Selection, use and interpretation of proficiency testing (PT) schemes by laboratories - 2000

This document has been produced by the Eurachem Nederland, task group "Proficiency testing schemes" (part of the working group on "Interlaboratory Studies"), and the Laboratory of the Government Chemist (LGC), United Kingdom.

Production of this Guide was in part supported under contract with the UK Department of Trade and Industry as part of the National Measurement System Valid Analytical Measurement (VAM) Programme.

Comments are invited on this document and should be sent to the editor:
Mr Nick Boley
Convenor, EURACHEM Proficiency Testing Mirror Group
LGC, Queens Road, Teddington, Middlesex, TW11 0LY, United Kingdom
☎Int +44 20 8943 7311, 🗐: Int +44 20 8943 0654, 💻🖃: npb@lgc.co.uk

English Edition 1.0, 2000

Ruling language

The text may be freely translated into other languages, but where such action results in a dispute over interpretation, the guidance given in this English version is taken as being the definitive version.

Copyright of text

Contents

Authors

Eurachem Nederland, task group "Proficiency testing schemes"

(part of the working group on "Interlaboratory Studies")

- Mr. G. Counotte: Gezondheidszorg voor Dieren, Deventer
- Mrs. D. Van Dijk: Wageningen Agricultural University, Wageningen
- Mrs. D.C. Van Loenen-Imming: KEMA, Arnhem
- Mr. W. Oussoren: Institute for Interlaboratory Studies, Dordrecht
- Mrs. B. Van der Vat: TNO Nutrition and Food Research Institute, Zeist
- Mr. R.G. Visser: Institute for Interlaboratory Studies, Dordrecht

Laboratory of the Government Chemist (LGC), Middlesex, United Kingdom

- Mr. N. Boley
- Mr. J. Day

Dr. R. Walker

Summary

This document represents the current (2000) state-of-the-art with respect to the selection and use of proficiency testing schemes, and the interpretation of results and evaluations given in proficiency testing schemes. Although this is aimed primarily at staff in analytical laboratories, it is also useful for customers of laboratories, assessors working for accreditation bodies and other external users of PT scheme results.

The correct use and interpretation of PT scheme results is complex. This guide is designed to enable users of these results to gain a better understanding of proficiency testing, and therefore how to use the results it produces in a more sophisticated and appropriate manner. It gives guidance to laboratory staff at all levels on how to place results in context, and how to use the results to gain an overall picture of the quality of performance.

Proficiency testing is a quality assurance tool which is evolving very quickly. Therefore this document should be viewed as a “living document” which will be subject to continuous revision over the next few years. As developments occur, the document will be revised, and updated versions made available subsequently.

Samenvatting

Dit document beschrijft de huidige (2000) 'state-of-the-art' situatie met betrekking tot de keuze en het gebruik van 'Proficiency Testing' (PT) schema's en de interpretatie van de resultaten en evaluaties gegeven in proficiency testing schema's. Hoewel het document zich met name richt op het management in analytische laboratoria is het ook bruikbaar voor klanten van laboratoria, assessoren die werkzaam zijn bij accrediterende instanties en andere externe gebruikers van PT schema resultaten.

Het correcte gebruik en de interpretatie van PT schema resultaten is complex. Deze handleiding is geschreven om gebruikers van de resultaten een beter inzicht te geven in proficiency testing. Dit om zo het juiste gebruik van de PT resultaten te bevorderen. De

handleiding geeft een handreiking aan het het gehele laboratoriummanagement over hoe de resultaten in de juiste context geplaatst moeten worden en hoe resultaten gebruikt kunnen worden om een totaalbeeld van het presteren van het laboratorium te verkrijgen.

Proficiency testing schema's vormen één van die gereedschappen van kwaliteitsborging ('quality control', QC) die zich zeer snel ontwikkelen. Dit document moet daarom gezien worden als een levend document waarvan de komende jaren regelmatig een revisie zal verschijnen. Zodra zich nieuwe ontwikkelingen voordoen zal het document worden bijgewerkt en de ge-update versies zullen daarna beschikbaar worden gemaakt.

1. INTRODUCTION AND SCOPE

1.1 Background and motivation to this document

A regular independent assessment of the technical performance of a laboratory is recommended as an important means of assuring the validity of analytical measurements, and as part of an overall quality strategy. A common approach to this assessment is the use of independent proficiency testing (PT) schemes. A PT scheme is a system for objectively evaluating laboratory results by external means, and includes regular comparison of a laboratory's results at intervals with those of other laboratories. This is achieved by the scheme co-ordinator regularly distributing homogeneous test samples to participating laboratories for analysis and reporting of the data. Each distribution of test samples is referred to as a round. The main objective of a PT scheme is to help the participating laboratory to assess the accuracy [14] of its test results. In addition, participation in an appropriate PT scheme is recommended for laboratories seeking accreditation to the standard of EN45001. Indeed, in some sectors such participation is mandatory.

This document is intended to give guidance on the selection and use Proficiency Testing (PT) schemes for laboratory analysts, senior managers and quality managers in participating laboratories. These guidelines aim to explain how PT scheme data is produced and how such data should be interpreted to give an objective picture of the performance of individual participating laboratories, a laboratory within its peer group and with respect to the PT scheme as a whole.

The large volume of data produced by many PT schemes can be particularly daunting for many users of that data, particularly those with a relatively limited technical background. In this situation it is easy to misinterpret or over-interpret the data, which can often give a misleading picture of the performance of participating laboratories. It is therefore essential that interpretation of proficiency testing scheme data is carried out in an appropriate and balanced manner.

Although primarily aimed at laboratories producing data, because of the increasing interest in PT performance being shown by the users of laboratory services, this document also provide some guidance from that perspective. This additional viewpoint is included as an Appendix to the main body of the report.

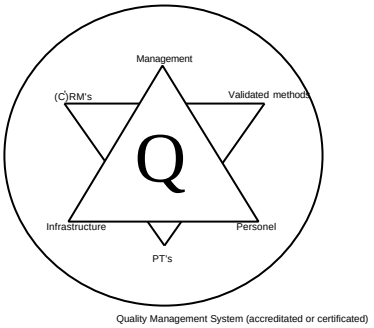
Although most PT schemes are concerned with quantitative analyses, there are also qualitative schemes (biological, etc.). However, this report is concerned principally with quantitative schemes. Qualitative schemes do not use the same statistical basis, and so some of the principles outlined in this report do not apply. However, the same degree of caution, and the placing of individual results in the overall context of the data should be applied to the interpretation of the data from qualitative schemes.

Proficiency Testing schemes are operated for the benefit of participating laboratories. However, other organisations are developing a legitimate interest in PT schemes. Interested organisations include accreditation bodies, regulatory authorities and customers of analytical laboratory services. It is important for scheme co-ordinators to bear in mind the needs of these organisations in order that they are able to use the data from PT schemes to aid their understanding of the capabilities and competence of laboratories (See Appendix 4).

2. BACKGROUND AND INTRODUCTION TO PROFICIENCY TESTING

Note: Some readers may be familiar with the subject of proficiency testing; this section has been included for the benefit of those readers whose background knowledge of proficiency testing is more limited.

2.1 Quality management system



One should never forget that proficiency testing (PT) schemes are just one of the tools in quality assurance and quality control. Other tools, for example certified reference materials (CRM's) and validated methods, address different topics of the quality management system and are therefore also of major importance.

2.1.1 External assessment

Because of the importance of producing analytical data fit for purpose, it is now increasingly necessary for a laboratory not only to produce such data, but also to demonstrate the accuracy and comparability of its data by some form of external assessment. Such a demonstration of a laboratory's competence is known as external quality assessment and there are two main ways by which a laboratory can do this, which are essentially complementary in nature.

The first is by physical inspection of the laboratory to ensure that its quality system procedures comply with well-established, recognised standards - in other words third party assessment or accreditation.

The second, known as proficiency testing, is by the assessment of its performance in interlaboratory comparisons using samples which have been distributed by the scheme's organiser. (Other names can be used to describe proficiency testing include quality assessment, external quality assessment or EQA and performance testing.)

2.2 Proficiency testing schemes: an introduction

2.2.1 Aims of proficiency testing schemes

Essentially a proficiency testing (PT) scheme checks the competence of participating laboratories by a statistical evaluation of the data they obtain on analysing centrally distributed materials. Each laboratory is then provided with a numerical indicator of its competence - a performance index or score - together with information on the performance of the group as a whole, enabling its proficiency relative to the group to be compared and evaluated. Participation in a proficiency testing scheme also reinforces an interest in quality assurance and provides the basis for any corrective action in those laboratories whose data does not meet the required level of acceptability.

Although the overall aim of proficiency testing is to encourage good performance, within the scope of this general aim a successful scheme must provide certain types of information for both the participants and the organisers.

Firstly, it must enable a laboratory to compare its performance at a particular time against an external standard of performance. How accurate is the data? It must also enable a laboratory to compare its performance at a particular time with its performance in the past. Is it getting better or worse? It must enable a laboratory to compare its performance with that of other laboratories at a particular time. Within its peer group, how well does it perform? It must enable the organisers to identify the participants whose performance is unsatisfactory. This is particularly important for regulatory authorities. Finally, it must enable the organisers to see whether there is a general improvement in performance with time. Is the scheme efficacious? Is it doing its job of improving the quality of chemical measurements.

