

Exploiting interlaboratory comparisons

Hendrik Emons

Published online: 13 July 2013
© Springer-Verlag Berlin Heidelberg 2013

Many individuals and laboratory teams, respectively, experience some kind of surprise when they are starting to compare their own measurement results with the ones from other laboratories. It could be a positive astonishment, namely how well the results are matching for a given measurand and sample. More often the contrary is observed and the origin of the discrepancies has to be identified.

However, the potential of an interlaboratory comparison (ILC) including its evaluation depends strongly on the ILC design and the amount of information provided by the participants. This is not different from any other scientific study. But there are still several myths about the ‘do and don’t’ for ILCs. For instance, opinions such as ‘never combine measurement results from an ILC to assign a property value to a reference material’ or ‘increase the number of ILC participants for better approaching the true value’ are voiced at conferences or in committee meetings.

Nevertheless, recent years have seen significant progress in the common understanding of underlying scientific principles and concepts of ILCs as well as in their design and execution. This was and continues to be driven by the needs to demonstrate the analytical capabilities (proficiency) of a set of laboratories, to identify the performance (limits) of a particular measurement method (actually a ‘measurement procedure’) or to characterize a specific property of the material under investigation. In the past, such judgments were left to some individuals or institutions, often based on a non-transparent division into ‘experts’ and ‘non-experts.’ It is interesting to note that the enhanced globalization of science, industry, trade, people’s mobility, communication, etc., has also put the traditional

way of defining experts and competences in question — at least partially. Nowadays, it is usually not sufficient anymore to be identified by a well-known scientific ‘heavy weight’ or to belong to a traditionally well-reputed institution. One has regularly to provide evidence for the claimed competence. ILCs can be a relatively independent route for demonstrating specific competencies in measurement tasks. This is increasingly recognized also by regulators. For instance, official control laboratories for food or environmental monitoring in the European Union are required by legislation to successfully participate in dedicated ILCs.

A prerequisite for the wider acceptance of ILC as a quality assurance tool was the combination of metrological principles for ILC designs with dedicated elements of standardization and internationally harmonized surveillance approaches via accreditation. This ranges from demanding participation in proficiency testing of laboratories which wish to obtain and keep an accreditation according to ISO/IEC 17025, to ensuring quality criteria for proficiency testing via application of ISO/IEC 17043. The latter document also contains a list of different goals which may be targeted by ILCs. In this respect, it is important to consider the interrelation between the three main components of a measurement exercise: the material (sample under investigation), the method(s) (measurement procedures used including all sample preparation and manipulation steps), and the laboratories participating. One cannot separately assess more than one of these components in the same ILC. For instance, in the framework of the characterization study of a new candidate reference material, the applied measurement methods have to be validated before applying them by laboratories of known (proven) competence. Measurement results obtained in such designed ILCs can be used for assigning property

H. Emons (✉)
Geel, Belgium
e-mail: JRC-IRMM-ACQUAL@ec.europa.eu

values to reference materials. However, one should not aim to test the proficiency of the participating laboratories within the same exercise!

Proficiency testing (PT) of laboratories, which forms a special subgroup among the ILCs, requires well-characterized test materials, and it is desirable to have PT materials which behave in the analytical process (including sample preparation) close to the routinely measured test samples. But it is essential that the PT samples are sufficiently homogeneous and stable within the period of the ILC. Consequently, PT materials should be qualified as reference materials, and the related ISO Guides provide helpful advice for their preparation and handling.

Now let us consider the second opinion from above, namely whether increased participant numbers in ILCs increases the chance of an assigned value approaching the true value. It is a frequently proven fact that science is not a democracy. Just inflating the number of ILC participants does not increase the reliability of a property value calculated from the submitted data. Therefore, using a ‘consensus value’ (the mean or robust mean) from results

of all PT participants as scientific basis for the ILC evaluation can be questionable. This holds especially true for more demanding measurement tasks, for instance determining the mass fraction of brominated flame retardants in contaminated sediments. Instead of pooling as many data as available, a fair and scientifically thorough ILC evaluation should make use of measurement results on the PT material which constitute sound quality benchmarks. In practice, such so-called reference values are obtained by one or a few highly skilled laboratories applying measurement methods of appropriate metrological level.

Overall, establishing the equivalence of measurement capabilities among trading partners or among globally distributed providers of crucial services such as healthcare diagnostics is a demanding task. *Accreditation and Quality Assurance* offers a forum for reporting on related concepts and new experiences. I am looking forward to reading even more submissions on this topic.

Hendrik Emons
Editor-in-Chief