown in Eqn. 2 with a derivation of the Z-score based on laboratory uncertainties rather than the PT statistics. The laboratory may still wish to conduct this assessment for internal evaluations.

$$x_{lab} - X_{ref} \tag{Eqn. (1)}$$

$$OWM_{historical} Z = \frac{x_{lab} - X_{ref}}{U_{lab}} \tag{Eqn. (2)}$$

3.3.2. Normalized Error, E_n

Normalized Error, E_n , is defined in ISO/IEC 17043:2010, Appendix B. E_n is a ratio of the difference between the reference value and the reported value compared to the root sum square of associated expanded uncertainties.

Passing results are < 1 . It is an indicator of accuracy/inaccuracy as compared to an assigned reference value with respect to the associated uncertainties.

$$E_n \text{ assessment: Is } (x_{lab} - X_{ref}) < \sqrt{U_{lab}^2 + U_{ref}^2} ? \quad \text{Eqn. (3)}$$

$$E_n = \left| \frac{x_{lab} - X_{ref}}{\sqrt{U_{lab}^2 + U_{ref}^2}} \right| \text{ Result must be } < 1 \text{ to pass.} \quad \text{Eqn. (4)}$$

A visual assessment of example (unitless) E_n results are shown in Figure 1. Assuming that the assigned reference value of 1.25 with a corresponding expanded uncertainty of 0.5 is correct and acceptable, and that submitted laboratory values vary in a normal distribution, laboratories A, B, and C were selected to illustrate the normalized error concept. In general, the E_n assessment determines the degree to which the values and associated uncertainties overlap each other.

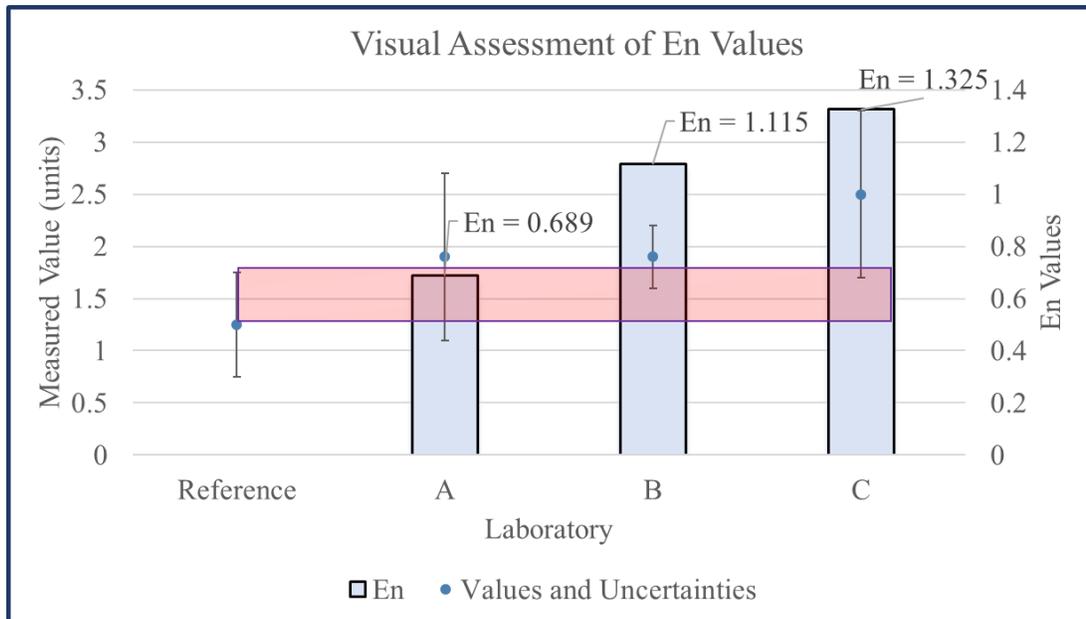


Figure 2. Visual Assessment of E_n Values.

- A: The value submitted by laboratory A is outside the uncertainty of the reference value although its uncertainty overlaps the reference value. Visually, there is a good amount of overlap of the uncertainty bars. The calculated E_n value of 0.689 is less than 1 and passes this assessment. An E_n value of 0.689 *might* still warrant further assessment of the laboratory accuracy by determining if the Bias (Difference) that is shown has been consistent in previous PTs or is observed in a laboratory control chart. Further evaluation would

depend on the applicable tolerances for the application and the desired level of accuracy needed by the laboratory.