2.2.2 Types of proficiency testing schemes

Over the years proficiency testing schemes have been introduced under a number of different circumstances, some are 'open' to any laboratory, usually for an annual fee, others are 'closed' or by invitation only. Regardless of the circumstances, there are basically two main types of scheme.

1. Those set up to measure the competence of a group of laboratories to undertake a very specific analysis, e.g. lead in blood or the number of asbestos fibres on membrane filters.
2. Those where there is a need to judge the competence of a laboratory across a field or type of analysis (e.g. trace metal analysis by atomic absorption spectroscopy or the analysis of food components by HPLC). Because of the wide range of possible analyte/matrix combinations it is not practicable to apply comprehensive testing and a representative cross-section of analyses is therefore usually chosen.

Each of these two main types of proficiency testing schemes can be sub-divided into three further categories:

1. Where the sample to be tested is circulated successively from one laboratory to the next. In this case the sample may be returned to a central laboratory sometimes before being passed on to the next testing laboratory in order to determine whether any changes have taken place to the sample (a less common approach - but may be necessary - e.g. quartz dust on membrane filters, HSE).
2. Where randomly selected sub-samples from a bulk homogeneous supply of material are distributed simultaneously to participating laboratories - by far the most common type of PT scheme.
3. Where samples of a product or a material are divided into several parts with each participating laboratory testing one part of each sample. This is frequently referred to as 'split sample' testing.

2.2.3 Other types of interlaboratory tests

We should also note that there are other types of interlaboratory comparisons which are distinct from proficiency testing, but which also have important roles to play in terms of quality assurance in general:

Interlaboratory Test	**Description**
proficiency testing	continuing assessment of technical competence
collaborative study	validation of a specific method
certification study	establishing the best estimate of the true value of an analyte in a reference material;
co-operative study (also known as a ring test or round-robin exercise)	laboratory assessment on a one-off basis,

3. SELECTION OF PT SCHEMES

Participating in a proficiency testing scheme provides a laboratory with an objective means of assessing and demonstrating the reliability of the data it produces. It thus supplements a laboratory's own internal quality control procedures by providing an additional external measure of its testing capability. Participation to proficiency tests is strongly advised for laboratories accredited according to ISO Guide 25, EN 45001 and EN 45003 and even considered mandatory in some cases. Thus, a laboratory operating under a (certified or accredited) quality system should (voluntarily) participate in appropriate proficiency tests.

Before, during and after participating in a proficiency testing scheme one has to think about all relevant aspects. This includes the selection of PT's and how to deal with its samples, its reports and actions to be taken on (deviating) results. Procedures in which this is formalised can be of great help.

A laboratory willing to participate in proficiency tests has to answer the following two questions:

- Does any proficiency testing scheme exist for the tests and samples my laboratory (usually) analyses?
- Is the organised proficiency testing scheme relevant for me?

3.1 How to find out if a specific proficiency testing scheme is organised?

Despite the fact that numerous institutes organise international proficiency tests for many years now, a summary of PT-organisers and/or the samples offered, is not available. Information on PT-organisers and/or the availability of proficiency tests can be found by:

- National accreditation bodies, who should be well informed about the existence of (international) proficiency tests.
- Peer laboratories who already participate in proficiency tests or do know about relevant PT's.
- The PT-organisers in ones own country, who probably will also have (summarised) information about the proficiency tests of other organisers.

- A search on the internet, which will probably provide the most actual information. One should include synonyms as “interlaboratory comparison”, “laboratory evaluating study”, ”External Quality Assessment (EQA) Schemes”, “proficiency test”, “cross check program”, “correlation program” and/or “ringversuch”.

A European project to inventory, summarise and present European proficiency tests is being conducted. At this moment however, no results are available yet (Concerted Action under the programme Standards, Measurement and Testing (SMT) "Information system and qualifying criteria for proficiency testing schemes", to inventory, summarise and present proficiency testing schemes is being conducted in the frame of a concerted action under the Standards, Measurement and Testing programme (Project SMT-4-CT-98-8002). This information is now available on www.eptis.bam.de. The project is managed by BAM, the German Federal Institute for Materials Research and Testing, in Berlin.

3.2. How to decide if a proficiency testing scheme is relevant?

First of all, one should be aware of the fact that participating in a proficiency testing scheme is almost always more beneficial than not participating at all. Furthermore, at this moment, the number of proficiency testing schemes is limited and often one has little choice of schemes.

If a number of similar proficiency testing schemes are organised and a choice has to be made, one should remember that a proficiency testing scheme with a perfect fit for its own laboratory rarely exists. Therefore, in practice, one should choose the PT best fit for use.

In order to decide if a proficiency testing scheme is relevant, one should compare the situation in the proficiency testing scheme with the routine situation in its own laboratory. Numerous topics can be addressed and the laboratory itself has to decide which are (most) relevant. It is the responsibility of the laboratory itself to decide about the topics to be addressed, to make the comparison and to judge the relevancy of the PT. A simple matrix can be helpful to systematically perform this process and the recording of it (see appendix 1). If the situation in the proficiency testing scheme and the laboratories is sufficiently comparable, one should seriously consider participating.
A number of PT-organisers allow participation in just one round. If one is not fully convinced of the relevancy of the proficiency testing scheme this is a good option. Sometimes old samples of PT's can be bought together with the PT report. This also is a good option to judge the relevancy of a specific proficiency testing scheme.

4. INTERPRETATION OF PT REPORTS

There are some basic points about the interpretation of PT results which it is worth stating before more detailed consideration of this topic is given. Proficiency testing is not about “passing” or “failing” a test; it is about taking part and learning from the results. One good round for a laboratory, where all measurements have produced satisfactory performance does not necessarily make that laboratory good. Consistent performance at this level is the goal. Neither, on the other hand, does one bad result in any round make a laboratory bad; this result needs to be studied and lessons learned from it so that it is not repeated.

4.1 Statistics

One of the basic elements in all proficiency test is the evaluation of the performance of each participant. In order to do so, the PT organiser has to establish two values: (1) the assigned value of the test material and (2) the acceptable range. Different methods can be used to establish these estimates [15, 16]. There is no standardised protocol, which describes or even prescribes in detail the strategies to be used. So, organisers make their own decisions and have developed their own procedures. The reader should be aware of the fact that - unfortunately - there is no single “best” approach for all situations. In general, one should trust the organiser of the proficiency testing scheme for their choice(s). Details about the decisions and procedures can normally be found in the protocol of the PT organiser.

There are essentially three methods available to obtain the assigned value, a working estimate of the true value:

1. The addition of a known amount or concentration of analyte to a base material containing none. This method is satisfactory in many cases, especially when it is the total amount of the analyte rather than the concentration that is subject to testing - but, of course, it may not simulate the difficulty of normal sample preparation procedures where recovery [which include, *inter alia*, extraction and speciation] problems may well arise.

2. The use of a consensus value produced by a group of expert or referee laboratories using best possible methods. This is probably the closest approach to obtaining true values for the test materials, but it may well be expensive to do so. Another problem is that it is often hard or even impossible to find a group of expert or reference laboratories which expertise is beyond doubt and accepted by all participants of the Proficiency Test. This is even more true for large, international tests with participants of many countries.

   For a number of analyses, the true value is, in principle, defined by the method used. In these cases, the expert or referee laboratory should all use the same method and should follow it in every detail. The expert laboratories should be declared before the proficiency test.

3. The use of a consensus value, produced in each round of the proficiency test, and based on the results obtained by the participants. The consensus is usually estimated as the mean of the test results after outliers have been rejected, but other techniques do exist also (see par. 3.1.1 and 3.1.2). The consensus approach is clearly the cheapest. There might be difficulties because there may not be a real consensus among the participants or the consensus may be biased because of the general use of inadequate methodology - neither of which is rare, for example, in the determination of trace constituents in natural matrices. When using natural matrix samples, this approach to establish a working estimate of the true value is often the only way. Another reason to choose this approach is an economic one. The first two procedures can be very costly. If (potential) participants, who pay for participation, do not want to pay for a proficiency testing scheme with high metrological aspects, than this latter approach should be used; it is better to have such a PT scheme, than no scheme at all.

   The choice between a consensus either from expert laboratories (method 2) or from the participants (method 3) depends on whether the aim of the proficiency test is to encourage the production of true results or merely to obtain agreement among the participants. In spite of the extra cost, method 2 is preferred if possible, because it is only by this approach that the scheme can hope to form part of the wider chemical measurement system within which comparable, traceable measurements are

desirable. For example, we can imagine a PT scheme using the consensus approach to determining the true value - method (3) - as a closed system within which there is an overall systematic bias in the results. Over time, the participants have learned to compensate for this bias in order to achieve good performance indices - and as a result build in the same bias for all their routine samples. Clearly, it would be impossible for a laboratory within the scheme to compare its results with a laboratory outside the scheme and hope to achieve comparability. If, on the other hand, the organisers of the PT scheme are attempting to obtain the best estimate of the true value, then it follows that comparability of measurement between separate closed chemical systems should follow.

