INTERNATIONAL ASSOCIATION OF GEOANALYSTS

PROTOCOL

FOR THE OPERATION

OF

Geo*PT*™

PROFICIENCY TESTING SCHEME

2nd Edition: May 2018

Revised: May 2020

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

CONTENTS

Foreword to the Second Edition 2018

The Geo*PT* proficiency testing scheme has gone from strength to strength since the first round was organised in 1996. The present document is the first major revision of the protocol. Whilst maintaining compatibility with IUPAC's International Harmonised Protocol for Proficiency Testing of Analytical Chemistry Laboratories[1], details in this revision take account of experience and expertise that has been accumulated over the last twenty years. The scheme is still focused on the needs of participating laboratories, with the aim of providing the best possible feedback with respect to their performance. To this end, significant developments have taken place in the use of *modes* as well as *means* and *medians* in the assessment of consensus values. Furthermore, significant advances have occurred in the assessment of measurement results by technique, so that an evaluation of data distributions now has the capability of detecting procedural bias, when present. Further confidence in the scheme has been obtained on the occasions when either certified reference materials have been used as Geo*PT* test materials, or Geo*PT* test materials have been subsequently certified on the basis of additional independent measurements. The excellent agreement that has been found on these occasions between Geo*PT* assigned values and certified values[2] provides unchallengeable evidence of the reliability of data analysis procedures used to estimate the true composition of Geo*PT* test materials. This confidence is further supported by the extraordinary stability of the Geo*PT* scheme as demonstrated by the statistically identical test results that were obtained when the same Geo*PT* test material was recirculated (anonymously) after an eighteen year period[3]. These developments demonstrate that the Geo*PT* scheme, as described in this revised protocol, continues to be fit for purpose in supporting the performance of participating laboratories.

To enhance the experience of participating laboratories and improve performance feedback, the IAG has invested in replacing the original paper-based scheme for reporting measurement results with a much more efficient and effective web-based management system. In addition, laboratories now receive a personalized certificate of performance, in addition to the full report of each round, facilitating an ‘at a glance’ assessment of performance and allowing easy comparison from one round to the next. All these developments justify the revised protocol presented below.

Phil Potts, April 2018
Chairman, Geo*PT*™ Proficiency Testing Steering Committee

Foreword to the First Edition 2002

The Millennium 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 disproportionate 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 are confident that it will continue 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 worldwide participate, demonstrates that the enterprise is worthwhile to them and to the scheme providers.

Michael Thompson, January 2002
Chairman, Geo*PT*™ Proficiency Testing Steering Committee

1. Overview of Geo*PT*™

Geo*PT*™ provides a proficiency testing service for analytical laboratories operating in the areas of both pure and applied geochemistry. It is concerned mainly with the chemical analysis of geological materials, especially silicate rocks, sediments and related materials.

The scheme offers an appropriate test material for analysis at twice yearly intervals. The principal test material may on occasion be accompanied by a supplementary test material. Participants report their measurement results by a published deadline, and the results of the proficiency test are made available to all participants in the form of a report once the data have been processed. The report enables participants to compare their quantity values, identified by code to maintain anonymity, with the scheme provider's best estimate of the true value for each measurand and to evaluate their performance both in relation to the scheme's fitness-for-purpose criterion, and their peers' current performance. Participants are also encouraged to review their results in relation to their own past performance.

The scheme is provided by the International Association of Geoanalysts (IAG), and its operation is managed by a Steering Committee. Feedback from participants is encouraged and should be referred in the first instance to the Scheme Administrator. Subscriptions and invoicing are the responsibility of the IAG's Subscription Manager.

2. Principles of proficiency testing

Proficiency testing[1,4–6] is a widely accepted quality assurance tool developed for analytical chemists. It provides an opportunity for analysts and quality managers to test the reliability of their analytical procedures. In its usual form, proficiency testing involves the distribution of identical samples of a test material to participant laboratories for analytical measurement, usually by a method of their own choice, employing their routine procedures. Results must be reported by a specified deadline. The scheme providers compare each participant's measurement result with the best available estimate of the true value for each measurand, and present the outcome as a score that represents the participant's analytical performance in terms of accuracy. The score is calculated on the basis of a performance criterion specified in advance and known to the participants. The test is repeated at regular intervals.

The main objective of proficiency tests is to provide a regular, independent and external check on the accuracy of measurement results, thereby allowing participants to detect, investigate and subsequently correct any unexpected sources of error in their routine analytical procedures. To achieve this objective the measurement results submitted should reflect the performance of the laboratory operating under normal routine conditions. Participants who employ non-routine procedures, or specially chosen analysts, or unusually careful methodology for measurement of proficiency testing samples are undermining the purpose of the scheme: they will not be able to discover deficiencies in their routine practices.

