PROFICIENCY TESTING FOR CALIBRATION LABORATORIES

Mr Philip Briggs

Proficiency Testing Australia, PO Box 7507, Silverwater, NSW, 2128 Australia

Abstract: In recent years proficiency testing has been used by laboratories to a greater extent to provide objective evidence of measurement capability. This paper details the different types of proficiency testing used by Proficiency Testing Australia (PTA) and explores the basic principles behind artefact selection and program design that are necessary in order to achieve an effective outcome. It also covers the analysis of results and the remedial and corrective actions necessary where problems are highlighted by a proficiency test.

Keywords: proficiency testing, measurement, uncertainty

1. INTRODUCTION

The competence of accredited laboratories is assessed by two complementary techniques. One technique is the on-site evaluation by a laboratory accreditation assessment team which examines the technical competence of the laboratories and their compliance with the requirements of ISO/IEC 17025 [1]. The other technique is by proficiency testing which involves the determination of laboratory testing performance by means of interlaboratory comparisons. The two techniques each have their own advantages and when combined give a high degree of confidence in the assessment of the overall competence of an accredited laboratory.

PTA operates its proficiency testing in accordance with ISO/IEC 17043 [2]. The proficiency testing activities are regularly evaluated as part of the proficiency testing provider accreditation process for the calibration programs.

ISO/IEC 17043 defines proficiency testing as *"evaluation of participant performance against pre-established criteria by means of interlaboratory comparisons"*. Also the document defines interlaboratory comparison as *"organisation, performance and evaluation of tests on the same or similar test items by two or more laboratories in accordance with predetermined conditions."*

2. ADVANTAGES OF PROFICIENCY TESTING

The advantages of proficiency testing are:
- for the accreditation body it gives a transparent demonstration of its laboratories technical competence;
- for the assessors it assist in making decisions on laboratory's proposed capabilities;
- for the laboratory the activity satisfies the ISO/IEC 17025 requirement for the monitoring of the quality of results;
- for the laboratory it can highlight otherwise undiscovered problems and can be used as a development and improvement tool.

3. AIMS OF PROFICIENCY TESTING

For calibration laboratories the main aim of proficiency testing is to test the participating laboratories ability to calibrate equipment within the accuracy for which they claim (i.e. their least uncertainty of measurement).

Secondary aims of proficiency testing are to ensure that laboratories are correctly:
- interpreting documentary standards;
- calculating results and uncertainties;
- reporting the results.

4. TYPES OF PROFICIENCY TESTING

There are two main types of proficiency test used by PTA:
- The interlaboratory comparison which involves the circulation of an artefact

sequentially around a group of laboratories;

- The measurement audit which involves the comparison of one participating laboratory with the reference laboratory.

The actual difference between the interlaboratory comparison and the measurement audit is not significant. In the majority of calibration interlaboratory comparisons, the laboratory's results are compared only to the reference results. So in effect the interlaboratory comparison is just a series of individual measurement audits.

Accredited laboratories in the Asia-Pacific region also may have the opportunity to participate in APLAC interlaboratory comparions coordinated by APLAC members. The purpose of APLAC is to build up and maintain mutual confidence between national calibration and testing services in order to reach and maintain a state of mutual agreement on the equivalence of the operation of the accreditation bodies and of the certificates issued by their accredited laboratories. This supports the removal of technical barriers to trade.

Calibration activities underpin all testing and measurement activities. APLAC calibration interlaboratory comparisons provide a forum for the comparability and traceability of measurements in the Asia-Pacific region and they are mandatory for APLAC members. They also provide confidence in the accreditation process of the APLAC members and also in their ability to take the appropriate corrective actions where the interlaboratory comparison reveals measurement deficiencies.

APLAC interlaboratory comparisons also result in a flow of know-how between the participating accreditation bodies and help establish a common high level of measuring capability within the Asia-Pacific region.

5. DESIGN STAGE

PTA's aim is to test if laboratories can achieve their claimed least uncertainties of measurement so the selected artefacts should be suitable for laboratories with differing accuracy levels. For laboratories with a higher claimed level of accuracy the artefact should have sufficient repeatability, reproducibility and resolution so that laboratories can report their least uncertainty of measurement.