A common way, at present, to establish the assigned value and the acceptable range, is the use of the participants PT results to calculate both values. A number of statistical methods have been developed to calculate the location and/or the spread of the PT results [1]. The statistics applied can be divided in two different categories: parametric statistics and non-parametric / robust statistics.

4.1.1 Parametric Statistics

Parametric statistics are the more commonly known approach. The (arithmetic) mean is used as the estimate of location of the test results and the standard deviation as an estimate of the spread.

This type of statistics requires a normal (Gaussian) distribution of test results. If this requirement is not fulfilled, the mean and standard deviation cannot be trusted as good descriptions of the test results. Therefore, before using parametric statistics, the presence of a normal distribution has to be proved. A number of test are available to do so (e.g. Kolmogorov-Smirnov). Despite its name, a normal distribution does not imply that it belongs to the normal situation. Numerous valid test results of day-to-day (chemical) measurements have shown an anormal distribution!

One of the characteristics of parametric statistics (e.g. the mean and standard deviation) is the great sensitivity for deviating results. Therefore, the use of classical statistical

techniques usually requires the application of outlier tests (e.g. Grubb's, Dixon's or Cochrane's tests) to remove the influence of deviating results (e.g. outliers, stragglers) before the mean and standard deviation are calculated.

Example 1: Mean and standard deviation are (very) sensitive to deviating results.
When twenty laboratories produce the test results beneath a mean of 12.52 and a standard deviation of 0.54 would be calculated. If just one laboratory made a writing error and reported 122 instead of 12.2 the mean would increase to 18.01 (+44%) and the standard deviation would increase to 24.48 (+400%). Apparently, both mean and standard deviation are extreme sensitive to deviating results.

test results	*12.20*	*12.5*	*12.3*	*12.2*	*11.9*	*11.6*	*11.4*	*12.4*	*12.6*	*13.2*
	13.20	*13.2*	*12.3*	*12.8*	*12.2*	*12.7*	*13.4*	*12.7*	*12.5*	*13.0*
n	*20*									
mean	*12.52*									
st.dev.	*0.54*									
test results	*122.0*	*12.5*	*12.3*	*12.2*	*11.9*	*11.6*	*11.4*	*12.4*	*12.6*	*13.2*
	13.20	*13.2*	*12.3*	*12.8*	*12.2*	*12.7*	*13.4*	*12.7*	*12.5*	*13.0*
n	*20*									
mean	*18.01*									
st.dev.	*24.48*									

Numerous tests have been developed to detect deviating results; how to deal with those results (outliers, stragglers) has been topic of discussion for many, many years now and an overall accepted procedure can not be given. Some organisations however have defined a procedure [13, 14, 17, 18].

4.1.2 Non-parametric and robust statistics

Extensive and complete descriptions of non-parametric and robust statistics have been published, but in general these are not advised to non-experts [10, 11]. A number of publications however, can be understood also by those not experienced in statistics and/or chemometrics [2, 3]. Some publications discuss the use of non-parametric and robust statistics in proficiency tests. [4, 5, 6]

Non-parametric and robust statistics are also valid with test results that do not have a normal (Gaussian) distribution. So, unlike parametric statistics, the presence of a normal

distribution is not required. This is advantageous in proficiency tests as anormal distributions are frequently encountered.

In contrast to the commonly used statistics (parametric statistics), deviating test results (outliers, stragglers) do not have a (great) influence on the estimate of the location and spread of the test results. Therefore, outlier techniques are not required. This is generally seen as one of the great advantages of these kinds of statistics.

There is not one estimator of location or spread, but a whole collection. In general, the estimators give all about the same value, but (slight) variations may exist The most common non-parametric counterpart of the mean is the median, the middle value of the sorted results (or in case of even number of results, the mean of the two middle values). One of the non-parametric estimators of the standard deviation is the median absolute deviation (MAD), which is given by the median of all absolute deviations of each test result and the median: MAD = median(|xi - median|). To confer equivalence to the standard deviation of a normal distribution, the MAD has to be multiplied by 1.483. This value is known as MADe.

Example 2: Median and MAD are (very) insensitive to deviating results.

When twenty laboratories produce the test results beneath a median of 12.50 and a MAD of 0.30 would be calculated. If just one laboratory made a writing error and reported 122 instead of 12.2 the median would increase to 12.55 (+1%) and the standard deviation would increase to 0.35 (+17%). The median and MAD are - in comparison to the mean and the standard deviation (see example 1) - much less sensitive to the presence of deviating results.

test results	*12.20*	*12.5*	*12.3*	*12.2*	*11.9*	*11.6*	*11.4*	*12.4*	*12.6*	*13.2*
	13.20	*13.2*	*12.3*	*12.8*	*12.2*	*12.7*	*13.4*	*12.7*	*12.5*	*13.0*
n	*20*									
median	*12.50*									
MAD	*0.30*									
test results	*122.0*	*12.5*	*12.3*	*12.2*	*11.9*	*11.6*	*11.4*	*12.4*	*12.6*	*13.2*
	13.20	*13.2*	*12.3*	*12.8*	*12.2*	*12.7*	*13.4*	*12.7*	*12.5*	*13.0*
n	*20*									
median	*12.55*									
MAD	*0.35*									

Remark: The mean and median are both estimators for the location of the test results and can be compared directly. The standard deviation and the MAD are both estimators for the spread in the test results, but can not be compared directly. For a fair comparison of the standard deviation and the MAD, the latter has to be divided by 0.6745 (for this example: standard deviation=0.54 or 24.84 MAD/0.6745=0.44 or 0.52).

More sophisticated robust procedures, calculate the robust estimate of location and spread together. One of these is “Huber proposal 2”, also known as “H15” [4, 5]

4.2 Split Level

In the case of split-level studies (sample I and II are similar) one can obtain information about systematic errors from the results of the participants by presenting the results in a Youden plot [19, 20]. The underlying rationale is that in any one laboratory, the systematic errors will be the same for each of the two samples. The difference between the two results will thus be free of systematic errors and will reflect only the true difference between the samples and within-laboratory random errors.

A Youden plot is a graphical presentation of the results of the participants by presenting the results of sample I and sample II in a XY-diagram (see appendix 2 for two typical examples).

When there are no systematic errors random errors will be expected to cause the scattering of the results of the participants. If that is the case the results of the participants should be equally divided in four groups.

The circle represents the equal division of the results over the four groups. Most times, however, the results are dominantly both low or both high due to systematic errors. This results in the circle becoming an ellipse around the 45° line. If random errors become extremely small the ellipse will become more and more elongated.

Participants in split-level studies can get the following information from the described presentation:

- graphical presentation of the performance of the participant towards the performance of the other participants
- impression of the repeatability and reproducibility

4.3 z scores

One of the basic elements in all proficiency tests (PT) is the use of a performance indicator to quantify the analytical performance of each participant [15, 16]. The z score is frequently advised as such performance indicator. The z score is a measure of the deviation of the result from the assigned value for that determinand and is calculated as:

$$z = (x_i - X)/\sigma$$

where σ is a standard deviation which is chosen either as an estimate of the actual variation between results encountered in a particular round of the scheme or a set target value representing the maximum allowed variation consistent with achieving valid data.

The main assumption in using the z score is that the individual z scores will have a Gaussian or normal distribution with a mean of zero and a standard deviation of one. On this basis analytical results can be described as 'well-behaved'. A common classification based on z scores can be made:

$z < 2$	is considered Satisfactory
$2 \leq z \leq 3$	is considered Questionable
$z > 3$	is considered Unsatisfactory