Participants should have in place a system for responding to unsatisfactory results identified in a round of a proficiency test. Where necessary, further diagnostic tests should be carried out to determine the specific source of any unexpected error. Accreditation assessors will look for evidence not just of the overall successful participation in proficiency tests, but also of an appropriate response to unsatisfactory results, including investigation, implementation of corrective action, and follow-up to ensure that the corrective action had been effective.

An important aspect of proficiency testing is that it should encourage participants to achieve a standard of performance in the quality of their routine results that is fit for purpose. 'Fitness for purpose' implies that the uncertainty in a result is of a magnitude appropriate to the use to which the data will be put.

3. Organisation of Geo*PT*™

Geo*PT*™ is managed for the IAG by a Steering Committee appointed by Council. Normally one member of the Steering Committee is the Scheme Administrator. At least half of the members of the Steering Committee must be members of the IAG and at least one member should be a member of Council. Members of the Steering Committee will usually be experts in geoanalysis, and at least one member must have expertise in statistics. Details of the current Management Team are provided in Appendix A, along with contact details for the Scheme Administrator and the Subscriptions Manager.

3.1 Terminology

Each distribution of material in Geo*PT*™ is known as a ***'round'***. There are normally two rounds per year, six months apart. The material sent to participants in a particular round is called the ***'test material'***. Individual packets of test material sent to participants are called ***'distribution units'***. The quantities of the test material weighed out by participants for analysis are called ***'test portions'***. The principal test material is intended to satisfy the PT requirements for the majority of participating laboratories and is generally referred to as the ***'routine'*** test material. An additional test material may, on occasion, be distributed, and is usually designed to test the performance of participating laboratories on a wider range of geological matrices. It is normally referred to as a ***'supplementary'*** material.

3.2 Test materials

The test materials are finely-ground powders, produced mostly from silicate rocks, but some other geological materials such as sediments, glasses, carbonate rocks, mineralised materials and soils may occasionally be provided as test materials. Materials prepared for distribution are thoroughly homogenised before being split into distribution units. A test for sufficient homogeneity of the test material is made before distribution, broadly conforming with recommendations in the International Harmonised Protocol[1]. The organisers may choose to take advantage of other evidence in assessing homogeneity, including particle size distributions. For some test materials a minimum size of test portion may be recommended.

3.3 Distribution of materials

Test materials are distributed by an appropriate means, normally at least 10 weeks before the reporting deadline in order to secure delivery in sufficient time to allow analytical measurement to be undertaken. For most test materials the distribution units comprise about 40 g of material. Test materials are accompanied by a ***Letter of Invitation*** to participants with ***Instructions to Analysts*** that provide full instructions for handling, measurement and reporting of results. Participants are notified when test materials have been dispatched and are required to inform the administrator if they have not arrived after 3 weeks have elapsed in Europe and North America; 4 weeks have elapsed in South America, Africa, Middle East and Asia; and when 5 weeks have elapsed in Australasia.

3.4 Analysis by participants and handling of PT distribution units

Analytical measurements are conducted in the participant's laboratory by any procedure or procedures selected by the participant, but the analytical protocol used should reflect the routine practice in that laboratory. Only one quantity value per analyte may be reported.

The scheme organisers recommend that participants adhere to good laboratory practice when handling and measuring the test materials provided by Geo*PT*™. Test materials should be handled with the usual care in respect of health and safety that is appropriate for geological materials containing silica and silicate minerals. This includes taking precautions against the inhalation of dust when handling powder, and the inhalation of fumes when fusing or roasting test portions. Material Safety Data Sheets applicable to the test materials provided are available on request from the scheme administrator. Occasionally, when the scheme organisers are aware that materials identified as being derived from mining operations or tailings sites, or believe that they could contain noteworthy amounts of toxic elements, a specific warning will accompany the test material. Analysts are also made aware that some samples, such as those containing significant amounts of sulfide or arsenide minerals, may attack platinum ware or other laboratory equipment. However, IAG cannot accept responsibility for any damage or misadventure occurring from handling or processing test materials.

3.5 Reporting of analytical results by participants

Participants must report their measurement results online via the Geo*PT*™ website according to the specifications stated in the ***Instructions to Analysts***. Before recording their measurement results, participants are required to provide details of their analytical procedures, including the analytical technique employed, the method of sample preparation, and the mass of the test portion used. Analytical results must be reported in the quantity units specified for the species identified, e.g. Au in mg kg^{-1}, CaO in g $100g^{-1}$, and be submitted by the deadline specified in the *Letter of Invitation* in order to take part in the proficiency test. Once submitted, results cannot be changed, either by the participant, or by the organisers, even if they are clearly in error. The reporting of measurement results is regarded as part of the measurement process.

3.6 Confidentiality

Participants' identity will normally remain confidential to the scheme administrator alone. In exceptional circumstances when assistance is required from members of the Geo*PT*™ Steering Committee or IAG Officers, those individuals will also be bound by confidentiality. Participants will be identified by a numeric code in all published tables of analytical results and z-scores, also on charts in reports and in any other publicly available document. Code identifiers are changed for every round of Geo*PT*™. The scheme organisers will not disclose the code identity of a participant to any third party without the written approval of the participant.

