INTERNATIONAL ASSOCIATION OF GEOANALYSTS*

PROTOCOL FOR THE OPERATION OF

GeoPT™

PROFICIENCY TESTING SCHEME

Revision: January 2002

*Secretary of the International Association of Geoanalysts: Ms J Cook, BGS, Keyworth, Nottingham NG12 5GG, UK. (j.a.cook@bgs.ac.uk)

©International Association of Geoanalysts 2001

CONTENTS

Foreword

The Millennium has witnessed the emergence of proficiency testing as a major influence for quality in analytical measurement. Proficiency tests have now been developed in virtually all areas where between−laboratory agreement is important. The geoanalytical sector was relatively late in the adoption of a proficiency testing scheme: this was in part because geoanalysts were unusually well provided with reference materials, a factor that to some extent offsets the need for a proficiency testing. However, even where reference materials are plentiful, there is no substitute for a proficiency test in the detection of unsuspected error in results obtained under typical conditions of analysis, from individual laboratories. Geo*PT*™ was initiated in 1994 by the embryonic International Association of Geoanalysts (IAG) with the intention of filling that need.

The main purpose of proficiency testing is to enable participants to detect unsuspected errors in their analytical systems. Of course errors will always be present − that is the nature of measurement. However, it is essential that errors are sufficiently small to make them unlikely to affect the interpretation of the data. On the other hand, we must recognise that a reduction in uncertainty is associated with a rapid escalation of costs, so it is equally important to avoid the production of data with unnecessarily small errors. This concept of appropriateness has long been recognised by geoanalysts, and is nowadays called 'fitness−for−purpose'. It is fitness−for−purpose that proficiency tests should strive to represent.

Proficiency tests, therefore, exist primarily to encourage laboratories to move towards fitness−for−purpose by instigating remedial action where error of inappropriate magnitude is detected. However, it cannot be denied that other, mainly commercial, factors now also depend on proficiency testing. It is recognised to be an essential ingredient of accreditation. Accreditation assessors will expect to see laboratories participating in a relevant proficiency testing scheme, if one exists in the sector, and will expect to see evidence of mainly satisfactory performance and of documented remedial activity in response to occasional lapses. Moreover, participants will want to demonstrate their capabilities to potential clients by showing that their proficiency test results have been largely satisfactory. While not part of the original ethos of proficiency testing, these later manifestations are simply a fact of current analytical life.

Geo*PT*™ was designed with all of the foregoing factors in mind, and we hope that it continues to fulfil a need in the geochemical community. It is a non−profit enterprise within the IAG. In an attempt to minimise costs, much of the work is done on a volunteer basis. Nevertheless, there are significant costs involved in running the scheme, including the preparation, packaging and checking of the test material, posting the material around the world, and the printing and distribution of the reports. These costs have to be passed on as a fee for participation. The fact that so many laboratories world−wide do participate, demonstrates that the enterprise is worthwhile to them and to the organisers.

Michael Thompson **January 2002**
Chairman, Geo*PT™* Proficiency Testing Steering Committee
International Association of Geoanalysts

Aims and intentions of Geo*PT*™

Geo*PT*™ provides a proficiency testing service to analytical laboratories operating in the areas of pure or applied geochemistry. It does not intend to compete with proficiency testing schemes established essentially for environmental analysis, and is mainly concerned with the analysis of rocks and sediments.

The scheme at present offers a single test material for analysis twice per year (the sample is changed from round to round), and provides the participants in advance with fitness−for−purpose criteria that establish a standard for the assessment of quality. Participants report their results by a deadline, and the processed results are made available to the participants as soon as possible after that time. The results are presented in a form that allows participants to compare their performance with the scheme's (or their own) fitness−for−purpose criterion, with their peers' current performance or with their own past performance.

The operation of the scheme is reviewed in detail every three years during the Geoanalysis™ Conference, but the organisers encourage comment at any time. Substantive comment can also be addressed at any time to the editor of the IAG Newsletter, or posted on the Bulletin Board of the IAG Website.