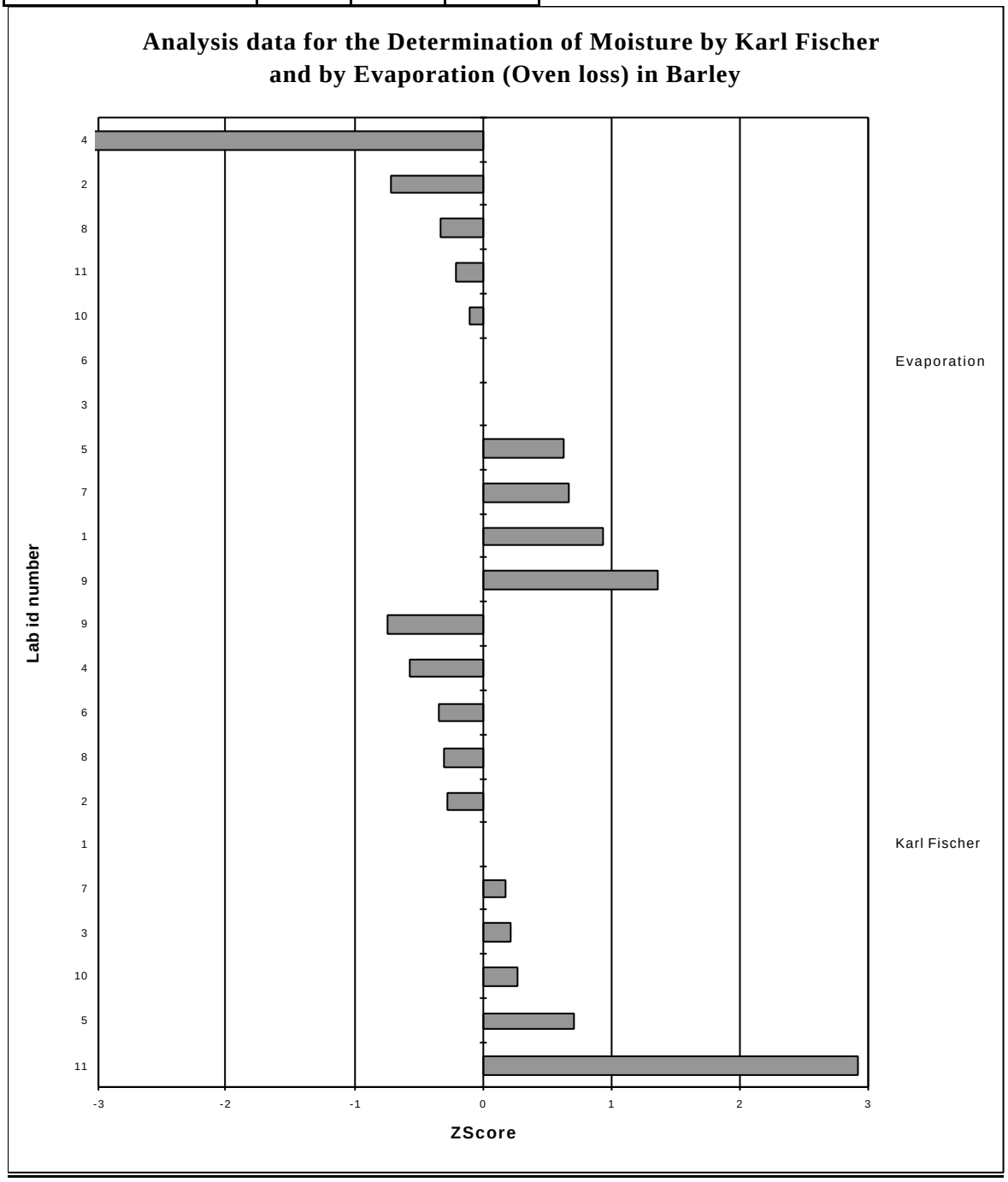
An alternative scoring, known as Q-scoring, is also sometimes used, but is based not on the standardised value, but on the relative bias. This type of score tends to be used when the participants in a scheme may have wide standards of performance and there is no basis for a common value of s.

In the majority of proficiency testing schemes a z score of between -2 and +2 is considered satisfactory, and a z score of greater than +3, or less than -3, is considered unsatisfactory. A z score of between +2 and +3 or -2 and -3 is usually considered as “Questionable”. Although some schemes may adopt slightly different definitions of the z score, in general a z score of 2 indicates a result which is 2 standard deviations from the assigned value for the chosen analyte. Normal variations due to the statistics of sampling mean that, for any laboratory, one result in 20 should generally be expected to lie between +2 and +3 or -2 and -3, and one in 100 should be greater than +3 or less than -3, even when the analysis is being carried out competently. Examples of z score plots are shown in Figure 1.

Figure 1: An Example of a z score Plot from a PT Scheme

Moisture

	Total	KF	Evap
No of Results	22	11	11
Robust Mean	26.25	25.3	27.2
Robust SD	1.41	0.65	1.33



5. LABORATORY ACTIONS

5.1 Evaluation

A participant that evaluates his performance in a proficiency test, will be interested in two things: First, he wants to know if he should undertake a corrective action. And second, when his laboratory already participated in one or more earlier rounds, he wants to compare his present performance with that in the other round(s). These two topics are addressed in the paragraph 5.1.2 (short term evaluation) and paragraph 5.1.3 (long term evaluation).

First of all it should be considered, that the calculations used by the organiser of a proficiency test to determine the performance of the participants (usually z scores), may not always be “fit for purpose”.

In all cases, the results of a laboratory should be studied in conjunction with the information of internal quality control. Intralaboratory (internal) quality control is a continuing and systematic in-house regimen intended to ensure the production of analytical data for continuing high validity. Proficiency testing enables the participants to assess the continuing capability and relative performance.

5.1.1 Short Term Evaluation

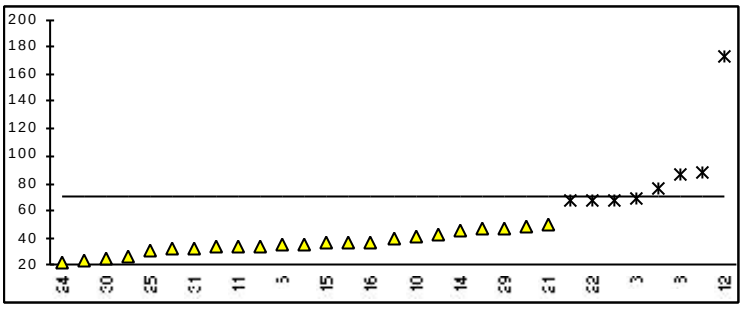
To decide whether or not corrective actions should be undertaken and in the case the participation was unique (no earlier participation(s)), the participant may use the judgement of the proficiency test-organiser or may decide to recalculate the performance score.

When recalculating the scores, one should consider the following options:

- use the mean of participants that used the same analytical method (see example 1)
- use the claimed or target value for uncertainty (this is applicable when the specific test is critical for the laboratory. In that case, the participant should judge if the standard deviation used in the performance evaluation, was acceptable. When the standard deviation is not acceptable, the participant should recalculate the performance indication using a more acceptable standard deviation)

- use the claimed or target value for uncertainty combined with the uncertainty or standard deviation of the assigned value
- correction of the results for recovery (this is applicable when the concentration is known by formulation and the recovery of the method used is less than 100%, e.g. Aldrin for the EOX determination); the uncertainty/standard deviation of the recovery should be low or added to the total uncertainty/standard deviation.

Example 1: Aluminium in white cabbage on CRM (69.0 ± 14 mg/kg)



The usual conclusion would be that eight high values are outliers. However in this round a CRM was used and the true value was known. Thus the conclusion was that only seven labs reported correct values. After investigation, the differences were explained by the different methods used by the participants. The labs that reported the high values used HF for digestion or used a non-destructive method (INAA or XRF), while the other labs used digestion procedures that did not dissolve the silicates in the sample.

5.1.2 Long Term Evaluation

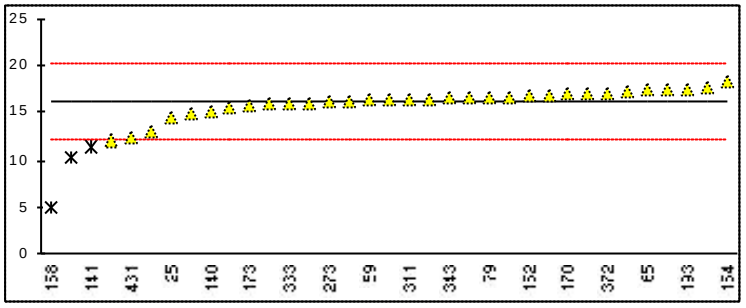
If a laboratory already participated in one or more earlier rounds, he wants to compare his present performance with that in the other round(s). In this case the laboratory wants to evaluate performance in time. Is ones performance improving during the last rounds or is it decreasing ?

To conclude the direction of progress in time from the data of subsequent interlaboratory studies, one has to have comparable data. However, data from the same test from two different PT's are often not directly comparable. The z scores are usually calculated, using a standard deviation, which is calculated from the round. This standard deviation is likely to vary from round to round, depending on the group of laboratories that participated and the difficulty of the sample used (see examples 2a and 2b, where the standard deviation varies from 0.58 - 1.5). However, when a fixed standard deviation is used for calculation of the z scores, the z scores from 2 different rounds are comparable and progress in time is made visible quite easily (see examples 2c and 2d) and quite different conclusions may be drawn.

Of course, it may be most convenient for the participants if the organiser of the proficiency test already calculated z scores, using target standard deviations. But when

this is not the case, the participant may calculate target z scores itself using a selected target standard deviation. As standard deviation, target values from literature (e.g. from a standard method like ISO, EN, ASTM, DIN) can be used. In case no literature values are available, own criteria may be chosen depending on the goal of the participation to the PT or the importance of the test (any realistic target value can be used, e.g. 10% of the consensus value). Note that the selected standard deviation does not have to be constant, but that it may be concentration dependent.