For the laboratories with a lower level of accuracy it may be possible to introduce an error (departure from nominal value) as the artefact may be better than their own reference equipment and the laboratory may be able to calculate from the artefacts specifications, the maximum departure from nominal.

Larger departures from nominal values may be introduced by the means of:
- add-on error devices;
- manipulation of calibration factors enetered in the software;
- use of stable, but lower accuracy grade artefacts, e.g. Grade 2 instead of Grade 00 gauge blocks, F1 instead of E1 masses.

Note for the higher accuracy level laboratories, a larger departure from nominal would not normally make a difference to their uncertainty of measurement.

Generally, PTA wants to evaluate a laboratory's ability to achieve its claimed least uncertainty of measurement using their normal practice. An assessment is made of the quality of the work that the laboratory is routinely carrying out for its clients.

However, when evaluating high accuracy level laboratories using an instrument, the test method and specific settings may have to be well defined in order to minimize the number of variables. It is important that the artefact will operate in exactly the same way as when it was calibrated by the reference laboratory.

Calibration points are usually specified in order to directly compare them with the points calibrated by the reference laboratory. These will normally include:
- points where the laboratory may have its best uncertainty
- difficult ranges
- lowest point
- highest point
- points where the techniques or equipment may change

In areas such as dimensional metrology, where laboratories take measurements at certain points that characterize the whole item, it is necessary to ensure that there will be some common points (e.g. centre of a setting ring).

6. INSTRUCTIONS

The participating laboratory needs to be instructed that they must not adjust the instrument. The instructions will usually

include a description of the artefact and a listing of associated accessories and manuals and usually a pro-forma results sheet for efficient processing of results.

7. ARTEFACT STABILITY

To be a valid interlaboratory comparison it is essential that the artefact does not significantly change its measured characteristic from the time the reference laboratory calibrates to when each participating laboratory undertakes the measurements.

In a measurement audit if there is any doubt about the artefact's stability then a check is usually carried out at the point in question.

In an interlaboratory comparison the reported results give confidence in the artefacts stability. Other strategies include:

- for sensitive devices a number of reference laboratory calibrations are carried out throughout the program. Successive reference values are then compared to monitor the artefact's stability. If there is any doubt the artefact is recalibrated;
- checks by the laboratory may be specified to determine if the artefact has been damaged or overloaded;
- checks by a pilot laboratory (not necessarily the reference laboratory) where only a few measurements are required to ensure the artefact's stability after each laboratory;
- scheduling of high accuracy level laboratories close to the reference laboratory calibrations;
- where possible, use of an artefact that has a calibration history or model/type of artefact whose stability has been proven previously.

In the analysis the mean of the reference laboratory calibration results would normally be used. However a secondary analysis may also be conducted against the reference results and where problems are evident the best-case results are used.

A review of all results is carried out. If participant's results are distributed fairly evenly above and below the reference values and/or there is good agreement with the higher level laboratories then this gives added confidence in the stability of the artefact and the reference values.

8. HANDLING OF ARTEFACTS

Stable artefacts with calibration histories can be worth considerably more than their replacement price because of the time and expense it can take to re-establish a calibration history.

Laboratories are instructed to report damage to the artefact or the packaging. There is also a need to ensure that the artefact is also adequately located within its protective container and that it cannot shake or vibrate during transport.

9. EVALUATION OF RESULTS

The calibration interlaboratory comparison programs are not trying to determine which laboratory can report results closest to the reference laboratory values. As stated earlier, the aim is to determine if the participant can achieve their claimed least uncertainties of measurement.

ISO 13528 [3] defines the internationally accepted measure for evaluation of participant performance is the En ratio. En stands for Error normalized and is defined by equation (1):

$$E_n = \frac{LAB - REF}{\sqrt{U^2_{LAB} + U^2_{REF}}}$$
(1)

LAB is the participating laboratory's result
REF is the reference laboratory's result
U_{LAB} is the participating laboratory's reported uncertainty
U_{REF} is the reference laboratory's reported uncertainty

For E_n numbers:

$|E_n| \leq 1.0$ = satisfactory

$|E_n| > 1.0$ = unsatisfactory

(The closer to zero the better.)