3.7 Assessment of laboratory performance

To assess laboratory performance, the scheme organisers usually convert each measurement result reported by participants into a z-score (see Section 4: Scoring and statistical methods), which provides a means of assessing performance for each analyte in relation to the fitness-for-purpose criterion employed by Geo*PT*™ and the quality option selected by the individual laboratory. For some analytes no z-scores are produced. This happens when it is not possible to derive either an assigned value (i.e., the best estimate of the true value of a measurand) or a provisional value (i.e., an estimate of the true value of a measurand that carries a greater degree of uncertainty, as further detailed below).

3.8 Reporting of laboratory performance by the scheme organiser

For each round of proficiency testing the scheme provides each participant with online access to a customised certificate of performance and a full report of results which contains:

(a) a description of the details relating to the particular round, including the type of material and its source, a summary of the data submitted and an outline of the type of statistical analysis carried out by the scheme organisers;

(b) a table of raw results as supplied by all participants – listed by laboratory code to ensure anonymity;

(c) a table listing all assigned and provisional values for the test material with corresponding uncertainty estimates, target precision and statistical details;

(d) a table of *z*-scores corresponding to the results supplied in (b) above, for those analytes credited with assigned or provisional values;

(e) bar charts of submitted data ordered incrementally and shown relative to the optimal consensus value and appropriate *z*-score benchmarks, and

(f) a multiple *z*-score chart highlighting results for participant laboratories that may not be satisfactory and require reviewing.

There may also be comments from the scheme organisers on particular analytical issues that have arisen. Reports will be made available to participants to download from the Geo*PT*™ website normally not more than 7 weeks after the reporting deadline.

The scheme organisers may publish Geo*PT*™ results and comment on them in other media. In any such publication, participants' results will remain anonymous and where necessary, identified only by their confidential code number (unless participants grant specific approval in writing for their identity to be disclosed).

3.9 Review by the scheme organiser

On behalf of the IAG, the Steering Committee will periodically review the efficacy of the scheme in print[7], or at Geoanalysis Conferences, and, should it be necessary, take any appropriate action to revise practices.

3.10 Correction of mistakes

Mistakes by participants

Mistakes in reporting made 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, which includes not only the analytical measurement result, but the whole measurement process, including maintaining the identity and integrity of the sample, and the reporting of results.

Mistakes by GeoPT™

Every reasonable effort is made by the scheme organisers to avoid mistakes in conducting each round of proficiency testing, from the provision of test materials to the calculation of *z*-scores. Participants should communicate any perceived problems to the scheme administrator, who will deal with them immediately if at all possible. If this is not possible, they will be subject to a thorough investigation by the Chairman of the Steering Committee. Geo*PT*™ will issue a correction statement to the affected participant relating to any mistake that is substantiated.

3.11 Disclaimer

Neither the IAG, as scheme provider, nor individuals involved in the management of Geo*PT*™ or in the processing of contributed data, 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. Full terms and conditions of participation in Geo*PT*™ are provided in Appendix D and are available from the Geo*PT*™ website.

4. Scoring and statistical methods

Scoring and statistical analysis in Geo*PT*™ is compliant with the ISO 13528: 2015 Standard[8] relating to statistical methods used in proficiency testing which is largely based on the earlier recommendations of the IUPAC International Harmonised Protocol[1].

4.1 The z-score

Participants' reported results (x_i) for each analyte will be converted into a '*z*-score', defined by $z = (x_i - x_{pt}) / \sigma_{pt}$, where x_{pt} is the organisers' best estimate of the true value of a measurand, and σ_{pt} is the corresponding standard deviation for proficiency testing (SDPT), a value based on a Geo*PT*™ fitness-for-purpose criterion as detailed below in Section 4.3. The SDPT was originally called the 'target value' in our 2002 protocol document, but is more conveniently referred to as the 'target precision' in this document. Thus, $(x_i - x_{pt})$ is the error in the measurement and the *z*-score provides a measure of the accuracy of the result submitted, scaled according to the SDPT, which is similar in function to a standard deviation that describes the acceptable range of variation among the results. Accordingly, a *z*-score outside the range ±3 implies that an unacceptable source of error may be present in the participant's analytical system and that the need for remedial action should be considered. *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 with the scheme's fitness-for-purpose criterion), so isolated values will not be especially noteworthy.