Example 2: 2 Rounds in 1996 & 1997: determination of Conradson Carbon Residue in Fuel Oil



Example 2a:

1996 round, using st.dev of the group of participants (= 1.5)

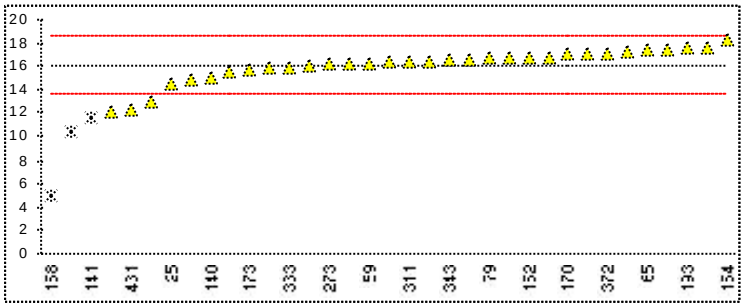
Lab 431: z score = -2.74

Example 2b:

1997 round, using st.dev of the group of participants (= 0.58)

Lab 431: z score = -2.08

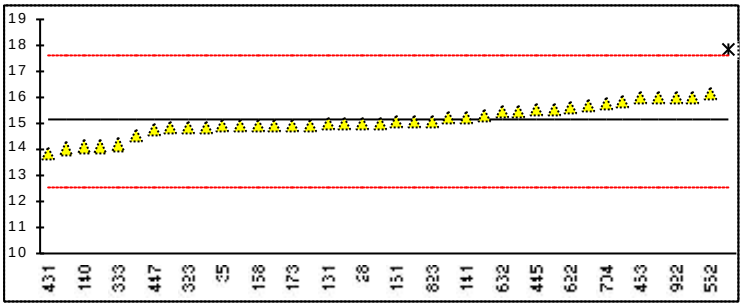
Conclusion: The performance of lab 431 improved only slightly and is still unsatisfactory



Example 2c:

1996 round, using st.dev of ASTM D189 (= 0.89)

Lab 431: z score = -4.45



Example 2d:

1997 round, using st.dev of ASTM D189 (= 0.89)

Lab 431: z score = -1.44

Conclusion: The performance of lab 431 improved to an acceptable level

Rolling Performance Indicators or Scores (RPIs) are used in a number of PT schemes to assess the performance of participants over a number of rounds. A Rolling Performance Indicator is a measure of PT scheme performance for an individual laboratory over a number of consecutive rounds. These may be designed to measure bias (constant deviation from the assigned value) and precision. A number of mechanisms for calculating and expressing these are used. The laboratory needs to be aware that one unsatisfactory result can distort an RPI value over a period of time until that result is no longer used in any calculation. However, some PT scheme participants graphically plot their performance scores from round to round in order to assess performance over time. This approach is very useful enabling unusual or unexpected results to be highlighted, as well as assisting in the identification of trends. A laboratory's internal quality control (IQC) procedures would normally be expected to identify trends associated with, for example, improper instrumental calibration or maintenance, or use of reagents. Monitoring PT performance over time acts as a backup system for this.

There are a number of RPIs which can be used, including RSZ, SSZ and RSSZ, which are used with z scores.

Any RPI should ideally look at single measurands over time, although they could take into account groups of like measurands (e.g. trace metals determined by ICP-OES). However, the value of RPIs based upon a basket of different measurands is limited and, in many cases, may have, at best, no value at all.

5.1.3 Overall performance

Laboratories should also study carefully the overall performance of all participants in the scheme. Unsatisfactory performance in the context of a round where the majority of participants performed to a satisfactory level should be contrasted with unsatisfactory performance where a significant number of participants performed poorly. Both situations should be viewed seriously, since both indicate problems regarding the methodology used for the test sample supplied. However, in the latter situation, the robustness, specificity, linearity and/or validity of the method should be investigated as well as the potential effect of interferences. The possibility that the criteria for satisfactory performance was incorrectly set should also be considered. In the former situation, it is more likely that the method followed was appropriate, but it had been incorrectly applied by the analysts during the proficiency test.

5.2 Corrective Actions

If the participant can assess its performance as unsatisfactory, corrective action should be made. The following list can provide guidance to establish at which point remedial action needs to be taken:

- after one unsatisfactory result or z score on one round if the analysis is key to the laboratory's business
- if at least three unsatisfactory results or z scores are recorded in one round, particularly if the analyses are linked
- after at least two consecutive unsatisfactory or questionable results or z scores for any analysis
- if results for any analysis show a clear trend towards unsatisfactory performance with time over a number of rounds
- if results for any analysis over a number of rounds show clear and consistent bias, of at least one standard deviation above or below the assigned value, even though results are satisfactory.

At the very least the laboratory should identify and document the problem, and decide whether corrective action is necessary.

However, before taking action the problem should be analysed thoroughly. A good procedure consists of several steps:

- analyse the quality problem based on the result of successive interlaboratory studies, internal quality control data and record the relevant measurements
- make a plan for corrective action
- execute and record the corrective action
- check whether the corrective action was successful.

The analysis of the internal quality control data should include checking of:

- the internal quality control samples (are they within specifications)
- the treatment, storage and handling of the samples (repeat the determination if possible)
- the apparatus (do other parameters done with the same apparatus show the same deviation)
- the environment of the lab (temperature storage of samples and reagents and of centrifuges, water baths and freezers)
- the treatment of the control samples is different from the normal samples (water, apparatus, pipettes, shaking, mixing)
- the calibrants and reference material samples (batch influence, preparation and storage, compare with older calibrants if possible)
- the calculations, dilutions, transcription and program errors.

An example of a checklist to analyse the internal quality control data and the corresponding flowchart is given in appendix 3. If somewhere is listed: "stop checklist" (implying that the problem is solved) someone always needs to think about correcting SOP's or other instructions.

5.3 Tasks and Functions

5.3.1 Laboratory analyst/manager

For the majority of Proficiency Testing schemes, reports are usually sent, in the first instance, to the laboratory manager or analyst responsible for the work, since they are familiar with the operation of the PT scheme. It is recommended, by most schemes, that the results from any distribution are studied in conjunction with the scheme protocol or participants' instructions. In particular, the performance scoring system used in the

scheme needs to be clearly understood. If this is unclear from any documentation provided by the co-ordinator, the co-ordinator should be contacted to clarify matters.

Laboratory staff should always check that the results in the report are those submitted by the laboratory, and whether the results in the report indicate satisfactory, unsatisfactory or questionable performance. If there is any doubt regarding the validity of the results in the report the co-ordinator should be contacted to resolve the situation.

Where an unsatisfactory performance score has been obtained for any analyte(s), it is recommended that the laboratory ascertain possible causes (see par. 5.2 for factors that might be of interest). Furthermore, the range, specificity and sensitivity of the method should be considered when interpreting a given z score. Data in the mid-range of the method, producing a questionable z score is of more concern than when the data are at or near the limit of detection, or at the top end of the range of the method. The precision of the method is usually less well-defined at the extremes of its range. The laboratory may be able to use results obtained in this situation to assess the validity of its current method. Poor performance could lead to the rejection of the method for a particular analysis.

5.3.2 Quality Manager

The Quality Manager of a laboratory or organisation participating in a PT scheme will usually receive a report on that laboratory's or organisation's performance, either from the scheme co-ordinator directly or via the laboratory manager. The Quality Manager will be concerned with the level of performance in the PT scheme, and will usually be seeking an explanation of unsatisfactory performance as well as evidence of any appropriate remedial action.

The Quality Manager needs to consider all the points which have previously been raised by the laboratory manager or analyst. The Quality Manager should also review any unsatisfactory result with respect to the laboratory's quality policy, and should decide whether such data is nevertheless fit for the laboratory's, and customer's, purpose before demanding remedial action. A result deemed "unsatisfactory" by the PT scheme's

statistical protocol may be satisfactory when taken in the context of the participating laboratory's own precision statement for any method, or in the context of its customers' requirements.

The Quality Manager should pay particular attention to the longer term performance trends in PT scheme participation. He or she should take note of any bias for particular determinations which may be apparent over several rounds. It is also important to note whether performance improves as a result of action taken when an unsatisfactory result has been obtained. The Quality Manager should bring any trends or bias to the attention of the analyst and the laboratory manager, and identify the cause of the errors.

- The Quality Manager should consider at which point remedial action needs to be taken following unsatisfactory PT scheme performance. The list in par. 5.2 in is intended to provide guidance

5.3.3 Senior manager

Senior managers within organisations participating in PT schemes have less day-to-day contact with PT matters than those within the laboratory with responsibility for analytical quality and PT scheme performance. There is therefore considerable potential for senior managers to misinterpret data regarding PT scheme performance within their organisation.

It is important for senior managers to appreciate the essentials of PT scheme participation. They should be aware that PT scheme participation is one of a number of quality assurance techniques available to laboratories, and that it only forms a part of the overall quality picture. They should also realise that PT scheme performance is not competitive. Individual scheme operators lay down criteria for the assessment of performance which are stated in the scheme protocol documentation. No difference in standard of performance should be implied between laboratories which fall within the category of "satisfactory performance" in any given round of that scheme.

It is advisable that senior managers also put PT scheme performance within their organisation into an appropriate context. In circumstances where a laboratory within their organisation has recorded an unsatisfactory performance for one or more determinations, this must be related to overall performance for the determinations in question. Unsatisfactory performance in the context of a round where the vast majority of participants achieved satisfactory performance should be viewed more seriously than a situation where unsatisfactory performance has been obtained by a high proportion of participants. Equally, one poor result for a given determination should not be viewed in the same manner as a series of poor results over time.

To interpret PT scheme data appropriately it is clear that senior managers should gain an appropriate level of understanding of all PT schemes in which their organisation is a participant. It is recommended that copies of scheme protocols or other documentation be available to senior managers for reference.

6. CONCLUSIONS

Proficiency testing is gaining increasing importance as a quality assurance tool for laboratories carrying out analytical measurements. The performance of laboratories in PT schemes is also being increasingly used, particularly by accreditation bodies, as a measure of the competence and quality of laboratories.