The nature and purpose of proficiency testing

Proficiency testing[1−3] is one of the quality assurance techniques available to analytical chemists. In its usual form, it involves the distribution of effectively identical samples of a material to the participant laboratories for analysis, usually by a method of their own choice. Results must be reported to the organisers by a published deadline. The organisers compare each participant's result with the best available estimate of the true value of the analyte concentration, and present the outcome as a score that represents the analytical performance in terms of accuracy. The score is calculated on the basis of a performance criterion known in advance by the participants. The test is repeated at regular intervals, between one and six months usually, depending on the particular scheme. Proficiency tests, therefore, provide a regular, independent, external check on accuracy, and thereby allow participants to detect and subsequently correct any unexpected source of error.

Proficiency testing is essentially a validation of individual laboratories, so it must be distinguished from methods of validating analytical methods, and should not be used by participants for that purpose. In particular, proficiency testing must not be seen as an alternative to normal method validation or routine internal quality control. Proficiency testing is also distinct from certification exercises for reference materials.

Participants should have in place a system for responding to unsatisfactory results in a round of a proficiency test. Where possible, further diagnostic tests should be carried out to determine the source of any unsuspected error. Accreditation assessors will look for evidence of not just overall successful participation in proficiency tests, but also to an appropriate response to unsatisfactory results, such as investigation, corrective action, and a follow−up to check that the corrective action was effective.

The important aspect of proficiency testing is that it should encourage in participants to seek fitness−for−purpose in their routine results. 'Fitness−for−purpose' implies that the uncertainty on a result is of a magnitude appropriate to the use to which the data will be put. This means that the standard of performance required is set independently of the results of the participants and, logically, must be known in advance of the analysis. To achieve the main objective, therefore, the results

submitted should reflect the performance of the laboratory operating under normal routine conditions. Participants who employ special methods, or particular analysts, or unusually careful methodology for proficiency testing samples are subverting the purpose of the scheme.

Organisation of Geo*PT*™

Terminology

Each type of material regularly distributed in Geo*PT*™ is known as a 'Series'. Series 1 comprises silicate rocks and allied materials for bulk analysis. Within each Series each distribution is known as a 'round'. There are normally two rounds per year. As well as 'Series' there may also be one−off studies reflecting particular interests. The material sent to participants in a particular round of a series is called the 'test material'. The individual packets of test material sent to participants are called the 'distribution units'. The quantities of the test material weighed out for analysis are called 'test portions'.

Test materials

The test materials for Series 1 are mostly silicate rocks, but some other materials such as sediments, glasses and soils may occasionally be included. Further Series, such as materials for microprobe analysis may be subsequently be launched. Materials prepared for bulk analysis are finely ground and thoroughly mixed before being split into packages for distribution. A test for effective homogeneity among the distribution units is made before distribution, according to the method given in the International Harmonised Protocol.[1]

Distribution of materials

Samples are distributed by post at least 8 weeks before the reporting deadline. For silicate rocks the distribution units comprise about 40 g of material.

Analysis

Analysis is conducted in the participant's laboratory by any suitable method, but the analytical protocol used should reflect the normal practice in that laboratory.

Reporting by participants

Participants report their results by e−mail (or alternatively by post or FAX) to the organiser on the results form provided and in the concentration units specified. The results form must reach the organiser by the deadline. Results are transcribed by Geo*PT*™ exactly as they appear on the result form, even if they are clearly in error.

Scoring

For assessment, the organisers normally convert the results to z−scores (see below). In some instances z−scores are not produced. This happens when it is not possible to obtain an assigned value (i.e., the best estimate of the true composition of the sample) with a sufficiently small uncertainty.

Reporting by organiser

The organiser provides each participant with a report comprising (a) a certificate of participation in the round including the participant's identification code, (b) a table of the participants' raw results as input by Geo*PT*™, (c) a table of z−scores corresponding to the results in (b) above, (d) a statement relating to the testing of the distributed material for sufficient homogeneity, and (e) various charts showing the z−scores obtained by all of the participants in the round. Usually only diagrams showing results for analytes that show unusual or noteworthy features will be included as individual barcharts. There may also be comment from the organiser on particular issues that have arisen as a result of the round. Reports will normally be issued to participants, by post or electronically, not more than 60 days after the reporting deadline.

