A reference value in a proficiency testing scheme should be as metrologically credible as possible

Paul De Bièvre

Published online: 31 August 2012
© Springer-Verlag 2012

To evaluate the performance of participants in a proficiency testing scheme (PTS), an average, modified average, or median of the participants' 'measurement results' (entry 2.9 in [1]) is often chosen as an assigned value which is then used as 'reference value' (entry 5.18 in [1]) with standard deviation, against which participants' performance is evaluated. Both are needed to serve as a measure of the spread of the data, possibly leading to identifying reasons of large deviations of some participants' results from others. The standard deviation mentioned is thereby considered as a sort of *measure of the reliability* of the mean of the participants' results. Frequently, outlier criteria are applied, reducing said standard deviation, although, normally, a fairly large number of data points with normal distribution must be obtained to apply criteria for removing outliers. Proof of this distribution is missing most of the time, whereas *deviation from normality* has been shown to be often the case. For well-documented cases see, for example, [2] and numerous graphs in the ongoing IMEP (International Measurement Evaluation Programme, 1986-to date) of the Institute for Reference Materials and Measurements (IRMM) [3]. Despite this counter-evidence, the data set from a PTS is frequently assumed to have a normal distribution. If any criterion for removing outliers is thereby used, the spread of the results may become dependent from that criterion. One can say that the resulting average (with associated standard deviation) becomes "variable" depending on the chosen criterion and

the particular mix of laboratories and measurement procedures in the PTS concerned. Conclusions from the PTS are therefore open to some arbitrariness of a "variable" reference value. Imagine participating in an archery contest where the location of the bull's eye varies from one round to the other and the archer does not know exactly where that is when the arrow is shot. The underlying thinking seems to be the belief that the greater the number of measurement results, the more confidence one can have in their average—which is true if the main concern is the extent to which that value would change in a repeated study using the same materials and participating laboratories. However, there is a risk to that: the approach is only valid when the participants' measurement results are not badly non-normally distributed. Frequently, normal distribution is just assumed and rarely accompanied by testing whether the data set concerned is indeed normally distributed. Visual review can confirm normality quite easily.

In a different view, the results from a PTS are used "to document participants' measurement performance" (Table 7.2-1 in [4]). Now, the choice of the reference value has a direct impact on conclusions about the 'measurement capability' (concept 7.3-1 in [4]) of the participating laboratories, and therefore, the most metrologically credible value should be sought. Instead of taking the average of the participants' results, the value obtained by one or several external laboratories with demonstrated metrological competence can be taken to serve as 'metrological reference' (concept 2.6-1 in [4]).

How can this approach be organized?

1. The organizer's laboratory can be deemed to be an appropriate laboratory, if it is in a position to display documented metrological traceability for its 'reference value' with associated measurement uncertainty (as

Disclaimer The opinions expressed in this Column do not necessarily represent the view of ACQUAL.

P. De Bièvre (✉)
Kasterlee, Belgium
e-mail: paul.de.bievre@skynet.be

described more in detail under point 2 below), at the same time meeting a pre-set target measurement uncertainty (entry 2.34 in [1]) for the intended use [5], against which each of the participant's measured values is checked when drafting the conclusions from the PTS. A rather comfortable situation exists when the reference value has an expanded uncertainty (as described in the Guide on the expression of uncertainty in measurement (GUM) [6]), which is considerably smaller than the expected spread of the participants' results. These characteristics provide the organizer's laboratory with a metrological authority, satisfactory for its measurement result to be used as a reference value for the PTS. Good examples in PTSs are the cases in which an independent 'primary measurement standard' (entry 5.4 in [1]) is made and the thus created embodied value is used as reference value in the PTS.

2. Another, maybe more independent, approach seems to be that a laboratory (or more than one laboratory—see ISO Guide 35) *external* to the entire PTS structure is selected. The recently released standard ISO/IEC 17043 provides the possibility for that [7]. However, such a laboratory must be able to show documented metrological traceability of its results to a valid reference, either agreed within the professional field concerned or accepted nationally or internationally. The result must be accompanied by an associated measurement uncertainty, which should be compatible with, or smaller than, a target measurement uncertainty, set a priori as measure of the fitness-for-intended-use of the participants' measurement capability. Such a laboratory can find the formally established references for 'metrological traceability' in the VIM (entry 2.41 Note 1 in [1]):
 - (a) a measurement unit (entry 1.9 in [1]),
 - (b) a reference measurement procedure (entry 2.7 in [1]), or
 - (c) a measurement standard (entry 5.1 in [1]).

The latter was described above. The case (b) will be discussed at another occasion.

It cannot be stressed enough that an independent and as metrologically traceable as possible reference value is important: participants expect such a reference, that is, a value against which their performance (their 'measurement capability') will be evaluated, to be established and documented in an undisputable, objective way.

Non-metrologically traceable references should be avoided in a PTS. That would suppress any kind of (suspicion of) arbitrariness in the assessment of the measurement capability of the PTS participants.

As usual, any comment, question, or amendment is welcome, preferably as a contribution to the Discussion Forum of this Journal.

References

1. BIPM, IEC, IFCC, ILAC, IUPAC, IUPAP, ISO, OIML, The international vocabulary of metrology—Basic and general concepts and associated terms (VIM), edn 3, JCGM 200:2012 at <http://www.bipm.org/vim>
2. De Bièvre P (2009) Comparing results from chemical measurements: some questions from practice, chapter 8. In: Pavese F, Forbes AB (eds) Data modeling for metrology and testing in measurement science. pp 255–273. ISBN 978-0-8176-4804-6 (© Birkhauser Boston, a part of Springer Science + Business Media, LLC)
3. http://www.irmm.jrc.be/interlaboratory_comparisons/imep/Pages/index.aspx
4. De Bièvre P, Dybkaer R, Fajgelj A, Hibbert B (2011) Metrological traceability of measurement results in chemistry—concepts and implementation. Pure Appl Chem 83:1873–1935 at <http://iupac.org/publications/pac/83/10/1873>
5. De Bièvre P (2010) Fitness for intended use is an important concept in measurement. Accred Qual Assur 15:545–546
6. BIPM, IEC, IFCC, IUPAC, IUPAP, ISO, OIML, Guide for the expression of uncertainty in measurement (GUM), JCGM 100:2008 at www.bipm.org/en/publications/guides/gum.html
7. ISO/IEC Standard 17043:2010, Conformity assessment—General requirements for proficiency testing, International organization for standardization, ISO, Geneva