It is important for laboratories to have comprehensive information on the scope and availability of PT schemes in the areas in which they work. This will enable them to make appropriate decisions about which scheme(s) they should participate in. It is important that this type of information is widely disseminated in order to more effectively inform such decisions.

Laboratories therefore need to develop a good working understanding of PT, what the objectives of PT are, and how the data from PT schemes should be evaluated and used. This is important not only for laboratory staff and management within laboratories, but also for those who use their results including accreditation bodies and the laboratory's customers.

There are a number of key principles which need to be addressed by all the above parties:

- The PT scheme in which a laboratory participates should resemble as closely as possible the laboratory's routine work in terms of test samples, analytes and levels; any differences should be noted and accounted for;
- Performance in a PT scheme should be placed in the correct context
- The performance of a laboratory over several rounds of a PT scheme should be looked at where possible
- The scheme documentation and statistical protocol should be read in order understand how the scheme operates
- Where appropriate talk to the scheme co-ordinator to gain a more accurate understanding of the scheme and its operation

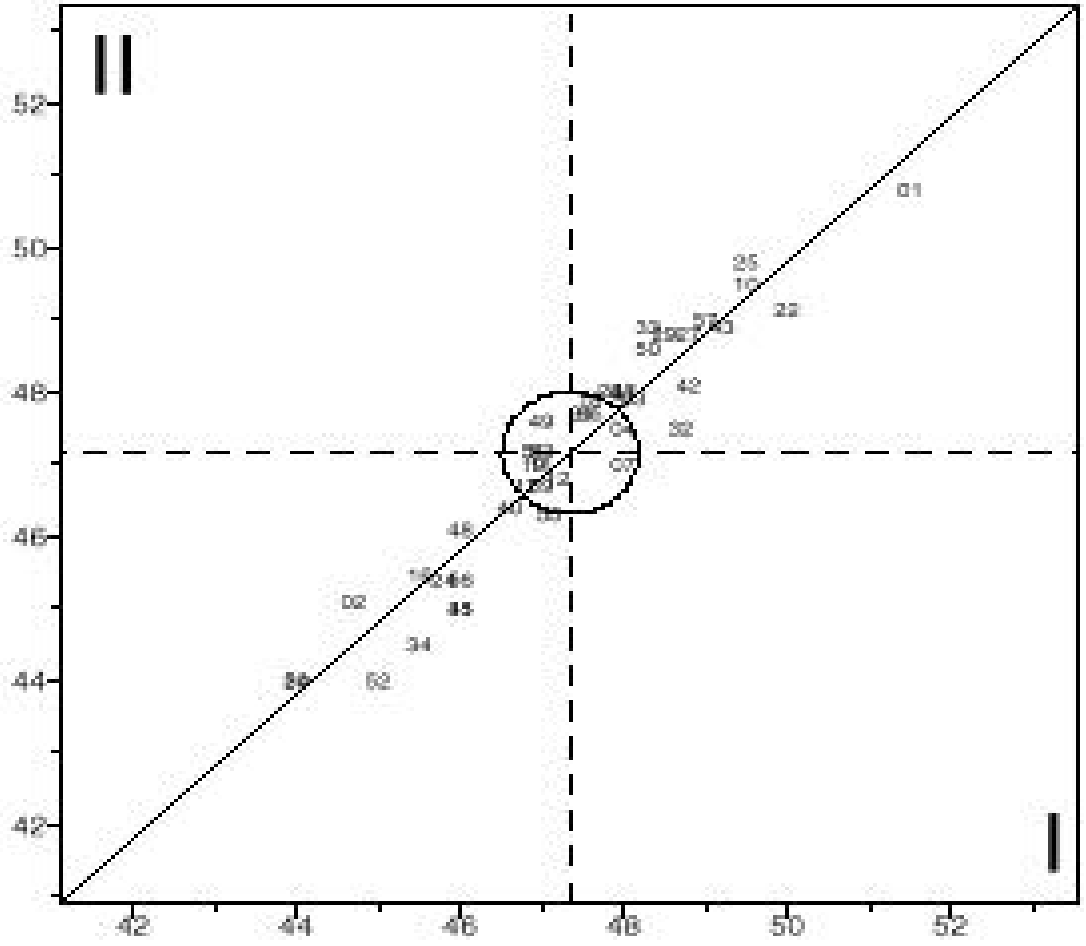
APPENDIX 1: JUDGEMENT OF RELEVANCY OF PROFICIENCY TEST

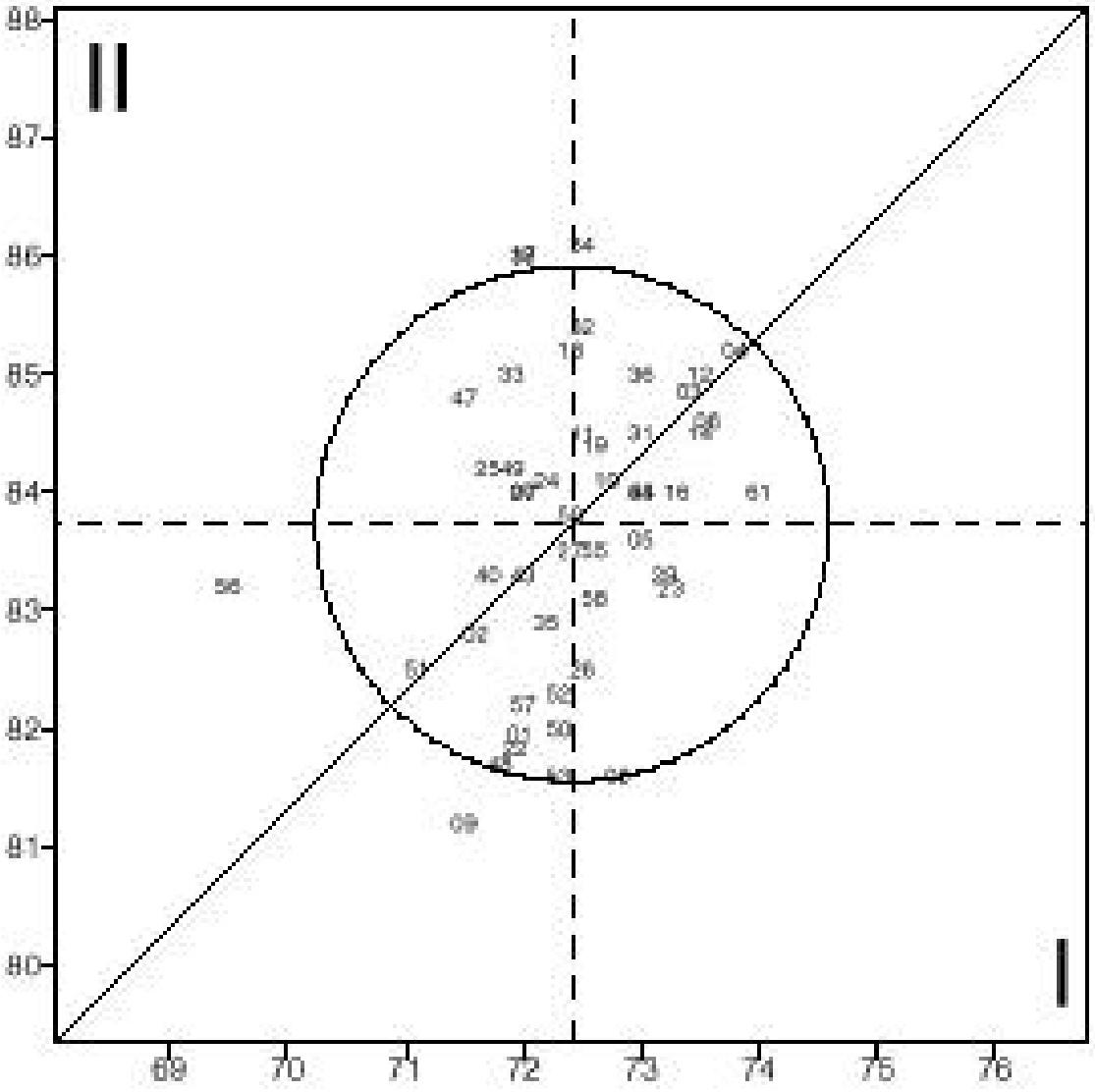
Analysis	Trace elements in Animal Feed
Proficiency test	Metals in Animal Feedstuufs (Organiser: UP, The Netherlands)

Situation in laboratory	Situation in proficiency test	Differences acceptable?
Compound(s) usually analysed: *Cu, Ag, Hg*	*Compound(s) offered:* *Ag, Al, Mg, Hg, Zn*	*Yes*
Matrix usually analysed: *Catfood*	*Matrix offered:* *Animal feedstuffs*	*Yes*
Concentration range(s) usually analysed: *0-200mg/kg*	*Concentration range(s):* *50-500 mg/kg*	*Yes*
Method(s) usually used: *Acid solubilisation after ashing*	*Method(s) prescribed:* *None (short descriptions of typical methods given)*	*Yes*
Type of laboratory: *Producer of Petfood*	*Type of laboratories participating:* *All kinds of laboratories in the Netherlands*	*Yes*
……..:	*……..:*	/
	Final conclusion: PT relevant?	***YES*** */ NO*

DATE: Jan 11, 1999……. APPROVED BY: Quality manager……. SIGNATURE: …….