It is planned that tables of results and z−scores and charts for all of the analytes with more than 10 reported results will be available from the electronic version of

Geostandards Newsletter. The organisers may also publish the results and general comments in other media. Participants' results will be identified only by confidential code number in any such publication, unless participants grant specific approval in writing..

Review by organiser
After each round the Steering Committee will review the efficacy of the scheme and take any appropriate action.

Correction of mistakes
Mistakes by participants
Mistakes in reporting by participants will not be corrected. The resultant z−score will stand regardless of the nature of the mistake. The z−score represents the performance of a participant's whole analytical system including reporting.

Mistakes by Geo*PT*™
Every reasonable effort is made to avoid mistakes in the transcription of the reported results onto a worksheet and in the calculation of z−scores. Participants should compare their reported results with data in the report. Any discrepancies should be reported to Geo*PT*™ immediately. Geo*PT*™ will issue a correction statement to the affected participant relating to any such mistake that is substantiated.

Disclaimer
Neither the IAG, nor individuals involved in processing the results, accept liability for the outcome of any mistakes in the operation of Geo*PT*™. Participation in Geo*PT*™ implies that the participant accepts this condition.

Confidentiality
Participants will be identified on Tables of results or of z−scores and any other public document only in the form of a numeric code, changed every round. Geo*PT*™ will not disclose the code identity of a participant to a third party, without the written approval of participants. A list of participants may be published from time to time and as part of the report for each round, although it is open for participants to request to be omitted from such a list.

Steering Committee
Geo*PT*™ is organised for IAG by a Committee of Council. The Chairman is appointed for five year periods by IAG Council and is *ex officio* a member of Council. The Committee is appointed by IAG Council following recommendations by the current chairman. At least half of the members must be members of the IAG and will usually be experts in geoanalysis. At least one member must be an expert in statistics. As of 1st January 2002, the Committee comprised: Prof M Thompson (Birkbeck College, University of London, UK) (Chairman), Dr P Potts (The Open University, UK), Dr P Webb (The Open University, UK), with other members coopted for specific rounds.

Scoring and statistical methods
Scoring and statistics in Geo*PT*™ is undertaken according to the recommendations of the International Harmonised Protocol[1] adopted by ISO, IUPAC, and AOAC International. Geo*PT*™ does not classify or rank participants on the basis of their performance.

The z−score

Participants' reported results (x) will usually be converted into a 'z−score', defined by $z = (x - x_a) / \sigma_p$

where x_a is the 'assigned value', that is, the organisers' best estimate of the true value of the concentration of the analyte, and σ_p is the 'target value', a value similar in function to a standard deviation that describes the acceptable range of variation among the results.

The assigned value

The function of the assigned value is to enable an estimate of the participant's error to be made. In Series 1 the assigned value is taken as the 'consensus' of the participants' results for a particular analyte. Where such results are unimodal and, outliers aside, roughly symmetrically distributed, the consensus is taken as the robust mean of the results and its uncertainty the robust standard deviation of the results divided by the square root of the number of results[3]. When the distribution is clearly skewed, it is sometimes preferable to use the median as the consensus. This choice is effected by the Steering Committee.

In some instances it is not possible to estimate a satisfactory assigned value, and then no z−scores can be calculated. Plots of the results are still useful and are produced when more than 6 results are available. Circumstances where this is likely to happen are as follows:

- There are few results and the uncertainty on the assigned value is high enough to affect the value of the z−scores. This commonly occurs when the number of results is less than about 10.
- The dispersion of the results is unusually wide, or multimodal. This can occur when participants use two or more methods that produce discordant results.
- The dispersion of the results is grossly skewed and no reasonable consensus can be identified.

The target value

The target value σ_p is a scaling factor for the participant's error and, in Geo*PT*™, its value is based on fitness−for−purpose criteria. It is set in advance of the test by the Steering Committee and describes what is judged by them to be the maximum acceptable level of uncertainty in the results. Accordingly, a z−score more extreme than ±3 implies that an unacceptable source of error may be present in the participant's analytical system and that remedial action should be taken. Z−scores more extreme than ±2 carry the same message to a lesser degree, but will occur by chance with reasonable frequency (about one in twenty results for a participant complying exactly), so isolated values will not signify much.