4.2 The assigned value

For a particular analyte, the assigned value is the scheme organisers' best estimate of the true value of the measurand in the test material and is evaluated as a consensus derived from all contributed measurement results. The consensus is recognized as the location on the measurement scale at which the density of contributed results is greatest. The function of the assigned value is to enable an estimate of error in a participant's measurements to be made. Estimation of the optimal consensus value and its associated uncertainty take account of the following:

- When the dataset is approximately symmetrical apart from a small proportion of outliers, the Huber H15 robust estimates of mean $(\hat{\mu})$ and standard deviation $(\hat{\sigma})$ of the *n* data are the statistics of choice. The consensus is taken as $\hat{\mu}$ and its uncertainty is taken as the standard error of $\hat{\mu}$, namely $\hat{\sigma}/\sqrt{n}$. (*Note*: in some instances it may be preferable to replace *n* by a slightly smaller value to account for the downweighting in the robust algorithm[9].)
- When the dataset is less symmetrical, but there is nevertheless a well-defined consensus, the median may be preferable to the Huber H15 robust mean as an estimate of the consensus. The uncertainty on the median can be taken as the simple standard error of the mean multiplied by $\sqrt{\pi/2}$, i.e. 1.2533.
- When the distribution is skewed, sometimes more strongly, but there remains a clear consensus, and the asymmetry is judged to originate from recognised technical deficiencies in measurement procedures, a mode may provide a suitable location estimate. Modes may be estimated in various ways, among them several described by Thompson[10]. A procedure involving resampling techniques ('bootstrapping'), as approved by the ISO Standard[8], provides an estimate of the standard error of the mode, a value that can be taken as the uncertainty of the consensus. In some circumstances, the mode provides a better definition of the consensus location than either the median or the robust mean.

The choice of location estimator to provide the appropriate consensus value is made by expert judgement of the Steering Committee.

Criteria taken into account for a consensus value to be credited with 'assigned' status normally include:

- At least 15 valid measurement results (i.e. excluding outliers) contribute to recognition of the consensus.
- These data conform closely to a random sample from a normal distribution.

- The ratio of the uncertainty in the location estimate to the standard deviation for proficiency testing, i.e. the target precision, calculated as $u(x_{pt}) / \sigma_{pt}$ is an acceptably small value (usually less than 0.5, see Section 4.3).
- An evaluation of measurement results by procedure – including both analytical technique and method of sample preparation employed – indicates no detectable procedural bias among measurement results from which the consensus is derived. If procedural bias is detected and its origin understood, it may be eliminated by using a judicious choice of procedure for deriving the consensus. Such considerations are reported accordingly.

Where these criteria are not fully met, but a well-defined consensus value can be derived from the dataset, it may be credited with 'provisional' rather than ‘assigned’ status to provide laboratories with some useful *z*-score feedback. Instances of provisional status may be recorded when:

- A relatively small number of measurement results (but at least 8) contribute to the consensus.
- The measurement results are unduly dispersed in relation to the target precision.
- The distribution of values is clearly skewed, but a meaningful consensus is still obtainable.
- An evaluation of measurement results by procedure indicates that bias is present and is either of unknown origin or cannot be entirely eliminated by judicious choice of procedure for deriving the consensus.

In some instances, it is not possible to estimate either a satisfactory assigned or provisional value, and then no *z*-scores are calculated. Bar charts of the results can still be useful, however, and are provided for information when more than 6 results are available. Circumstances where this is likely to happen are when:

- Too few measurement results are contributed and the uncertainty on the assigned value is high enough to affect the value of the *z*-scores (see Section 4.3). This commonly occurs when the number of results is less than about 10.
- The dispersion of the measurement results is unusually wide in relation to the target value, or the distribution of results is multimodal.
- The dispersion of the measurement results is grossly skewed and no meaningful consensus can be identified.
- An evaluation of measurement results by procedure indicates that bias is present and cannot be eliminated by judicious choice of procedure for deriving the consensus.

4.3 The 'standard deviation for proficiency testing (SDPT)' or 'target precision'

The SDPT or target precision, σ_{pt}, is a scaling factor which enables the error in individual measurement results reported by a participant to be represented as a score. In Geo*PT*™ its value is based on a fitness-for-purpose criterion; the principle on which it is based is described below. The target precision effectively describes what is judged to be an acceptable level of uncertainty in measurement results taking account of fitness for purpose and factors such as cost.

It must be emphasised that σ_{pt} is not a descriptor of the results. In Geo*PT*™ this parameter is not derived directly from the participants’ results. Consequently, there is *no prior expectation* that the participants’ results will be normally distributed or that about 95% of the *z*-score results for an analyte should fall within the range of ±2.

The value of σ_{pt} used in Geo*PT*™ is derived from the Horwitz function[11,12], $\sigma_H = 0.02c^{0.8495}$, where σ_H is the reproducibility (between laboratory) standard deviation observed at a mass fraction *c*, and both are expressed as mass ratios (for example, 1 mg kg^{-1} (i.e. 1 ppm) = 10^{-6}). The Horwitz function was originally derived from empirical observations that were found to apply over a wide range of measurands, test materials, analytes and physical principles underlying the analytical measurement procedure[11].