- B: The value submitted by laboratory B is identical to the value from laboratory A, thus the Bias (Difference) from Equation 1 is also identical. However, the uncertainty for this laboratory is smaller than the laboratory A uncertainty and also smaller than the reference value uncertainty. While the uncertainty values still overlap slightly, the laboratory uncertainty does not overlap the reference value, the uncertainty of the reference value does not overlap the submitted laboratory B value, and the E_n value of 1.115 fails the assessment. As noted, the Bias (Difference) for both laboratories A and B are identical, but the Uncertainty for laboratory B does not support this level of bias. Either the uncertainty is too small if all other laboratories performed the same procedure and submitted uncertainties comparable to Laboratories A and C (likely) or the laboratory needs to identify the root cause of this failure (e.g., a systematic error of some type or the need for recalibration of standards to bring values closer to the reference value).
- C: The value submitted by laboratory C is not inside reference value uncertainty and its uncertainty is the same as that of laboratory A. In this case, there is very minor overlap of uncertainty values, but the overlap is not enough and the calculated E_n value of 1.325 fails this assessment and corrective action is needed to identify the cause for the inaccuracy shown in the results. Some laboratories working with larger tolerances might suggest that the offset does not matter and the failure is not significant, which is counter to the purpose of PTs. If the tolerances are significantly larger than the offset shown, a larger uncertainty to cover the gap and pass the E_n assessment is likely warranted.

Note that none of the observed biases for laboratories A, B, and C pass criteria in SOP 29 to incorporate bias into the uncertainty in any of these cases!

3.3.3. *Normalized Precision, P_n*

The Normalized Precision, P_n , is an assessment of suitability of the laboratory uncertainty compared to applicable documentary standards where relevant and passing values are < 1 . This statistic is unique to OWM assessments and is not part of the ISO/IEC 17025 or ISO/IEC 17043 system of PT assessments, except indirectly in the context of decision risk evaluations. Documentary standards used in legal metrology generally specify appropriate uncertainty to tolerance (or maximum permissible errors, m.p.e.) ratios on which to base decision risks. In this supplemental report, tolerances and m.p.e. terminology is used interchangeably.

Documented decision risks are required by the ISO/IEC 17025:2017 standard. Many of the OWM published procedures and documentary standards that are referenced for legal metrology include uncertainty to m.p.e. ratios of 1/3:1, where the uncertainty must be less than the applicable m.p.e. or the uncertainty must be less than one-third of the m.p.e. The 1/3 ratio is common in international legal metrology documentary standards such as those from the International Organization of Legal Metrology (OIML) and a number of the NIST Handbook 150-x series documentary standards. ASTM E617 for mass standards also includes this common ratio of uncertainty to tolerances. In some cases the ratio will be 1:1, where the uncertainty must simply be smaller than the applicable tolerance or m.p.e..

The P_n value should be assessed by the laboratory prior to participation in all applicable PTs with corrective action taken prior to participation. Failures of the P_n assessment are preventable with appropriate risk mitigation methods. Therefore, failures of the P_n statistic in a PT always require suitable follow up corrective action and may immediately impact laboratory Recognition and or Accreditation status.

$$P_n \text{ assessment: Is } U_{lab} < \frac{1}{3} m.p.e.? \quad \text{Eqn. (5)}$$

$$P_n = \left| \frac{U_{lab}}{\frac{1}{3} m.p.e.} \right| \text{ Result must be } < 1 \text{ to pass.}$$

alternative ratios that may be used: Eqn. (6)

$$P_n = \left| \frac{U_{lab}}{m.p.e.} \right|, P_n = \left| \frac{U_{lab}}{\text{fraction or \% of } m.p.e.} \right|$$

A visual assessment of example (unitless) P_n results are shown in Five examples are shown to illustrate the relationship between the maximum permissible error (m.p.e.) or tolerances and the uncertainties submitted by the laboratory. In the P_n assessment, the actual values are not what is being assessed. For example, laboratory A is exactly the same as the reference nominal value of zero error, yet the calculated value of its normalized precision is 3, and fails the requirements of being less than one-third of the m.p.e. Also, in the case of laboratories D and E, they have identical passing P_n results even though laboratory D reported a result identical to the reference nominal value and laboratory E is significantly away from the reference value (and would likely fail an E_n assessment).