(It is emphasised that σ_p is not meant to be a *descriptor* of the results. It is not (in Geo*PT*™) derived from the participants' results at all. Consequently there is no prior expectation that about 95% of the results for an analyte will fall within the range of ±2.)

The value of σ_p used in Geo*PT*™ is derived from the Horwitz function[4,5], $R_H = 0.02c^{0.8495}$, where R_H is the reproducibility (between laboratory) standard deviation observed at an analyte concentration c, and both are expressed as mass ratios (for example, 1 ppm = 10^{-6}). The Horwitz function is an empirical observation that applies over a wide range of concentrations, test materials, analytes and physical principles underlying the analytical procedure.* In Geo*PT*™, two

* The exact form of the function used in Geo*PT*™ is under review: there is evidence that it is too generous both at high (> 10^{-1}) and low (< 10^{-8}) concentrations.[5]

levels of uncertainty are recognised as fit−for−purpose: 'Class 1', which is appropriate for high precision analysis for 'pure' geological research where $\sigma_p = R_H / 2$; and 'Class 2', more appropriate for 'applied geochemistry' where $\sigma_p = R_H$. (The terms 'pure' and 'applied' are used for guidance: which σ_p value is used must be determined by each participant according to their objective needs.) Some values of relative standard deviation based on σ_p, over the normal concentration range, are given in Table 1.

The value of σ_p also acts as a benchmark for the uncertainty on the assigned value $u(x_a)$. If $u(x_a)/\sigma_p > 0.6$ the z−scores may be unduly affected, so they are not usually calculated. (Sometimes, where it is felt worthwhile, z−scores are calculated and marked 'provisional − for guidance only': this is at the discretion of the Steering Committee.)

Table 1. Relative standard deviations implied by the target value σ_p.

Concentration	%RSD (Class 1)	%RSD (Class 2)
100% *m/m*	1	2
10% *m/m*	1.4	2.8
1% *m/m*	2	4
1000 μg g^{-1}	2.8	5.7
100 μg g^{-1}	4	8
10 μg g^{-1}	5.7	11.3
1 μg g^{-1}	8	16
0.1 μg g^{-1}	11.3	22.6
0.01 μg g^{-1}	16	32

Testing for sufficient homogeneity

Heterogeneity contributes to the uncertainty on the assigned value, and this is tested separately and also related to the value of σ_p. The term 'sufficient homogeneity' recognises that only solutions can be truly homogeneous and that most rock powders will be multi−mineralic and, therefore, heterogeneous in that true sense. A properly devised test can always detect such heterogeneity. Sufficient homogeneity means that the contents of the distributed units of the test material do not differ among themselves sufficiently to affect the outcome of a proficiency test based on bulk analysis: that is, the z−scores will not be affected to any noticeable degree. Clearly participants in a proficiency test must be confident that the material they are dealing with is sufficiently homogeneous. It should be noted that a material can be sufficiently homogeneous for some analytes and not for others, and hence multi−analyte homogeneity tests are needed for Geo*PT*™.

Tests for sufficient homogeneity are based on the analysis of a number of distributed units. The Harmonised Protocol[1] outlines a specific method for carrying out this procedure, as follows.*

- Crush, grind and mix the bulk material to a degree that is expected to produce effective homogeneity.
- Split the material into the distribution units (sealed containers with sufficient material for the participants needs, exactly as will be distributed), taking whatever measures are needed to minimise segregation.
- Select *at random* a number ($n > 10$) of the distribution units. (Ten is an absolute minimum number: ideally a much larger number should be used, but

* The procedure recommended in the Harmonised Protocol is recognised as being too stringent, in that it is too likely to reject materials that are, in fact, satisfactory. A modification of the procedure has been proposed[7] and a revised version of the Harmonised Protocol is expected in 2002.

this is often not economically feasible in proficiency tests.)