In calibration interlaboratory comparisons the E_n numbers indicate whether laboratories are within their particular uncertainty of measurement of the reference value (assigned value).

The E_n numbers do not necessarily indicate which laboratory's result is closest to the reference value. Consequently, calibration laboratories reporting small uncertainties may

have a similar E_n number to laboratories working to a much lower level of accuracy (i.e. larger uncertainties).

In a series of similar measurements a normal distribution of E_n ratios would be expected. So when considering the significance of any results with $|E_n|$ marginally greater than 1, all the results from that laboratory are evaluated to see if there is a systematic bias e.g. consistently positive or consistently negative values of E_n.

A sample table of results from a Mass interrerlaboratory comparison, their corresponding reported uncertainties and E_n ratios are tabulated below. The result for laboratory 5 is considered unsatisfactory.

Lab Code	Results	U_{Lab}	En
Ref	1.000177	0.000008	-
1	1.000162	0.0000499	-0.30
2	1.000174	0.000006	-0.30
3	1.000176	0.00001	-0.08
4	1.000179	0.000015	0.12
5	1.000245	0.000023	**2.79**
6	1.000174	0.000007	-0.28

Table 1. Interlaboratory results

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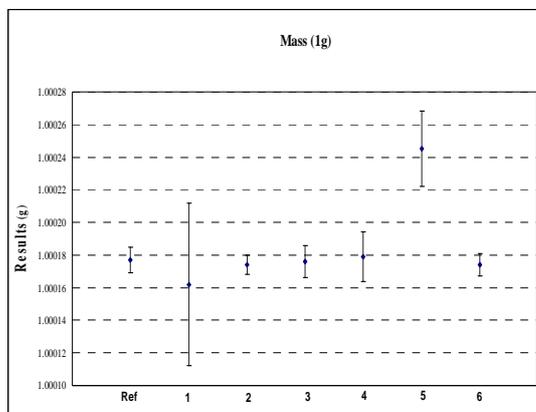

Figure 1. Graph of results

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

Applied Pressure	Ref value MPa	Ref U_{95} MPa	Lab Mean MPa	Lab U_{95} MPa	En Ratio
5.0000	4.8983	0.0014	4.8982	0.002	-0.03
7.5000	7.3478	0.0014	7.3466	0.002	-0.46
9.9999	9.7973	0.0019	9.7970	0.004	-0.08
7.5000	7.3605	0.0031	7.3462	0.002	**-3.88**
5.0000	4.9074	0.0025	4.8971	0.002	**-3.51**

Table 2. Measurement audit results

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

10. ACTIONS FOR UNSATISFACTORY PERFORMANCE

Where there are results with $|E_n| > 1.0$ the laboratory should investigate. Where a laboratory has not performed satisfactorily, the participant should ensure that appropriate action is taken to:
- correct the problem;
- review other works that could have been affected by the same problem and where necessary recall reports;
- take action to prevent recurrence of the problem.

Accreditation bodies may work with the laboratory to identify and overcome the problem:
- where there are results with $|E_n| > 1.0$ the laboratory may be asked to investigate (if accredited for the specific test) and report details of their findings and corrective action;
- seek advice where necessary from the appointed program technical adviser for a PTA response;
- discuss the matter further at the laboratories next scheduled reassessment.

11. CONCLUSIONS

Proficiency testing provides objective evidence of calibration laboratories' competence and complements the laboratory accreditation process. The effectiveness of such proficiency tests requires careful artefact selection, program design, evaluation of results and

where necessary, appropriate follow-up action to be taken by laboratories reporting unsatisfactory results.

12. REFERENCES

[1]   ISO/IEC 17025: 2005 General requirements for the competence of testing and calibration laboratories

[2]   ISO/IEC 17043:2010 Conformity assessment – General requirements for proficiency testing

[3]   ISO 13528: 2005 Statistical methods for use in proficiency testing by interlaboratory comparisons