

# ISO/IEC 17025:2017 ENSURING THE VALIDITY OF RESULTS – ARE THESE NEW REQUIREMENTS?

---

Trevor Thompson of [www.bestmeasurement.com](http://www.bestmeasurement.com) offers advice about the use of Interlaboratory Comparison and Proficiency Testing activities as required by ISO/IEC 17025:2017

Previous versions of ISO/IEC 17025 used different terminology but the concept that has existed for many years remains. If you consider that compliance arises from

- a. Technical Competence – the right people and equipment
- b. Consistency provided by a reliable management system and
- c. Actually getting a valid result demonstrably in practice

then one needs to consider how to achieve the demonstration of validity that c) above requires.

In the latest version of the standard, the use of the term “quality” has been replaced in several places by “validity”. This better reflects that measured values and test results all come with an uncertainty of measurement and if correct, the result is valid. This avoids confusing smaller and larger uncertainties with high and low quality of work.

A further enhancement in this version of the Standard is to include an extensive list of possible activities that might be undertaken to ensure the validity of results. All laboratories should expect to find something appropriate and attractive to them in the list. This activity shall be undertaken, where appropriate, but not be limited to the items in Clause 7.7.1:

- a. use of reference materials or quality control materials;
- b. use of alternative instrumentation that has been calibrated to provide traceable results;
- c. functional check(s) of measuring and testing equipment;
- d. use of check or working standards with control charts, where applicable;
- e. intermediate checks on measuring equipment;
- f. replicate tests or calibrations using the same or different methods;
- g. retesting or recalibration of retained items;
- h. correlation of results for different characteristics of an item;
- i. review of reported results;
- j. intralaboratory comparisons;
- k. testing of blind samples

It is not expected that all laboratories undertake all these activities. The list includes items particular to various types of laboratory activity and the laboratory should choose items that are relevant to them and which would bring the most useful information in their circumstances. Some of the suggestions overlap with each other in their application and usefulness. Most of these possibilities would usually be considered “internal” to the laboratory or to a group of laboratories having common features.

There are distinct weaknesses arising from purely internal activity. If there was deficient traceability of measurement in a single external source, this might not be seen. Furthermore, any deficiency in the documented procedure for doing a test or calibration, followed by everyone would not be seen and the same might apply to bad training practices.

## **Something external is essential!**

Hence we have Clause 7.7.2 which clearly requires some external involvement. One sometimes hears of laboratories who undertake some unique test or calibration for which there is no comparable other laboratory. I would not doubt that this may happen, but the trick then is to pick parts of the test or range, maybe looking at some inputs rather than at the final combined output.

The Standard splits this into “Proficiency Testing” which is taken to be participation in a scheme run for that purpose and “participation in interlaboratory comparisons other than proficiency testing” which often entails making your own arrangements with other laboratories.

Let’s now consider some of the possibilities in more detail:

**Use of reference materials.** The use of any stable item, similar to the items to be tested but considered especially stable, may have a well-deserved place in a laboratory’s repertoire of activities. If it does not have a value ascribed to it, then it is still useful to use, maybe just after equipment is calibrated, and retain it for future use at intervals, or when malfunction is suspected. This technique is often relatively cheap to implement but remember that it will not show systematic components of uncertainty arising from an external source, or defective procedures, unless it is a certified reference material with a known value. **Retesting of stable retained items cited in g) is a similar technique.**

**Use of alternative instruments.** If one has alternative and equal instruments both having external calibrations, then this technique is excellent because it will show faulty equipment. If the calibrations come from different sources all the better. If the lab has a bad technique always used then this will not help. Sometimes it is possible to lengthen and stagger the external calibrations by undertaking internal comparisons between the duplicated instruments. This also helps with removing downtime for calibration, but for some laboratories with very expensive measuring equipment or environments it may be prohibitively expensive and other ideas might be more appealing.

**Replicate or Repeat Tests** may be applied in a variety of useful situations. This may be to compare operators where operator skill is central to the validity. It may be used also to compare equipment and environments. The technique may also be used to establish or check random components of uncertainty under static conditions and to investigate variables by constraining all but one and swinging that one between set extremes.