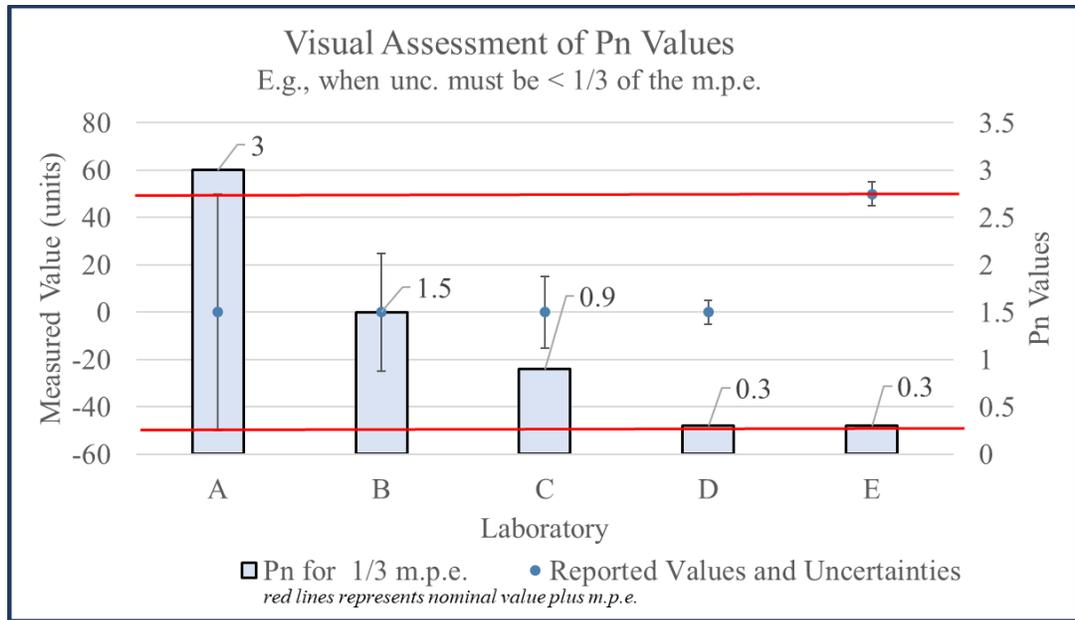


Figure 3. Visual Assessment of P_n Values.

Laboratory B fails the P_n assessment because the uncertainty is one-half of the tolerance instead of one-third. Laboratory C passes this assessment but is very nearly at the limit of 1 and may want to evaluate the uncertainty further.

3.3.4. Z Score

This statistical evaluation of Z Score comes from ISO/IEC 13528, 3.7 as: “*standardized measure of performance, calculated using the participant result, assigned value and the standard deviation for proficiency assessments*”. The Z-score may be used in combination with the adjustment statistics (trimmed mean and associated uncertainty) described earlier.

OWM reports this value in the tables of the final report, but does *not* use this statistic for pass/fail criteria in PTs because it uses PT statistics and does not consider the laboratory reported uncertainty in analysis. Satisfactory performance is generally indicated as $Z \leq 2$; unsatisfactory performance is indicated as $Z > 3$. and marginal performance is anything between $Z > 2$ and $Z \leq 3$. Further evaluation of the Z-score requires an assessment of the offset from the assigned reference value that includes the laboratory uncertainty, such as is provided by the E_n assessment. However, the directionality (positive or negative values) of this statistic can provide additional insights to the laboratory for ongoing evaluation of bias/offsets in measurement results.

$$Z - score = \frac{x_{lab} - X_{ref}}{s_{PT}} \quad \text{Eqn. (7)}$$

The Z-Score statistic and analysis is very similar to that of control charts where plus and minus two standard deviations serve as warning limits and plus and minus three standard deviations are the control or action limits. In the case of the PT, however, the standard deviation of the PT is based on the final statistics of the official values as any adjustments (if needed) have already been completed. In the graph shown in Figure 3, the Z values for each laboratory are given on the X-axis with the laboratory identification. It can be seen that the values are sequentially placed on one standard deviation intervals. Again, the assumption must be made that these laboratory values were selected for illustration purposes and the submitted values are all normally distributed around the assigned reference value.