- Take two independent test portions of the material from each selected distribution unit. (This might involve further crushing and mixing of the distribution units separately, if the material is palpably grainy, to reduce the risk of segregation under storage.)
- Analyse the 2*n* test portion *in a random order*, if possible, in a single run of analysis, by a method with a sufficiently good precision. Record the result with sufficient significant figures to represent adequately the variability of the measurement. If in doubt, collect more significant figures than is normally justified. (Ideally the analytical repeatability standard deviation σ_a should be smaller than about $0.4\sigma_p$: if it is not, heterogeneity that is significant in the interpretation of Geo*PT*™ results may be undetectable.)
- Inspect the results graphically, paying attention to possible (a) outlying analyses (indicated by an exceptionally large difference between duplicated results for a distribution unit), (b) outlying distribution units, or (c) non−random patterns among the results. (Problem (a) indicates analytical blunders: such results, after confirmation by an outlier test, should be deleted from the data. If they are not deleted they could cause a heterogeneous material to *pass* the test. Problem (b) indicates that the material may really be heterogeneous. Such outliers must never be deleted before the statistical test. Problem (c) should be referred to the statistical expert, but may mean that the data should be abandoned and the whole test repeated.)
- Calculate, by analysis of variance, *MSW* (the mean square within samples (i.e, between analyses)), *MSB*, the mean square between samples, and calculate the estimated analytical standard deviation, $s_a = \sqrt{MSW}$, and the sampling standard deviation component, $s_s = \sqrt{(MSB - MSW)/m}$, where $m = 2$ for duplicate analysis.
- If the probability associated with the value $F = MSB / MSW$ is greater than 0.05, then no significant heterogeneity has been detected. So long as $s_a < 0.4\sigma_p$, the material is taken as sufficiently homogeneous. Even if the material is significantly heterogeneous, it is taken as sufficiently homogeneous if $s_s < 0.4\sigma_p$. (If the analytical method has poor precision, the test may be incapable of detecting an important degree of heterogeneity. If the analytical method is very precise, even very small and unimportant heterogeneities could be statistically significant.)

This procedure differs from practice in the preparation of geological reference materials. In particular, no account is taken of the heterogeneity within a distribution unit, as long as the average contents of packages are sufficiently homogeneous. This is partly because it is often difficult to distinguish between heterogeneity within distribution units and analytical problems. In Geo*PT*™, the test material is normally so finely comminuted that further grinding of the distribution unit is superfluous. However, the ethos of proficiency testing is that part of the participants task is to ensure that the test portion is sufficiently representative of the whope distribution unit, just as it is in routine analytical practice.

How to use the information obtained from Geo*PT*™

Proficiency tests are for the use of participants, primarily (i) to check for unexpected sources of error in results and (ii) to check that any remedial action to reduce errors has been successful and (iii) to check, in general, that the laboratory is

working to an expected level of uncertainty. For the increasing proportion of laboratories becoming involved in accreditation, however, there is now an obligation to participate in a relevant proficiency test if one is available and, moreover, to demonstrate overall appropriate performance and the effectiveness of procedures to deal with occasional inappropriate performance. Further, in preparing tenders for analytical work, or otherwise advertising commercial analytical services, it is now commonplace for a laboratory to cite performance indices based on proficiency test results to convince the customer that the prescribed analytical performance can be achieved. All of these circumstances require the judicious use of the results.

Such activities are essentially the responsibility of the participant. Geo*PT*™ does not have the resources for activities beyond preparing the reports as detailed above. However, some suggestions for optimal use of the data are provided here.

How to assess your results

Geo*PT*™ is not designed to be diagnostic: it provides no direct information for the participant to determine the sources within the analytical system of any inaccuracy in the result. The participant will have to devise additional tests taken from the normal range of quality assurance practices to find such information. Nevertheless, because the scheme is multianalyte, some limited diagnostics can be extracted from the results themselves.

If nearly all of the z−scores obtained in a round are within the range ±2 and the remaining few are outside the range by a small margin, then probably all is well with the analytical system. So if only a small proportion of the analytes give rise to suspect results, the investigation should first consider whether they could have plausibly arisen by chance. For example, although Geo*PT*™ z−scores cannot be interpreted in strict terms of confidence limits, it would be reasonable to expect about two results outside the range ±2 for ever 40 analytes determined. No further investigation would be called for.