In Geo*PT*™, two levels of uncertainty are recognised as fitness-for-purpose criteria: 'Quality 1', where $\sigma_{pt} = \sigma_H / 2$, which is judged to be appropriate for high precision measurement in 'pure' geological research, where care is taken to provide data of high precision and accuracy, sometimes at the expense of a reduced sample throughput rate; and 'Quality 2', where $\sigma_{pt} = \sigma_H$, which is judged to be more appropriate for 'applied' geochemistry, where, although precision and accuracy are still important, the main objective is to provide measurement results on large numbers of samples collected, for example, as part of geochemical mapping projects or geochemical exploration programmes. The terms 'pure' and 'applied' are used only for guidance. Which of the σ_{pt} values, i.e. Quality 1 or 2, is used must be selected by each participant according to their objective needs. Some values of relative standard deviation based on σ_{pt}, over the range of mass fractions routinely encountered, are given in Table 1.

The value of σ_{pt} also acts as a benchmark for judging the uncertainty on the assigned value $u(x_{pt})$. If $u(x_{pt}) / \sigma_{pt} > 0.6$ the *z*-scores may be unduly affected by the relatively high uncertainty, so they are not usually calculated. Sometimes, however, when considered justified, *z*-scores are calculated and credited with 'provisional' status, at the discretion of the Steering Committee.

Table 1. Relative standard deviations implied by the target precision, σ_{pt}.

Quantity values	**Mass fraction**	**σ_{pt} %RSD (Quality 1)**	**σ_{pt} %RSD (Quality 2)**
100 g/100g	1	1	2
10 g/100g	0.1	1.4	2.8
1 g/100g	0.01	2	4
1000 mg/kg	0.001	2.8	5.7
100 mg/kg	0.0001	4	8
10 mg/kg	0.00001	5.7	11.3
1 mg/kg	0.000001	8	16
0.1 mg/kg	0.0000001	11.3	22.6
0.01 mg/kg	0.00000001	16	32

5. Testing for sufficient homogeneity

Conventional homogeneity testing is geared to ensuring that no statistically significant difference can be detected between distribution units. Such tests are described in the IUPAC Harmonised Protocol[1] and are outlined in Appendix B. They ensure that participants receive distribution units of essentially the same material. For geological materials, which consist of discrete mineral grains of different compositions and are inherently inhomogeneous on the scale of individual milled particles, it is important to establish the minimum quantity of test material which could be expected to be representative of the distribution unit. This quantity of test material varies with analyte, and the distribution of elemental mass fractions varies with mineralogy of the test material. While it is accepted that some laboratories may routinely measure test portions of less than 200 mg, it is considered that this quantity would be an appropriate minimum quantity for most test materials. For some test materials a minimum size of test portion may be recommended.

6. Using the information and materials supplied by Geo*PT*™

Proficiency test results are primarily for the use of participants to:

(a) identify unexpected sources of error in their results;
(b) establish whether any remedial action previously taken to reduce errors had been successful; and
(c) check, in general, that the laboratory is working to an expected level of uncertainty.

For the increasing proportion of laboratories becoming involved in accreditation, there is an obligation to participate in a relevant proficiency test if one is available and, moreover, to demonstrate both overall appropriate performance and the effectiveness of procedures to deal with occasional inappropriate performance. It is now commonplace for a laboratory to use proficiency test results to demonstrate that a particular level of 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 the preparation of reports and certificates as detailed above. However, some suggestions for optimal use of the data are provided here.

6.1 How to assess your results

The following account assumes that the chosen Geo*PT*™ fitness-for-purpose criterion (Quality 1 or Quality 2, Section 4.3), is appropriate for assessing the performance of the laboratory. 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. If only a small proportion of measurands give rise to *z*-score results outside the range ±2, it should first be considered whether they could plausibly have arisen by chance. For example, although Geo*PT*™ *z*-scores cannot be interpreted in terms of strict confidence limits, it would be reasonable to expect about one in twenty results to be outside the range ±2, and no further investigation would be necessary. A control chart approach is best practice for the long-term assessment of *z*-scores. Two successive *z*-scores outside the range ±2, or one result somewhat more extreme, would call for action.

A higher proportion of such results would call for further investigation and possibly the installation of a more comprehensive internal quality control system. The form that corrective action might take depends on the nature of the error. 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 any unsatisfactory results.

Geo*PT*™ is not designed to be diagnostic: it provides no direct information for the participant to determine the sources of error within the analytical system. The participant will need to devise additional tests taken from the normal range of quality assurance practices to obtain such information. Nevertheless, because the scheme involves multiple analytes, some limited diagnostic indicators can be derived from the results themselves. Further advice on diagnostic indicators is provided in Appendix C.

6.2 Proficiency testing in the overall context of quality assurance

Proficiency testing, being an occasional check on the accuracy of measurements from an analytical system, must not be confused either with internal quality control (IQC) or with validation, both of which provide ways of monitoring routine analytical operations[13] on a routine basis.

6.3 Use of excess test material

Geo*PT*™ test material remaining after analysis can be used in a number of ways by the participant, with reference to the assigned value and its uncertainty. Some test materials are available for sale through IAGeo Limited at reduced rates for IAG members.

