amc technical brief

Analytical Methods Committee No. 2. Nov 2000

How to combine proficiency test results with your own uncertainty estimate - the zeta score

Background

Proficiency testing is a method for regularly assessing the accuracy of laboratories in conducting particular measurements. In analytical chemistry, proficiency testing usually comprises the distribution of effectively identical portions of the test material to each participant for analysis as an unknown. The laboratories conduct the test under routine conditions, and report the result to the organiser by a deadline. The organiser then converts the result to a score which helps the participant assess the accuracy of the result in relation to a fitness for purpose criterion.

The primary purpose of the proficiency test (PT) is to allow the participants to confirm that they are complying with the external criterion or, failing that, to detect unexpected errors in their results. Unexpected errors should trigger an investigation of causes of the problem and, if necessary, remedial activity. Proficiency tests have also acquired secondary purposes beyond the original self-help ethos. Accreditation agencies usually require that candidate laboratories (a) participate in appropriate proficiency tests where available, (b) perform satisfactorily overall, and (c) have a procedure for investigating exceptional errors when they occur. Moreover, laboratories are increasingly using PT results to demonstrate competence in their bids for contract analytical work.

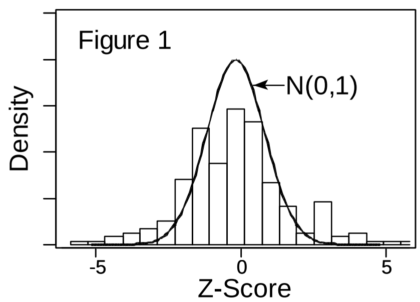
Scoring systems

Most proficiency testing schemes in analytical chemistry use the scoring system recommended in the Harmonised Protocol.[1] In this system, the participant's result x is converted into a 'z-score' given by the equation:

$$z = (x - x_{ass}) / s_p ,$$

where x_{ass} is the assigned value, the organiser's best estimate of the true value, and s_p is the so-called 'target value of standard deviation'. $(x - x_{ass})$ is the estimate of the error in the result, and z is the same error scaled in a particular way.

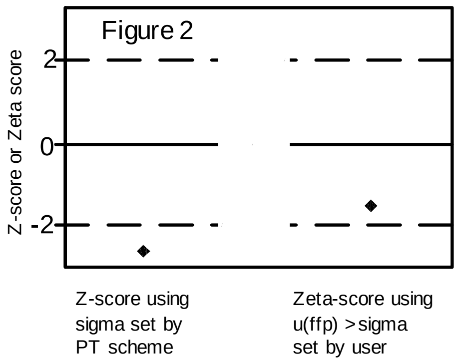
In an ideal PT scheme, the value given to s_p is determined by fitness for purpose: it represents the amount of uncertainty in the result that is tolerable in relation to the purpose of the data. Notice that here s_p describes the end-user's requirements, not the data. It is then up to the organiser to put arbitrary limits of acceptability on the value of z. If the participants as a whole complied with the criterion, but no better, we would expect z to be roughly like a random normal variable with a mean of zero and a standard deviation of unity [N(0,1)]. That is why many organisers regard a value of z falling between ±2 as indicating satisfactory performance. Fig. 1 shows an example of overall non-compliant data.



If you make s_p describe the *data* (rather than the requirements), for example by setting the value to the robust standard deviation of the participants' results, the z-scores will have a unit standard deviation and around 95% will fall into the 'satisfactory' category, irrespective of any fitness-for-purpose considerations.

Do-it-yourself scoring

A problem sometimes encountered by participants is that their customer's fitness-for purpose criterion is different from that of the PT scheme that they are using. This can easily happen: the PT scheme sets its criterion for the sector of analysis in general, while the participant deals with specialised applications. A hypothetical outcome is shown in Fig 2.



The participant gets a poor z-score in the PT scheme, but would do better if the target value were compatible with the customers requirements. Moreover, analysts are being encouraged to estimate the uncertainty of their results[2] and are beginning to wonder whether such information can be incorporated into proficiency test scoring. These possibilities were first envisaged by the AMC.[3,4]

The recommended course of action is for the participant to calculate an auxiliary score called the 'zeta-score', given by:

$$z = (x - x_{ass}) / u_{ffp} .$$

In this equation the denominator is u_{ffp}, which is the participant's or the end-user's definition of an uncertainty that is fit for purpose. The laboratory should take the PT scheme's assigned value for the calculation. The zeta-score is therefore a customised z-score that applies to the participant's individual circumstances. For accreditation or contractual purposes, the participant can list the zeta scores obtained and show the u_{ffp} values on which they are based. The value of u_{ffp} would have to be demonstrably justifiable, but may vary with the concentration of the analyte.

ISO Guide 43 suggests a similar strategy,[5] presumably to be implemented by the PT Scheme itself. This idea is formulated in terms of an 'En' number given by:

$$En = (x - x_{ass}) \Big/ \sqrt{u_{ass}^2 + u_x^2} .$$

However, this idea as it stands calls for two caveats. First, the formula does not refer to fitness for purpose because it uses u_x (the uncertainty of the laboratory's result) rather than u_{ffp}. Second, the inclusion of the term u_{ass} (the uncertainty of the assigned value) is technically correct, but can lull the user into a false sense of security. Essentially, if u_{ass} is big enough to matter in the equation, it is big enough to make us question the validity of the proficiency test. Some proficiency test protocols already include the constraint that u_{ass} should be negligible. The use of the En number is therefore not recommended here. (Note that ISO Guide 43 was drafted to be completely general: not all of the methods mentioned are necessarily suitable for analytical chemistry.)

It would be difficult for the PT organisers to carry out zeta-score calculations based on submitted uncertainties. The organisers would have no control over whether the uncertainties were appropriate, and would be unable to attribute any meaning to the scores based on them. In addition, participants might have several different fitness-for-purpose criteria for different customers, each of which could generate a different score. It is therefore more appropriate for the individual participants to calculate their own zeta-scores, in consultation with their customers.

References

1 M Thompson, R Wood, *Pure Appl. Chem.*, 1993, **65**, 2123.
2 *Quantifying Uncertainty in Analytical Measurement*, www.measurementuncertainty.org
3 AMC, *Analyst*, 1995,**120**, 2303.
4 R E Lawn, M Thompson, R F Walker, *Proficiency Testing in Analytical Chemistry*, Royal Society of Chemistry, Cambridge, 1997.
5 ISO Guide 43, *Proficiency Testing by Interlaboratory Comparisons*, ISO, Geneva, 1997.

AMC Technical Briefs (Editor M Thompson) are informal but authoritative bulletins on technical matters of interest to the analytical community. They are prepared by the Analytical Methods Committee of the Analytical Division of the RSC, and are carefully scrutinised for accuracy. Correspondence should be addressed to: The Analytical Methods Committee, The Royal Society of Chemistry, Burlington House, Piccadilly, London W1V 0BN. Other AMC Technical Briefs can be found on **www.rsc.org/lap/rsccom/amc**