**Blind Testing.** This has a slightly different meaning across industries but can be taken to be when an operator or even a whole lab does not know the source or sometimes the identity of the test item. Difficult to arrange, the technique is good for spotting corruption, conscious or unconscious bias and the taking of shortcuts. Aren’t we all aware of taking greater care if we know the test is for someone rather special in some way?

**Proficiency Testing Schemes** are common in many testing areas, especially those involving operator acumen, eyesight or other skills and where a feature present or not may be involved. It is less commonly available for calibration, especially in the UK. In an ideal PT Scheme there would be a higher order laboratory with a reference value and artifact(s) are circulated between participants or sent to each from a known batch. There is a comprehensive directory of such schemes at <http://WWW.EPTIS.ORG> which is maintained by the most respected German Laboratory BAM in Berlin. Some PT Schemes are accredited to ISO/IEC 17043 and these would be preferable. PT Schemes vary in their composition and in the analysis of results. Some have clear reference values, some may use participant consensus values and may be affected by the performance of other participants. Laboratories should strive to find a suitable PT Scheme but may sometimes find, although worthwhile, that the application is not an exact fit for their work in terms of ranges and materials offered.

**Interlaboratory Comparisons** (other than PT) may just be something arranged privately between two or more laboratories or may be a scheme typically for calibration and run, for example, by a National Measurement Institute.

**Comparing the results from ILCs and PT** is usually done by using the relatively simple use of a formula known as En Ratio, Zed or Zeta Score. These are all similar but be aware that some testing schemes do not handle uncertainty of measurement in the complete strict sense that is expected in calibration labs. This formula may be used to compare any two measurements to see if, taking the Uncertainty of Measurement into account, the two values are potentially valid and come from the same population. A En ratio of less than 1 is valid. Some PT Schemes expect Zed or Zeta to be less than 2 or 3 for a satisfactory result.

$$E_N \text{ ratio} = \frac{LV - RV}{\sqrt{U_{LV}^2 + U_{RV}^2}}$$

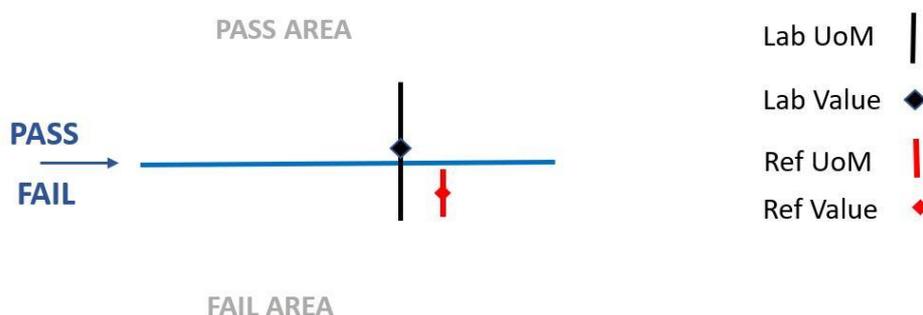
LV is the first laboratory value for any single measurement point

RV is the same measurement point taken at another laboratory (may be a “better” reference laboratory)

ULV and URV are the associated Uncertainties of Measurement for the values LV and RV.

Simply taking the difference between the two values and dividing by the square root of the sum of the squares of the two uncertainties gives a measure of the validity of the result LV relative to RV and is therefore very effective for the LV lab if the RV lab has a lower UoM. It can be used to compare similar laboratories but is a less critical measure in that case.

To see this graphically consider the diagram below where two laboratories with different UoM measure the same item. You can see the possible range of values for each lab and how they are compatible and potentially come from the same population, in this case because all possible values from both labs fall within the uncertainty ULV of the value LV. One may conclude therefore that both labs achieved a valid result.



What if a Pass/Fail mark lies between the two lab values? In that case one lab considers the product passes, the other fails. Both may be valid results, but that topic (Decision Rules, taking uncertainty into account) is a subject for another day!

Trevor Thompson who has enjoyed a long career in Metrology and Accreditation now offers training and consultancy in his semi-retirement at [www.bestmeasurement.com](http://www.bestmeasurement.com). He was the British BSI representative for the writing of the current version of ISO 17025 at ISO and a member of the Drafting Group. He welcomes questions from members at [questions@bestmeasurement.com](mailto:questions@bestmeasurement.com)