APPENDIX 2: TWO EXAMPLES OF TYPICAL YOUDEN PLOTS




II
I
88
87
86
85
84
83
82
81
80
69
70
71
72
73
74
75
76

APPENDIX 3: CHECKLIST FOR ANALYSING THE INTERNAL QUALITY CONTROL DATA

A Check the internal control-sample(s): are they within specifications?:

YES: proceed with A1

NO: proceed with B

A1 Check calculations, dilutions, transcription errors and spreadsheet-programs. If there have alterations been made in the calculations, check validation of the calculations with known values. Is the problem solved?

YES: stop checklist

NO: proceed with A2

A2 Check the treatment, storage and handling of the control-sample: repeat the determination (if possible). Problem solved?

YES: stop checklist

NO: proceed with A3

A3 Check apparatus: check if other parameters done with the same apparatus show the same deviation.

YES: proceed with C

NO: proceed with A4

A4 Check the environment of the lab: temperature storage samples, storage reagents, assistant-apparatus (centrifuges, water baths, freezers). Are there any deviations?

YES: correct the problem and check the determination

NO: proceed with A5

A5 Check if the control-sample has been treated in a different way than the normal samples: used water, apparatus, pipettes, shaking, mixing. Repeat (if possible) the determination with internal control-samples the same way as the external control samples have been treated. Is the problem solved?

YES: stop checklist

NO: proceed with A6

A6 Check calibrants and reference samples: check: batch-influence, preparation of calibrants and reference samples, storage of calibrants. Compare with older calibrants (if possible), check response-values of calibrants, dilutions of calibrants. Is the problem solved?

YES: stop checklist

NO: proceed with A7

A7 Try to get an other control-sample of reference sample and repeat all steps to check if the problems are not caused by matrix-effects or other sample-related effects. Other control-sample (or reference sample) gives good results?

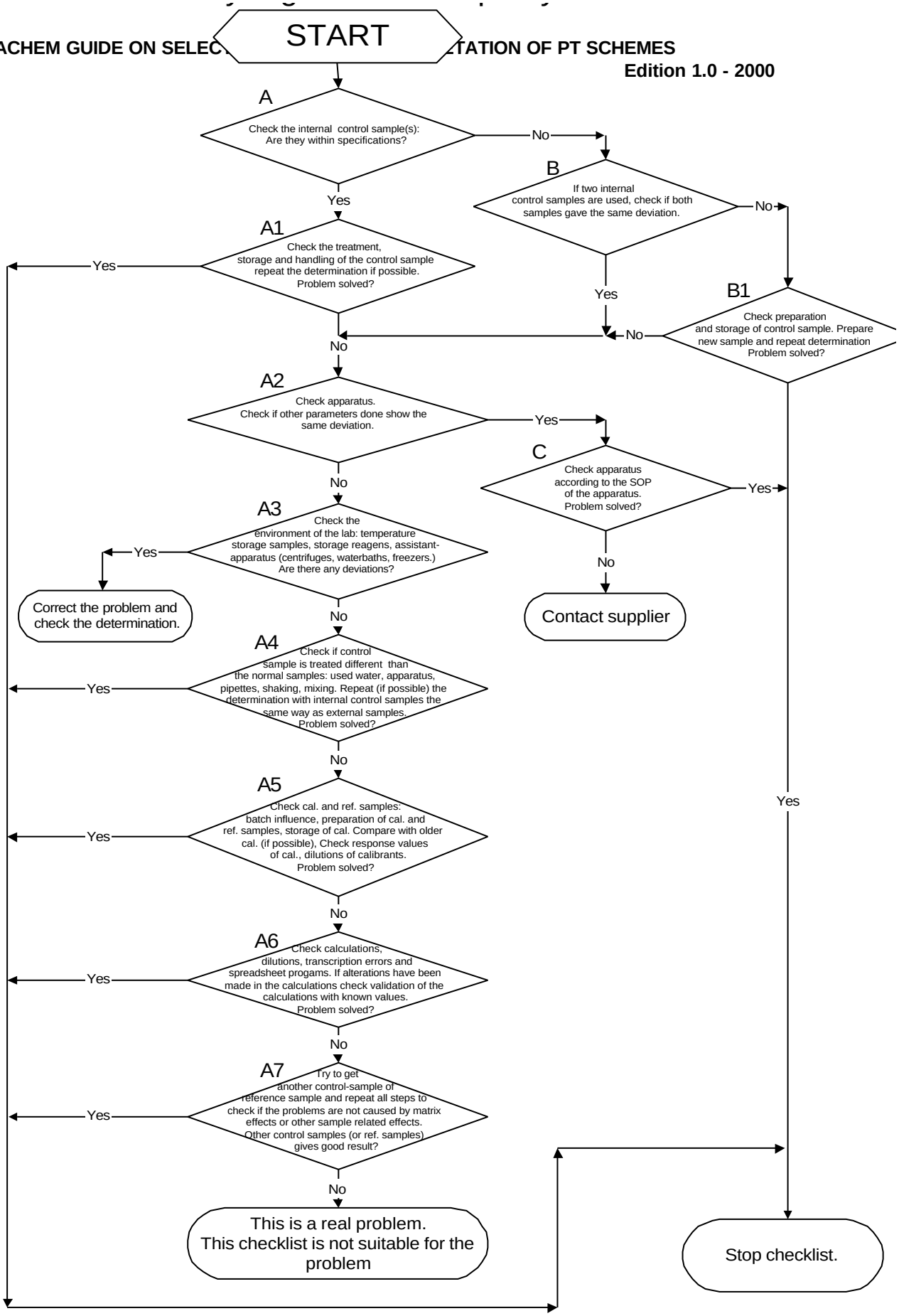
YES: stop checklist

NO: this is a real problem. This checklist is not suited for the problem.

B If two internal control samples are used, check if both samples gave the same deviation:
YES: proceed with A2
NO: proceed with B1

B1 Check preparation and storage of control sample. Prepare new sample and repeat determination: problem solved?
YES: stop checklist
NO: proceed with A2

C Check apparatus according to the SOP of the apparatus: problem solved?
YES: stop checklist
NO: contact supplier


START
A
Check the internal control sample(s): Are they within specifications?
No
Yes
B
If two internal control samples are used, check if both samples gave the same deviation.
No
Yes
A1
Check the treatment, storage and handling of the control sample repeat the determination if possible. Problem solved?
Yes
No
B1
Check preparation and storage of control sample. Prepare new sample and repeat determination Problem solved?
No
Yes
A2
Check apparatus. Check if other parameters done show the same deviation.
Yes
No
C
Check apparatus according to the SOP of the apparatus. Problem solved?
Yes
No
A3
Check the environment of the lab: temperature storage samples, storage reagens, assistant-apparatus (centrifuges, waterbaths, freezers.) Are there any deviations?
Yes
No
Correct the problem and check the determination.
Contact supplier
A4
Check if control sample is treated different than the normal samples: used water, apparatus, pipettes, shaking, mixing. Repeat (if possible) the determination with internal control samples the same way as external samples. Problem solved?
Yes
No
A5
Check cal. and ref. samples: batch influence, preparation of cal. and ref. samples, storage of cal. Compare with older cal. (if possible), Check response values of cal., dilutions of calibrants. Problem solved?
Yes
No
A6
Check calculations, dilutions, transcription errors and spreadsheet progams. If alterations have been made in the calculations check validation of the calculations with known values. Problem solved?
Yes
No
A7
Try to get another control-sample of reference sample and repeat all steps to check if the problems are not caused by matrix effects or other sample related effects. Other control samples (or ref. samples) gives good result?
Yes
No
This is a real problem. This checklist is not suitable for the problem
Yes
Stop checklist.

APPENDIX 4: INTERPRETATION OF DATA FOR END-USERS OF DATA OF ANALYTICAL LABORATORIES

A4.1 Introduction

A number of organisations outside the participating laboratory may wish to review the performance of a laboratory in a PT scheme; these include accreditation bodies, regulatory bodies and customers. These groups are increasingly recommending or demanding participation of laboratories in appropriate PT schemes as one means of demonstrating measurement quality.

As a general rule, laboratories who participate in PT schemes do so because they have a commitment to quality. This holds regardless of whether or not the laboratory has any formal accreditation status. However, participation in a PT scheme neither implies satisfactory performance on that scheme, nor does it guarantee competence in analysis in general. Also, failure to participate in a voluntary PT scheme does not necessarily imply incompetence. Customers of laboratories need to understand this, particularly when drafting specifications and tenders. Care also needs to be taken to ensure that unrealistic demands are not made regarding the level of performance in any scheme. In these circumstances, it is advisable to consult the scheme co-ordinator to ensure that any PT performance requirements placed on the laboratory are consistent with the aims, scope and history of the scheme.

