

AMC TECHNICAL BRIEFS

CrossMark
click for updatesCite this: *Anal. Methods*, 2015, 7, 7404

Fitness for purpose: the key feature in analytical proficiency testing

Analytical Methods Committee, AMCTB No 68

Received 15th June 2015

DOI: 10.1039/c5ay90052b

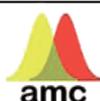
www.rsc.org/methods

The essential purpose of proficiency testing is to give participants reassurance on the quality of their performance and enable them to identify any trends or errors of consequential magnitude—those likely to affect decisions—in their reported results and then eliminate the causes of those errors. We must recognise, however, that *all* results of measurements include error. It's only a question of whether the errors are of acceptable size—in effect, whether results are fit for purpose. Fitness for purpose emerges therefore as the key feature of proficiency testing.

Most proficiency testing schemes in chemical measurement convert a participant's result x into a z -score (or an equivalent procedure), that is,

$$z = (x - x_A)/\sigma_p,$$

where the assigned value x_A is the scheme provider's best estimate of the true value, while σ_p is the standard deviation for proficiency testing (SDPT). The motivation behind scoring is to harmonise the outcome, for different analytes, matrices, and indeed schemes, in such a way as consistently to indicate to the participant what action, if any, would be an appropriate response to the result obtained. For example, a z -score falling between ± 2 is usually taken to show that there is no reason to suspect that the participant's analytical procedure calls for revision. On the other hand, a score of say 5.9 demands an investigation of the analytical system and, where necessary, elimination of the feature causing the error. But as a guide to action z -scores would be valid only if the SDPT σ_p were an uncertainty that was fit for purpose in the particular application sector.

amc technical briefswww.rsc.org/amc

AMC Technical Briefs are produced by the Analytical Methods Committee, the Technical Subcommittee of the Analytical Division of the Royal Society of Chemistry.

Selecting a standard deviation for proficiency testing

The original ISO/IUPAC/AOAC Harmonised Protocol for Proficiency Testing of Analytical Laboratories, and subsequent ISO Guides and Standards, mention several ways in which a provider could in principle determine an SDPT. These ways (see for example ISO 13528), although overlapping to a degree, fall into two main categories—those based on how participants *actually perform* and those based on how they *ought to perform* if they are best to fulfil their customers' needs. The revised Harmonised Protocol comes out overwhelmingly in favour of the latter, that is, with SDPT values based on fitness for purpose.

What difference does it make?

Proficiency tests based on how participants actually perform typically equate the SDPT σ_p with the (robust) standard deviation of the results in that round of the test. This statistic, however, simply describes the dispersion of most of the results—it brings nothing new to the discussion. It always ensures that a great majority of the participants receive a z -score between ± 2 , about 95% of them, or somewhat fewer when (as almost invariably) the results are heavy-tailed or include outliers. This strategy certainly allows the identification of discrepant participants, but has several disadvantages and considerable scope to mislead.

- It allows most of the laboratories to receive a respectable-looking score on most occasions, regardless of whether or not their uncertainties are sufficiently small to satisfy their customers' requirements. It does not encourage participants to move towards fitness for purpose.

- It is inconsistent, as the observed SD varies round-to-round. It therefore does not allow an individual participant to track performance over time and thereby identify trends and determine whether remedial changes to equipment or procedures have been successful.

- Overall, scores are influenced at least as much by the performance of other participants, and the same laboratory participating in more than one scheme will probably receive different performance scores.

- It neither allows the provider to assess the overall utility of the scheme nor legitimately to establish whether the scheme is helping participants to improve or to maintain a good performance.

A minor improvement in this type of criterion can be obtained by summarising the pattern of dispersion over many rounds of a test, because the resulting criterion, if well derived, allows a more consistent valid comparison round-to-round. Even so, it does not address fitness for purpose. Moreover, it demands the collection of statistics from a long sequence of rounds before a useful criterion can be determined. This is because (a) in a fixed type of test material the observed dispersion will depend on the concentration of the analyte, and (b) for different test materials, even those closely similar in matrix and analyte concentration, the dispersion of results can vary markedly.

A fitness-for-purpose criterion

These problems vanish when a fitness-for-purpose uncertainty is used as the SDPT. The criterion can be set before the scheme is inaugurated, either as a fixed value or a fixed function of concentration. The *z*-scores for a particular participant can then be meaningfully compared between rounds, between analytes and test materials. The scheme provider can monitor overall performance as a function of time. Best of all, the criterion leads directly to meaningful action limits when the *z*-scores are interpreted according to the standard normal distribution. Forming a score by using a fitness-for-purpose criterion adds value to the plain result.

A note on fitness for purpose

The exact meaning of fitness for purpose eluded analysts for many years, but a study of the needs of very different types of application sector provided the answer. If the uncertainty on the result is too large, the customer will make too many inept decisions. That can be costly—sometimes extremely costly—in financial terms or in harm to the public. If the uncertainty is too small, the customer pays an exorbitant price for the analytical result. The optimum (fit-for-purpose) uncertainty minimises the customer's average total outlay per result.

This conceptual fitness for purpose, however, is usually too difficult to calculate so, in most instances, analysts and their customers arrive at the fit-for-purpose uncertainty by an evolutionary process, which of course differs among various application sectors. Providers of proficiency tests should be well aware of this process and able to follow the appropriate pattern for their sector.

Conclusions

Criteria based on fitness for purpose are well designed for calculating *z*-scores that are of maximum use to participants in a proficiency testing scheme. Such criteria should be determined solely on the basis of the customers' requirements in the particular analytical application sector. The choice should be made by a panel of experts in the field, with the help of the scheme's advisory committee.

David G Bullock (UK NEQAS) and Michael Thompson.

This Technical Brief was prepared for the Analytical Methods Committee and approved on 15/06/15.

Further reading

- 1 International Harmonised Protocol for Proficiency Testing in Analytical Laboratories, *Pure Appl. Chem.*, 2006, **78**, 145–196.
- 2 M. Thompson and S. L. R. Ellison, *Accredit. Qual. Assur.*, 2006, **11**, 373–378.
- 3 ISO/IEC 17043, *Conformity assessment—general requirements for proficiency testing*, 2010.
- 4 D. G. Bullock, External quality assessment in clinical chemistry: an examination of requirements, applications and benefits, PhD thesis, 1988, <http://ethos.bl.uk/OrderDetails.do?did=1&uin=uk.bl.ethos.365386>.

CPD Certification I certify that I have studied this document as a contribution to Continuing Professional Development.

Name.....

Signature.....Date.....

Name of supervisor.....

Signature.....Date.....