Any other situation requires investigation and the possible installation of a more effective internal quality control system. Accreditation bodies would expect to see a mechanism for responding to the outcome of each round, so participants should adopt and document a systematic way of investigating such suspect results.

Several results outside ±3, or one result somewhat more extreme, however, calls for action. First, the participant should attempt to find whether the error is due to a systematic or a random effect. This can be ascertained by a few repeated analyses of the proficiency test material, in successive runs. (Analysis of a certified reference material at the same time would help to reinforce the interpretation.) A variable result from such a test suggests a random error, which could be due to a number of problems, such as using the method too close to the detection limit, or insufficient care with the manipulations of the test material or the instrumental determination. A persistent deviation from the assigned value (of roughly the same magnitude over several runs) suggests a systematic problem. This could be due to a number of causes and should be investigated further. One possibility is incomplete chemical decomposition of the test material, which would always give a negative error, but would relatively easy to track down, as the responses of most common minerals to standard attacks is well understood. Another possibility is the encountering of an unexpected matrix effect or other interference. Again, the nature of the analyte might suggest a possible culprit. A further possibility (which is quite common) is a mistake in preparing a standard solution for calibration, so calibration solutions should be checked.

If, in a multielement analysis, a number of analytes give results that are simultaneously suspect, the fault is probably systematic and must arise at that part of the analytical system where the affected analytes are all involved. For example, if the errors are nearly all in the same direction, a common action such as a mistake in weighing out the test portion may be implicated. It should be noted that some matrix interferences can affect different elements to different degrees or possibly in different directions, while incomplete dissolution may affect different groups of elements differently, according to the mineralogy of the test material.

For accreditation purposes it is important to document the procedures used for investigating problems, and to keep records of any actions taken and the apparent effect of such actions.

In the long term, it is beneficial for a participant to record z−scores in a graphical way so that results can be compared both by round and by element. This can be simply and quickly done with a hand−drawn chart similar in form to the 'Multiple Z−score Chart' used by Geo*PT*™ for comparing within−round results.

Where an analytical laboratory and its customer have agreed a fitness−for−purpose uncertainty u_{ffp} that differs from either value of σ_p used in Geo*PT*™, use can be made of an alternative 'do−it−yourself' score, the 'zeta−score' given by $\varsigma = (x - x_a) / u_{ffp}$.
The zeta−score acts as a simultaneous check on accuracy and uncertainty[8].

Proficiency testing in the overall context of quality assurance
Proficiency testing, being an occasional check on accuracy, must not be confused with internal quality control (IQC), which is a way of monitoring routine analytical operations.[6] The relationship between theses two activities is interesting, because it is known that poor performance in proficiency tests is related to poor design of the IQC system. An important point in IQC is that the control material used should have a chain of traceability independent from that of the calibrators (materials of known concentration used in calibration). If, for example, a method is calibrated with certified reference materials, the IQC should be conducted with a different material (that is, not a member of the calibration set), the reference value of which is not derived from the calibration set.

Use of excess test material
Any test material remaining after analysis in a round of Geo*PT*™ can be used in a number of ways by the participant, by reference to the assigned value and its uncertainty. (In some instances Geo*PT*™ may be able to supply unused distribution units at a cost.) Some valid uses include:

- to test whether remedial measures in the wake of unsatisfactory performance are necessary or, if they have been taken, successful;
- for internal review purposes (for example 'mock' proficiency tests for a trainee analyst);
- as an occasional supplement to routine internal quality control;
- for testing the efficacy of newly introduced methods;
- for a customer of an analytical laboratory, for blind testing the quality of the results.

In all such uses, due caution should be exercised, and Geo*PT*™ accepts no responsibility for the outcome. The assigned value will be reasonably safe where one is issued by Geo*PT*™, although the traceability of such a value has been questioned in some quarters. The assigned values do not, therefore, have the same status as those of certified reference materials, and should not be used for calibration.