The Bias (Difference) determined with Equation 1 is observed in these values and may impact which values are used in the selection of the assigned reference values, but further evaluation requires consideration of accuracy in conjunction with the E_n assessment, the reported uncertainty, and any applicable tolerances.

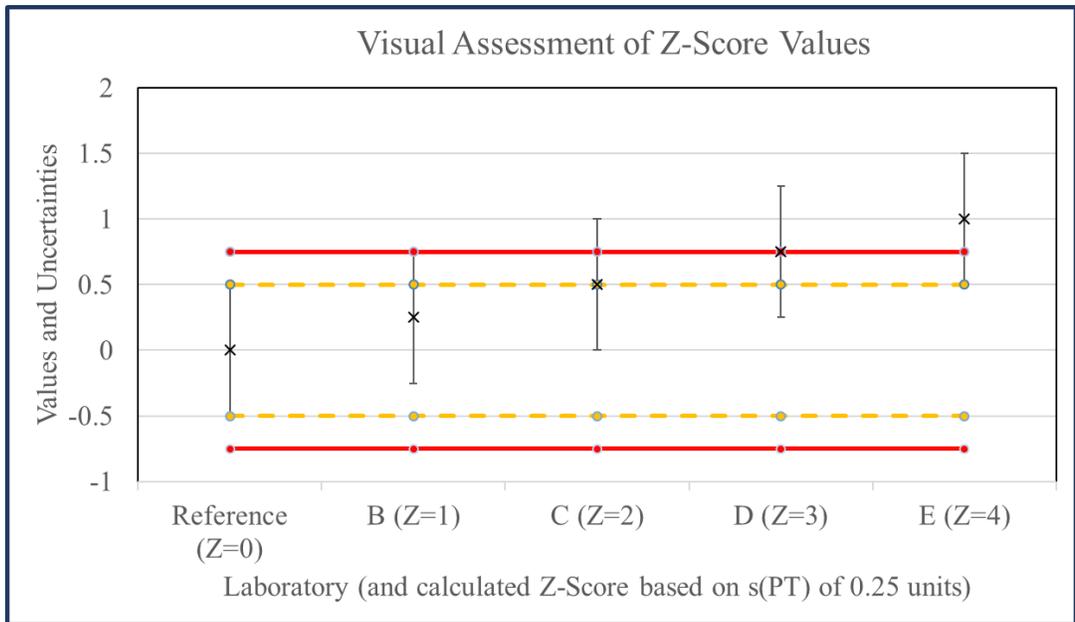


Figure 4. Visual Assessment of Z-Score Values.

4. Non-statistical Pass/Fail Criteria

Some PTs will have been planned and designed to assess laboratory participation for non-statistical evaluations. Additional non-statistical pass/fail criteria might include any or all of the following items that are explained in the PT report:

- Compliance of the certificate to ISO/IEC 17025, Section 7.8;
- Errors on submitted certificates and/or data sheets;

- [Unreasonable] time delays on standard/artifact shipments and/or report submission (e.g., communicating with the coordinator; reports within 2 weeks);
- Improper packaging and shipping (and handling);
- Deviations from the Approved/Accepted PT Plan (e.g., using a different SOP);
- Switching or substituting standards or artifacts, unapproved cleaning;
- Uncertainty Components: Uncertainty components specified in the SOP and Plan were not included; uncertainty reported smaller than what is on the Scope (for Accredited labs).

5. Follow-up Actions (Corrective, Improvement, Tracking)

Pass/fail status of each standard evaluated in the PT is not the only thing a laboratory should consider when participating in a PT. The OWM PT Follow-Up form is designed to enable a thorough follow up assessment that creates an Executive Summary that can be used in a Management Review as well as guiding the laboratory in performing a thorough assessment of the report, even when all indicators were successful. Ongoing tracking and evaluating of PTs is a part of ensuring the validity of measurement results provided by the laboratory and result assessment should be integrated into evaluating laboratory measurement assurance data from other sources such as periodic calibrations, internal evaluations of reference standards, similar past PTs, control charts, repeatability charts, and other statistical analyses. Regular assessment of PT data, even when successful, can mitigate risk and provide opportunities for continual laboratory improvement.