A4.2 Accreditation Bodies

Accreditation bodies, and technical sassessors employed by accreditation bodies, generally have a good understanding of the role of proficiency testing, and are skilled in the interpretation of PT scheme results obtained by laboratories who are either accredited or seeking accreditation. In following good practice accreditation bodies will familiarise themselves with any PT scheme in which the laboratory participates. Scheme protocols and other documentation will be studied and, if appropriate, the scheme co-ordinator contacted to discuss or clarify any outstanding issues. The level of performance on a PT scheme for any laboratory will be determined against the standard laid down by

the scheme co-ordinator. In some cases, what constitutes unsatisfactory performance within a PT scheme may still be acceptable or fit for purpose within the scope of the laboratory's accreditation.

A4.3 Regulatory Bodies

Regulatory bodies are less likely to need to investigate the PT scheme performance of individual laboratories. However, in certain circumstances, regulatory bodies have the need to satisfy themselves that measurements made in laboratories which are covered by regulations or directives are of satisfactory quality. Regulatory bodies may use PT scheme performance as one of the ways of assessing quality in addition to other approaches including having referee analyses undertaken or submitting check samples for analysis.

Where a regulatory body has been involved in the development of a PT scheme, that scheme will incorporate features which are of direct relevance to that body, and will be readily understood. For those situations where the regulatory body is using an independent scheme for their own purposes, it is recommended that they discuss fully the scope and operational parameters of the scheme with the co-ordinator. This will enable them to put results obtained by any laboratory of interest into context. The statistical processes used by the co-ordinator for calculation of performance need to be understood, in order that a laboratory's performance may be judged in relation to any tolerances allowed in regulations. Advice may be required from the co-ordinator in such situations in order that PT scheme performance data is not misinterpreted.

A4.4 Customers of Participant Laboratories

The customer of a laboratory participating in a PT scheme can use the performance in the scheme as one tool with which to monitor the quality of that laboratory. The customer needs to have a good understanding of how the scheme operates and how performance within the scheme is calculated by the co-ordinator. Although some systems for determining performance in a PT scheme are widespread, such as the use of the z score, there are many different systems in use. In addition, customers should be aware

that the way in which z scores and other performance indicators are calculated can vary from scheme to scheme.

Customers are increasingly including PT scheme performance criteria in tender documents, and are using information about PT scheme performance supplied by potential contractors to assist in the decision as to which laboratory is awarded the contract. When using PT scheme performance as a criterion in a tender, customers should ensure that, where they are setting a “performance standard”, the standard is realistic and achievable. For example, asking laboratories to achieve satisfactory results for all analytes in all rounds of a scheme is unrealistic. Scheme co-ordinators should always be willing to provide appropriate information on the overall performance of the scheme, so that a good benchmark may be set. Customers should also take care to ensure that the determinations in which they have an interest are clearly stated, as the scheme may have a broader scope, and performance of laboratories in determinations not of direct interest may be irrelevant.

Customers must place any data relating to PT scheme performance from a contract laboratory into the proper context; laboratories could present data to a customer in a way which paints an unrealistically positive picture. Customers are recommended to carry out the following, as appropriate, in order to gain an accurate picture of the laboratory’s true performance:

- Obtain information on the scope and operation of the scheme (e.g. scheme protocol) from the laboratory or the co-ordinator.
- Remember that one distribution or round in a PT scheme only gives a brief snapshot of the laboratory’s performance, so look at performance over time.
- Look at the overall performance of all participating laboratories in order to judge how the laboratory is performing.
- Ask for copies of scheme reports (where confidentiality is not an issue) to confirm any data summarising PT scheme performance. This data should usually be provided by the scheme co-ordinator; for many PT schemes the agreement of the participating laboratory will also be required.

- One unsatisfactory result in any round does not make a laboratory poor, neither does the achievement of 100% satisfactory results in any round make a laboratory necessarily good.
- The way in which a laboratory responds to an unsatisfactory result will usually give more information about that laboratory than the occurrence of the unsatisfactory result.

APPENDIX 5: ABOUT THE AUTHORS OF THIS GUIDE

Eurachem Nederland and the task group "Proficiency testing schemes"

Eurachem was established in 1989 and its current membership is spread over more than 15 European countries. Eurachem is:

- a cooperative network of European laboratories
- a forum for discussions on common problems, solutions and strategies
- a framework for laboratories to develop agreements and to compare results

Eurachem tries to promote:

- the reliability of chemical analyses and the international acceptance of analysis results
- interest in quality issues in laboratories
- the application of international accepted quality standards
- the development and application of validated methods
- the establishment of traceability of measurement results, based on reference materials
- the development of mutually recognised method for the quality management of analytical methods, e.g. by participation in national and international interlaboratory studies
- the awareness of users of chemical analysis information of the objectives (including relevance, sensitivity and uncertainty).

Eurachem Nederland was established in 1990 and now includes over 100 delegates who provide a good reflection of the chemical sector in The Netherlands. Many members are active in working groups, amongst them the working group on "Interlaboratory Studies". The now reported study was carried on by some members of this working group.

Laboratory of the Government Chemist (LGC)

LGC is one of the foremost laboratories in Europe with a staff of over 300. LGC is committed to QA within the whole of the analytical community; QA and demonstration of the production of accurate data is seen to be of paramount importance. The Laboratory is responsible for the management of

DTI's Valid Analytical Measurement (VAM) programme. The central aim of VAM is to promote as an integral part of the national and international measurement system, a UK and European infrastructure which will allow analytical laboratories to demonstrate the validity of their measurements and facilitate the mutual recognition of analytical data across national boundaries. The programme promotes the use of recognised third party accreditation and QA systems, e.g. EN 45000, EN 29000, NAMAS, ISO 9000 and GLP, and promotes the formation of new systems along the guidelines as set out in document ISO Guide 25 and the OECD. LGC plays a leading role in promotional, training and harmonising activities, including chairmanship of European and world chemistry forum. This places LGC at the forefront of international harmonised valid analytical measurements.

LGC has 7 years' experience in all aspects the operation of proficiency testing (PT) schemes. LGC currently provides 4 PT schemes to laboratories in over 30 countries, and can demonstrate competence in sample preparation, homogeneity and stability testing, sample packing and dispatch, data handling, result evaluation and report production. All LGC PT activities are carried out with the maximum regard to confidentiality. LGC holds ISO 9000 registration for these activities, and also complies with ISO/IEC Guide 43: 1997 Part 1, and the ILAC document "Requirements for the Competence of Providers of Proficiency Testing". LGC is seeking accreditation from the United Kingdom Accreditation Service (UKAS) for its PT activities when this service is available.

References

[1] Eurachem Nederland, working group on 'Interlaboratory Studies", "Statistics and assessment of interlaboratory studies", December 1996

[2] Vankeerberghen et al, "Some robust statistical procedures applied to the analysis of chemical data", Chemometrics and Intelligent Laboratory Systems, 12 (1991) 3-13

[3] Rousseeuw, P.J., "Tutorial to robus statistics", Journal of Chemometrics, 5 (1991) 1-20

[4] AMC, "Robust statistics - How not to reject outliers. Part 1. Basic concepts", Analyst, 114 (1989) 1699-1702

[5] AMC, "Robust statistics - How not to reject outliers. Part 2. Inter-laboratory trials", Analyst, 114 (1989) 1693-1697

[6] Lischer, P.L., "Robust statistical methods in interlaboratory analytical studies"

[7] Central Science Laboratory (SCL) - Food Science laboratory (Norfolk, United Kingdom), "Protocol for the Food Analysis Performance Assessment Scheme (FAPAS): Organisation and analysis of data", fifth edition, april 1997

[8] Montford, M.A.J. van, "Statistical remarks on laboratory-evaluating programs for comparing laboratories and methods", Commun. Soil Sci. Plant Anal. 27 (1996) 463-478

[9] Oussoren, W., Visser, R.G., Van der Kaaden, A., "Interlaboratory Studies: Protocol for the Organisation, Statistics and Evaluation", Institute for Interlaboratory Studies (I.I.S.), August 1998, Dordrecht (The Netherlands)

[10] F.R. Hampel et al, "Robust Statistics. The approach based on influence functions", John Wiley & Sons (1986) New York

[11] P.J. Huber, "Robust statistics", John Wiley & Sons (1981) New York

[12] LGC, "Protocol for the proficiency testing scheme for the determination of alcoholic strength of beverages (ProTAS)", October 1997, Teddington (United Kingdom)

[13] ASTM standard E 178 - 80: "Standard practice for dealing with outlying observations", American Society for Testing and Materials, West Conshohocken, Pa.

[14] ISO 5725-1994 (E), "Accuracy (trueness and precision) of measurement methods and results"

[15] ISO 43-1995 (E), "Part 1: Selection and use of proficiency testing scheme by laboratory accreditation bodies" and "Part 2: Development and operation of proficiency testing programs"

[16] Tholen, D.W., “Statistical treatment of proficiency testing data”, Accred. Qual. Assur., 3 (1998) 362-366

[17] ASTM Guide E1301-96 (1196), “Standard guide for the development and operating of laboratory proficiency testing programs”, American Society for Testing and Materials, West Conshohocken, Pa.

[18] ISO 4259-1992 (E), “Petroleum Products: Determination and application of precision data in relation to methods of tests”

[19] W.J. Youden, "Statistical Techniques for Collaborative Tests", 1975, 10-11

[20] Wernimont G.T., "Use of statistics to develop and evaluate analytical methods", 1985, 82-85