Comments on classification and ranking

Classification in proficiency testing is the assignment of a participant to a named class based on z−scores in some way. For instance individual results could be classified as 'satisfactory' if $-2 \leq z \leq 2$ and 'unsatisfactory' otherwise. Such classification may be useful for a participant's internal purposes, but must be used with caution. Even the names used need consideration: a preferable classification to the above would be 'inside (or outside) action limits'. Generally, in passing from z−scores to a classification, participants are throwing information away, so such action is scientifically undesirable and, indeed, potentially misleading.

In a multianalyte proficiency test like Geo*PT*™, participants sometimes express a wish for a single index to summarise performance in a round. Such an index is, for example, essential for classification. Some statistically sound indices of this type are discussed in the Harmonised Protocol.[1] However, all such indices have a flaw: persistent (round−to−round) poor performance in a one analyte can be masked by good performance in several others. Therefore simple indices, which are some kind of average of the (absolute) z−scores, may be unsuitable. A possible good definition of 'no action needed' after a round could be of the form:

'less than 10% of the z−scores in the round are greater than 2, none are greater than 3, and none that are greater than 2 were greater than 2 in the previous round for the same analyte'.

However, the following:

'the mean absolute z−score is less than 1.5'

might indeed be true, but could be misleading.

The exact wording would depend on the needs of the participant. Classification should be used with due care to avoid misleading information in promotional or advertising material or in bids for contract work.

Ranking is the creation of a 'league table' by placing participants in order of (say) average absolute z−score. Apart from the above−mentioned problems with single indices of performance, ranking is scientifically undesirable because the derived rank is very variable and likely to change drastically between rounds, despite the fact that there is no underlying change in the performance of the participant. Therefore, calling a participant, for example, 'best' on the basis of the current rank is almost meaningless. Ranking based on Geo*PT*™ results should not be used in publicity or other promotional material.

Ethical considerations

Geo*PT*™ is offered on the understanding that participants are using the results to monitor their routine analytical activities and that the results submitted reflect that usage. Therefore, the results should represent errors from all of the normal sources. This means that the proficiency test material should be treated exactly like a routine sample, with no special attention paid to it, no particular analyst assigned to its handling, and no more than the routine number of separate results averaged to form the submitted result.

Collusion must be avoided and, while not suspected in Geo*PT*™, it has been detected in other proficiency tests where accreditation puts pressure on participants to perform well. Accordingly, Geo*PT*™ will employ occasional split rounds to make collusion unproductive. In split rounds, two distinct but rather similar materials are distributed within a round, but the participants are not informed whether this has happened or which they have received until the z−scores are published.

Participants must be careful to avoid giving any false impression based on the results of proficiency tests when advertising their services.

The situation sometimes arises that a member of the organising committee is part of the staff of a participating laboratory. Geo*PT*™ ensures that such laboratories are treated in exactly the same fashion as other participants, and do not become aware of any privileged information regarding the test material, or other participant's results.

References

1 *'International harmonised protocol for the proficiency testing of chemical analytical laboratories'*. M Thompson and R Wood, *Pure Appl. Chem.*, 1993, **65**, 2123−2144, and simultaneously in *J. AOAC Internat.*, 1993, **76**, 926−940.

2 ISO Guide 43: *'Proficiency testing by interlaboratory comparisons'*, ISO, Geneva, 1997.

3 *'Proficiency testing in analytical chemistry'* R E Lawn, M Thompson and R F Walker, The Royal Society of Chemistry, Thomas Graham House, Science Park, Milton Road, Cambridge CB4 4WF, UK, 1997, 110pp. (1993).

4 W Horwitz, L R Kamps and K W Boyer, *J. Assoc. Off. Anal. Chem.*, 1980, **63**, 1344.

5 M Thompson, *Analyst*, 2000, **125**, 385−386.

6 *'Harmonised Guidelines for Internal Quality Control in Analytical Chemistry Laboratories'*, M Thompson and R Wood, *Pure Appl. Chem*. 1995, **67**, 649−666.

7 T Fearn and M Thompson, Analyst, 2001, **126,** 1414−1417.

8 Analytical Methods Committee. www.rsc.org/lap/rsccom/amc/amc_index.htm