Geo*PT*™ accepts no responsibility for the outcome of any use to which a test material is put. Although there are good grounds for believing assigned values to be reliable indicators of composition, it is generally considered that they do not, without further evaluation, have the same status as certified reference values.

6.4 Comments on classification and ranking

There is no merit in converting *z*-scores into named classes—it destroys the information content. However, it is useful to define warning and action limits in terms of *z*-scores. Geo*PT*™ does not rank the performance of participating laboratories. A discussion of the serious shortcomings of ranking methods can be found in the IUPAC Harmonised Protocol[1].

7. Ethical considerations

Geo*PT*™ is offered on the understanding that participants are using the results to check their routine analytical activities and that the results submitted reflect that usage. Therefore, the results should incorporate 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 amongst participants must be avoided. Whilst not suspected in Geo*PT*™, it has been detected in other proficiency tests where accreditation puts pressure on participants to perform well. Measures are in place to check submissions for signs of collusion. The IAG reserves the right to exclude from the Geo*PT*™ scheme any laboratory for which a *prima facie* case of collusion can be established.

Participants must be careful to avoid giving any false impression based on the results of proficiency tests when advertising their services. Wider decisions based on *z*-scores should be made only with expert consideration of the analytical and statistical principles involved.

The situation sometimes arises that a member of the Steering Committee is a member of staff at a participating laboratory. Geo*PT*™ undertakes to ensure that such laboratories are treated in exactly the same way as other participants, and do not become aware of any privileged information regarding the test material, or other participants' results.

APPENDICES

Appendix A

Constitution of the GeoPT management team

As of 1st May 2018, the permanent members of the Steering Committee are: Prof. P J Potts – Chairman (The Open University, UK), Dr P C Webb – Scheme Administrator (formerly of The Open University, UK), Prof. M Thompson – Statistician (Birkbeck College, University of London, UK), Dr C J B Gowing (British Geological Survey, UK). Individuals involved in the provision of test materials are co-opted to the Steering Committee for specific rounds.

The Scheme Administrator may be contacted by email: *geopt.iag@gmail.com*. The Geo*PT* Subscriptions Manager is Mr C Jackson, who may be contacted by email: *iag-treasurer@virginmedia.com*.

Appendix B

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_{pt}. 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. 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*™. Concern has been expressed relating to usefulness of homogeneity tests, as the tests seldom detect significant heterogeneity because experimental designs that are economically feasible have insufficient power[14].

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:

- 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 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 2n test portions 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_r should be smaller than about $0.4\sigma_{pt}$: if it is not, heterogeneity that is significant in the interpretation of Geo*PT*™ results may be undetectable.)
- Inspect the results graphically, paying attention where necessary to:
  - (a) outlying analyses (indicated by an exceptionally large difference between duplicated results for a distribution unit), which 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.
  - (b) outlying distribution units, which indicates that the material may really be heterogeneous. Such outliers must never be deleted before the statistical test.
  - (c) non-random patterns among the results, which 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 distribution units, i.e. between analyses), MSB, the mean square between distribution units, and calculate the estimated analytical standard deviation, $s_r = \sqrt{\text{MSW}}$, and the sampling standard deviation component, $s_s = \sqrt{(\text{MSB} - \text{MSW}) / m}$, 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 $s_r < 0.4\sigma_{pt}$, the material is taken as sufficiently homogeneous. Even if the material is significantly heterogeneous, it is taken as sufficiently homogeneous if $s_s < 0.4\sigma_{pt}$. 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 unnecessary.

Appendix C

Advice for investigation of unsatisfactory results

When there is evidence (as indicated in Section 6.1: How to assess your results) that *z*-score results are unsatisfactory, the participant should attempt to discover whether the error is systematic or random. This can be ascertained by obtaining a few repeated measurement results for the test material in successive runs. Analysis of a certified reference material at the same time would help to reinforce the interpretation. Variability of measurements from such a test suggests a random effect, which could be due to a number of problems, such as determining quantity values too close to the detection limit of the method, or taking insufficient care with the manipulation of the test material or the operation of the instrument. 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[15], which would always give a negative discrepancy, but would be relatively easy to track down, as the responses of most common minerals to dissolution procedures are well understood. Another possibility is encountering an unexpected matrix effect or other interference. Again, the nature of the element might suggest a possible cause. 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 measurands are simultaneously suspect, the fault is probably systematic and must arise at that part of the analytical system where all of the affected analytes are involved. For example, if the errors are nearly all in the same direction, a common action such as a mistake in weighing of 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 whole groups of elements and one group may behave differently from another group, according to the mineralogy of the material.

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

In the long term, it is beneficial for a participant to record their *z*-scores graphically so as to compare results both by round and by element. This can be done very effectively with a chart similar 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_f that differs from either value of σ_{pt} used in Geo*PT*™, use can be made of an alternative 'do-it-yourself' score, the 'z_L-score' as given by $z_L = (x_i - x_{pt}) / u_f$.