6. Proficiency Testing Follow-up Form

Instructions: complete one PT Follow-up Form per PT and per laboratory (do not complete one form for each staff member). Questions and directions are written to elicit descriptions, observations, and explanations and not Yes/No answers. This form may be used to summarize critical PT highlights that will be used in laboratory Management Reviews.

| | |
|---|--|
| Laboratory | |
| Date | |
| Completed By | |
| PT Measurement Parameter, Range, and Scope Description | |
| PT Identification (OWM Code) | |
| List of Participating Personnel (6.2) | |

| Assessment | Results and Evidence |
|---|-----------------------------|
| Executive Summary. Include summary and highlight the PT findings that can be used in the Management Review. (For example, total number of points, number passing/failing percentages, lessons learned, opportunities, corrective action already taken, etc.) | |
| PT Failures. Describe all laboratory failures that were identified in the final PT report (or additional failures or concerns identified outside the report). | |
| Analysis and Action Plan. Describe the analysis and investigation of Root Cause Analysis, Risk/Opportunities, Improvement Action, , and Corrective Action(s). (Section 8.5, 8.6, 8.7) | |
| Deadlines. List the deadlines for the completion of each action item and identify the personnel responsible for implementing and monitoring the results of each action. | |

| | |
|--|--|
| <p>Uncertainties (7.6., 7.8.6). Describe the uncertainty assessment. Questions to consider include: How did the reported uncertainty compare to other participating laboratory values? Was the correct k factor used? If the laboratory (or laboratory participant(s)) uncertainty value(s) were at the high end of the uncertainties, explain why. Could a better procedure or instrument have been used? If at the low end, was the value calculated correctly? Why is it smaller than the values reported by other laboratories? Explain if all appropriate uncertainty components were included (or why they were not included). Describe the planned measurement process and/or actual procedure used for the PT (higher/lower echelon procedure).</p> | |
| <p>Uncertainty versus Applicable Tolerances (7.8.6). Describe the Precision assessment results for this proficiency test. This assessment reviews the laboratory Uncertainty compared to the applicable Tolerances. Analysis questions to consider include: Was a precision test conducted as a part of the analysis? If yes, explain why there any ranges with unacceptable results. If no, conduct the precision analysis now. The calculation evaluates the reported uncertainty(expanded at $k=2$) against the tolerances required for the equipment with any uncertainty to tolerance ratios considered. Was the uncertainty reported acceptable/appropriate for the level of work? Could the uncertainty be improved with different equipment or procedures?</p> | |
| <p>Offset/Bias Assessment (7.7). Was bias observed in the PT also observed in other types of measurement assurance charts in the laboratory? Describe or summarize the bias and offset of the laboratory PT. Consider the following analysis questions: Was a Youden analysis or E_n analysis conducted? Were the reported values outside the Youden</p> | |

| | |
|--|--|
| <p>chart circle or was the E_n value greater than 1? If values were outside the circle or the E_n value is greater than 1, a measurement bias was indicated. If a bias was present, are there any overriding reasons for it? An investigation generally needs to be conducted looking for common errors and problems: e.g., apparent mass vs brass rather than conventional mass, incorrect values for the standard used, errors in software used for calculations, deviations from SOPs (using tap water for gravimetric calibration), need for calibration of standards. Conduct an investigation of bias (even if values passed the Youden and E_n analyses) against internal calibrations, control charts, PMAP charts, previous PT results or recent calibrations to find out if there is correlation of the ILC data with internal laboratory data. Was the bias for all laboratory participants similar? If not, describe why.</p> | |
| <p>Records (7.5, 7.8, 8.4). Describe the assessment in place to track PT data over time within the laboratory and evaluate the data against previous results and other data. Ensure that the final results were entered in laboratory PT log and identify the high-level summary data that will be included in the Management Review (8.9).</p> | |
| <p>Non-Measurement Result Observations or Failures. Describe any additional feedback related to the PT planning, scheduling, evaluation (e.g., delays) or reporting results (e.g., calibration certificate review, 7.8) that were provided as a part of this PT.</p> | |