USE OF INTERLABORATORY COMPARISON DATA BY LABORATORIES

Relevance of interlaboratory comparisons

Interlaboratory comparisons (ILCs) are performed for various reasons [1], e.g.

- to validate test methods,
- to certify reference materials,
- to assess the performance of laboratories (proficiency testing),
- or more general, to investigate the degree of equivalence among laboratories.

Irrespective of the specific aim(s) of an ILC, the results can be used by a participating laboratory

- to check the performance of its test methods and / or its staff,
- to demonstrate its competence towards clients and accreditation bodies,
- to gain useful information for the evaluation of its measurement uncertainty.

The scoring of ILC data

In proficiency tests (PTs) the PT providers often evaluate scores as a quantitative measure of the laboratory performance. There is a number of different scores, two of which are used most commonly [2]:

z-score:

$$z_i = \frac{\left(x_i - x_{pt}\right)}{\sigma_{pt}}$$

E_n number (mainly used in calibration):

$$\left(E_n\right)_i = \frac{x_i - x_{pt}}{\sqrt{U^2\left(x_i\right) + U^2\left(x_{pt}\right)}}$$

(with x_i: result of laboratory (i), x_{pt}: assigned value, σ_{pt}: standard deviation of proficiency assessment, $U(x_{pt})$:
expanded uncertainty of the assigned value, $U(x_i)$: expanded uncertainty of the laboratory (i) result)

The numerator of both scores gives the difference between the laboratory result and the assigned value, which can either be established by one or more reference laboratories (usual practice in calibration) or be derived as consensus value from the group of participating laboratories. The standard deviation σ_{pt} in the denominator of the z-score is a measure of the actual or accepted variability of the results. The denominator of the E_n number represents the expanded combined uncertainty associated with the difference in the numerator. Thus, both scores have in common that the actual difference between the laboratory result and the assigned value is assessed against an estimate of the (expected or acceptable) spread of results. The two scores differ in that the z-score assesses all laboratories against the same numerical value, while the E_n number allows the individually claimed accuracy of a laboratory to be taken into account.

PT providers often use the following classification for the result of a participating laboratory:

satisfactory result: $|z| \le 2$ or $|E_n| \le 1$respectively,
questionable result: $2 < |z| < 3$,

unsatisfactory result: $|z| \geq 3$ or $|E_n| > 1$ respectively[1].

The analysis of the ILC data by the laboratory

In order to use the result of an ILC for the purposes mentioned in clause 1, the laboratory, after participating in an ILC, should carefully analyse its result, taking into account existing information, such as

- statements on the measurement uncertainty of the test method used in standards, literature etc.,
- their own evaluation of this measurement uncertainty,
- standard deviation of the results of all laboratories participating in this ILC,
- measurement uncertainty acceptable for the laboratory and its clients.

Even if the organiser of an ILC provides a classification of the results as satisfactory or unsatisfactory the laboratory should not simply rely on this judgement. If e.g. the organiser uses a standard deviation σ_{pt} for the z-score, which the laboratory considers to be not fit-for-purpose (ffp), it might calculate a modified z-score using an $\sigma_{pt,ffp}$ according to its own or its clients' needs [3].

In the case of an unsatisfactory result, the laboratory should perform a root cause analysis and, based on its result, should take corrective actions. Sometimes the organiser of the ILC might offer advice. After implementation of the corrective actions, the laboratory should prove their effectiveness, e.g. by

- use of a suitable reference material,
- participation in another ILC.

Furthermore, the results of ILCs are an important tool for verifying the evaluation of the measurement uncertainty of the used test methods [4, 5, 6]. If the laboratory's estimates of the measurement uncertainty turn out to be too conservative or too optimistic, the laboratory should adapt them accordingly.

Conclusions

Irrespective of an eventual classification of the ILC results by the organiser of an ILC as satisfactory or unsatisfactory, each participating laboratory should carefully analyse its results on the basis of its own criteria. If a result then turns out to be unsatisfactory, the laboratory should take appropriate corrective actions and verify that these actions have been effective.

Additionally, the results of an ILC should be used to verify or improve the estimates of measurement uncertainty of the test methods used.

References

[1] ISO/IEC 17043:2010, Conformity assessment – General requirements for proficiency testing
[2] ISO 13528:2015, Statistical methods for use in proficiency testing by interlaboratory comparisons
[3] IUPAC, The international harmonized protocol for the proficiency testing of analytical chemistry laboratories, Pure Appl. Chem., 78 (2006), 145 - 196
[4] NORDTEST Technical Report 537 ed.3.1 (2012), Handbook for Calculation of Measurement Uncertainty in Environmental Laboratories, 2003, http://www.nordtest.info/
[5] EUROLAB Technical Report 1/2006, Guide to the Evaluation of Measurement Uncertainty for Quantitative Results, www.eurolab.org

---

[1] When the expanded uncertainties are calculated using a coverage factor of 2.0, a critical value of 1.0 for an E_n number is equivalent to the critical value of 2.0 used in z-scores [2].

[6] EUROLAB Technical Report 1/2007, Measurement uncertainty revisited: Alternative approaches to uncertainty evaluation, www.eurolab.org