Appendix D

Terms and Conditions of Participation in GeoPT

The Geo*PT* programme is operated by the International Association of Geoanalysts (IAG) for the benefit of the geoanalytical community. By taking part in Geo*PT*, participants must accept the following:

General terms and conditions

A. The IAG shall not be liable for any loss, damage, personal injury or death (other than death or personal injury suffered as a result of negligence on the part of the IAG) which results from the operations of the participant whether or not in relation to Geo*PT*.

B. The IAG shall not be liable to the participant for loss (whether direct or indirect) of reputation, profits, business or anticipated savings or for any indirect or consequential loss or damage whatsoever even if previously advised thereof and whether arising from negligence, breach of these Terms and Conditions or howsoever.

C. In any event, and notwithstanding anything contained in these Terms and Conditions, IAG's liability in contract, tort (including negligence or breach of statutory duty) or otherwise arising by reason of or in connection with these Terms and Conditions shall be limited to the price for the proficiency test giving rise to such liability.

D. The IAG does not grant any warranties in relation to Geo*PT* products or the supply of analytical services or distribution of the proficiency test, and all other conditions, warranties, stipulations or other statements whatsoever, whether express or implied, by statute, at common law or otherwise howsoever, relating to the Geo*PT* products, analytical services or proficiency tests are hereby excluded. In particular, (but without limitation to the foregoing) no warranties are granted regarding the fitness for purpose, performance, use, quality or merchantability of the Geo*PT* products, whether express or implied, by statute, at common law or otherwise howsoever.

Specific terms and conditions

1. Each round of the Geo*PT* programme is conducted as far as possible in accordance with the published *Protocol for the Operation of the Geo*PT *Proficiency Testing Scheme* (2018), available for download at http://www.geoanalyst.org/documents/GeoPT-protocol.pdf. If variations arise in a particular round, they are documented in the relevant report. Whilst every effort is made to ensure that the operation of Geo*PT* conforms to the published protocol and that results appearing in Geo*PT* reports provide an accurate account of the results submitted, neither the IAG nor any individuals undertaking activities on behalf of the IAG can be held liable for deficiencies in the operation of Geo*PT* nor for errors made in the reporting of results nor for the consequences of any errors that might occur. Participation in Geo*PT* implies acceptance of this condition.

2. Participation in Geo*PT* is open to any commercial enterprise, academic institution or governmental organisation making advance payment to the IAG at the current rate. The administrators of Geo*PT* reserve the right to exclude laboratories whose subscription remains unpaid by the reporting deadline. Geo*PT* will provide paid-up subscribers with a certificate of performance and a report that expresses their results in the form of a *z*-score for each oxide/element that has been given an assigned or provisional value.

3. Geo*PT* undertakes to supply subscribers twice per year with a test sample that has the composition of a regular silicate rock; this may be accompanied by a supplementary sample. Test samples are dispatched to addresses recorded on the Geo*PT* website, allowing ample time for delivery and for analysis to be undertaken. Participants are notified by email at the time of dispatch. If a sample has not arrived at a destination in Europe or North America by 2 weeks following dispatch, or in the rest of the world by 5 weeks following notice of dispatch, the organisers should be informed (*geopt.iag@gmail.com*) and a replacement will be sent by courier. While Geo*PT* will make every effort to ensure that samples are received in good time, Geo*PT* cannot be held responsible for non-arrival of test samples. However if the non-arrival of a sample prevents participation in a round of testing, Geo*PT* will apply a credit to the subscriber's account.

4. The identity of laboratories submitting results to the Geo*PT* programme and their account details are maintained as confidential by the IAG. The IAG reserves the right to publish reports or any other investigations involving Geo*PT* data (containing anonymised details of analytical results and/or procedures undertaken) by any appropriate means. Neither the identity of a participating laboratory nor the results submitted will be communicated to any third party without the formal approval of that laboratory. Full details of the data protection policy of the IAG are available at: *http://www.geoanalyst.org/data-protection-policy/*
   Note: To maintain confidentiality, it is incumbent upon participating organisations to inform us when contact personnel are no longer active. Only when we have such notification can we ensure that former contact personnel cannot continue to access their results.

5. Paid-up participants of the Geo*PT* programme have the right to use the Geo*PT* website to enter and/or import data, to access reports, to amend their own details for communication and for delivery of samples and to receive news from the organisers. The organisers reserve the right to deny access to any participant who abuses the system.

6. Participants are expected to ensure that their personal and institutional data as recorded on the Geo*PT* website are correct and up to date so that they continue to have access to the website and are contactable by the organisers. In addition, they should ensure that receipt of emails from *geopt.iag@gmail.com* (for direct contact with the administrator) and *noreply@geopt.info* (for automated notifications) is permitted and they are not intercepted or trapped as junk or spam.

7. Participants are expected to supply via the Geo*PT* website analytical data and information about their procedures that are correct to the best of their knowledge, and to ensure that the data have been obtained and reported in the manner requested in the *Instructions to Analysts*. Participants are requested to check that the results they have recorded on the system are correct before submitting them. Should there be any apparent discrepancies in the data recorded in the report, participants should notify us immediately. If incorrect data have been submitted, *z*-scores can be provided for the revised data. However, results cannot be corrected in the report produced for any round as the reporting of data is considered to be part of the proficiency test.

8. Participants are expected to adhere to good laboratory practice when analysing the test samples provided by Geo*PT*. In particular, test samples should be handled with appropriate care in respect of health and safety that is compatible with geological samples. This includes taking precautions against the inhalation or ingestion of dust when handling test samples, and taking precautions to avoid inhalation of fumes when fusing or roasting samples. Samples identified as being derived from mining operations or tailings sites or certain categories of environmental sample could contain significant amounts of toxic elements that are hazardous to health. Some samples, e.g., those containing sulfide or arsenide minerals, may attack platinum-ware or other laboratory equipment. IAG cannot accept responsibility for any damage or misadventure occurring when handling or processing the test samples.

The International Association of Geoanalysts, May 2020

References

1 M Thompson, S L R Ellison and R Wood, 2006. The international harmonised protocol for the proficiency testing of analytical chemistry laboratories'. (IUPAC Technical Report) *Pure Appl. Chem*., **78**, 145-196.

2 M Thompson, P C Webb, P J Potts and S Wilson, 2018. The stability of 57 consensus values in a proficiency test material re-issued blind after an interval of 18 years. *Analytical Methods*, **10**, 1547.

3 P J Potts, M Thompson and P C Webb, 2015. The Reliability of Assigned Values from the GeoPT Proficiency Testing Programme from an Evaluation of Data for Six Test Materials that have been Characterised as Certified Reference Materials. *Geostandards and Geoanalytical Research*, **39**, 407-417.

4 ISO Guide 43, 1997. Proficiency testing by interlaboratory comparisons, *ISO*, Geneva.

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

6 ISO/IEC 17043, 2010. Conformity assessment - General requirements for proficiency testing, *ISO*, Geneva.

7 M Thompson, P C Webb and P J Potts, 2014. The Geo*PT* Proficiency Testing Scheme for Laboratories Routinely Analysing Silicate Rocks: A Review of the Operating Protocol and Proposals for its Modification. *Geostandards and Geoanalytical Research,* **39**, 433-442.

8 ISO 13528: 2015. Statistical methods for use in proficiency testing by interlaboratory comparison, *ISO*, Geneva, 2015.

9 M. Thompson, 2006. The variance of a consensus. *Accred. Qual. Assur.*, **10**, 54-575.

10 M. Thompson, 2017. On the role of the mode as a location parameter for the results of proficiency tests in chemical measurement. *Analytical Methods*, **9**, 5534–5540.

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

12 M Thompson, 2000. *Analyst*, **125**, 385-386.

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

14 M Thompson, 2015. Is your 'homogeneity test' really useful? *Anal. Methods*, **7**, 1627.

15 P J Potts, P C Webb and M Thompson, 2015. Bias in the Determination of Zr, Y and Rare Earth Element Concentrations in Selected Silicate Rocks by ICP-MS when Using Some Routine Acid Dissolution Procedures: Evidence from the Geo*PT* Proficiency Testing Programme, *Geostandards and Geoanalytical Research*, **39**, 3, 315-327.

Additional references relevant to the GeoPT™ programme

Analytical Methods Committee. z-Scores and other scores in chemical proficiency testing—their meanings, and some common misconceptions. *AMC Technical Briefs* No. 74. (*Anal. Methods*, 2016, 8, 5553)

Analytical Methods Committee. Fitness for purpose: the key feature in analytical proficiency testing. *AMC Technical Briefs* No. 68. (*Anal. Methods*, 2015, 7, 7404)

Analytical Methods Committee. The amazing Horwitz function. *AMC Technical Briefs* No. 17.

Analytical Methods Committee. Proficiency testing: assessing z-scores in the longer term. *AMC Technical Briefs* No. 16.

Analytical Methods Committee. The J-chart: a simple plot that combines the capabilities of Shewhart and cusum charts, for use in analytical quality control. *AMC Technical Briefs* No. 12.

Analytical Methods Committee. Understanding and acting on scores obtained in proficiency testing schemes *AMC Technical Briefs* No. 11.

Analytical Methods Committee. Robust statistics: a method of coping with outliers. *AMC Technical Briefs* No. 6.

(*AMC Technical Briefs* are short articles covering a wide range of technical issues affecting the analytical chemist. They are produced by the Analytical Division of the Royal Society of Chemistry and can be downloaded gratis from the website www.rsc.org/amc. From issue No. 50, they can also be downloaded gratis from the website of the RSC journal *Analytical Methods.*)