Designation: E 1301 – 95 (Reapproved 2003)

An American National Standard

Standard Guide for Proficiency Testing by Interlaboratory Comparisons[1]

This standard is issued under the fixed designation E 1301; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

INTRODUCTION

Proficiency testing is the use of interlaboratory test comparisons to determine the performance of individual laboratories for specific tests and to monitor the consistency and comparability of a laboratory's test data.

Interlaboratory test comparisons are conducted for a number of other purposes including:

(*1*) Check the consistency and comparability of data for individual testing personnel;

(*2*) Assist in maintaining the calibration of instrumentation;

(*3*) Establish the effectiveness and comparability of new test methods;

(*4*) Achieve commercial improvement;

(*5*) Assist in determining reasons for interlaboratory differences;

(*6*) Determine the precision of a test method—often known as interlaboratory studies (see Practice E 691), collaborative trials, or round-robins; and

(*7*) Assign values to certified reference materials (CRMs).

Participation in proficiency testing programs provides laboratories with an objective means of assessing and demonstrating the reliability of the data they are producing. Although there are several types of proficiency testing programs, they all share the common feature of the comparison of test results obtained by two or more laboratories.

One of the main uses of proficiency testing programs is to assess laboratories' ability to perform tests competently. It thus supplements laboratories' own internal quality control procedures by providing an additional external evaluation of their testing capability. These activities also complement the technique of on-site laboratory assessment by technical specialists usually used by laboratory accrediting bodies. Confidence that a testing or calibration laboratory consistently obtains reliable results is of major importance to users of laboratory services. Users seeking such an assurance may undertake their own evaluation or may use the evaluation of other bodies.

Bodies assessing the technical competence of testing laboratories normally require or expect satisfactory participation in proficiency testing as evidence of a laboratory's ability to produce reliable test results, except where proficiency testing is inappropriate. However, it is emphasized that a major distinction exists between:

(*1*) The evaluation of the competence of a laboratory by the assessment of its total operation against pre-determined requirements, and

(*2*) The examination of the results of a laboratory's participation in proficiency testing which may only be considered as giving information about the technical competence of the testing laboratory at a single point of time under the specific conditions of the test for tests involved in a particular proficiency testing program.

1. Scope

1.1 While there are a number of uses for interlaboratory tests, and variations in their design and implementation, it is still possible to specify the essential principles that need to be considered when organizing such tests. Part A of this guide defines those principles and describes the factors that should be taken into account in the organization and conduct of proficiency testing programs.

1.2 This guide also covers how laboratory accrediting bodies, which assess technical competence of testing laboratories, should select and use proficiency testing programs (refer to Part B).

[1] This guide is under the jurisdiction of ASTM Committee e36 on Laboratory Accreditation and is the direct responsibility of Subcommittee E36.60 on Accreditation Systems.

Current edition approved October, 10, 1995. Published January 1996. Originally published as E 1301 – 89. Last previous edition E 1301 – 89.

1.3 Part A of the guide is intended for use by various parties, such as accrediting bodies, regulatory authorities and clients of laboratory services which have a need to assess the technical competence of laboratories. It is also useful for laboratories in self-evaluation, but recognizes that proficiency testing is only one mechanism that can contribute to establishing equivalent confidence among users of different testing laboratories.

1.4 It is currently a condition of some accreditation bodies that laboratories participate regularly in "approved" proficiency testing programs. Therefore, it is essential that program operators comply with principles for conduct of professionally managed proficiency programs, both in terms of technical requirements and quality management (see Annex A1 and Annex A2).

1.5 The methods of operation within different proficiency testing organizations are not expected to be identical and this guide does not give specific operational details for interlaboratory test comparisons. It does, however, cover both measurement comparison and testing programs in which large numbers of laboratories (over 20) or small groups of laboratories (1 to 20) are tested. Therefore, the contents of this guide are intended only as a framework to be modified appropriately for particular situations.

1.6 A list of some relevant references is given in Appendix X1.

2. Referenced Documents

2.1 *ASTM Standards:*

E 178 Practice for Dealing with Outlying Observations[2]

E 456 Terminology Relating to Quality and Statistics[2]

E 548 Guide for General Criteria Used for Evaluating Laboratory Competence[2]

E 691 Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method[2]

E 1187 Terminology Relating to Laboratory Accreditation[2]

2.2 *ANSI Standard:*[3]

ANSI/ISO/ASQC Q9000 Series: Quality Management and Quality Assurance Standards

2.3 *ISO Standards:*

ISO/IEC Guide 2, General Terms and Their Definitions Concerning Standardization and Related Activities[3]

ISO/IEC Guide 25, General Requirements for the Competence of Calibration and Testing Laboratories[3]

ISO Guide 30, Terms and Definitions Used in Connection with Reference Materials[3]

3. Terminology

3.1 *Definitions*—For formal definitions related to laboratory accreditation, Terminology E 1187 applies. For formal definitions related to quality and statistics, Terminology E 456 applies. In addition, the following terms and their definitions are provided for ease of reference.

3.1.1 *accuracy*—the closeness of agreement between a test result and an accepted reference value (Terminology E 456 without the note).

3.1.2 *bias*—the difference between the population mean of the test results and an accepted reference value (Terminology E 456 without the discussion).

3.1.3 *certified reference material (CRM)*—a reference material, accompanied by a certificate, one or more of whose property values are certified by a procedure that establishes traceability to an accurate realization of the unit in which the property values are expressed, and for which each certified value is accompanied by an uncertainty at a stated level of confidence (ISO Guide 30 without the notes).

3.1.4 *precision*—the closeness of agreement between test results obtained under prescribed conditions (Terminology E 456 without the three notes).

3.1.5 *proficiency testing (laboratory)*—determination of laboratory testing performance by means of interlaboratory comparisons (ISO/IEC Guide 2).

3.1.6 *reference material*—a material or substance, one or more of whose property values are sufficiently homogeneous and well established to be used for the calibration of an apparatus, the assessment of a measurement method, or for assigning values to materials (ISO Guide 30 without the note).

3.1.7 *repeatability*—the closeness of agreement between test results obtained under repeatability conditions (that is, conditions under which test results are obtained with the same test method in the same laboratory by the same operator with the same equipment in the shortest practical period of time using test units or test specimens taken at random from a single quantity of material that is as nearly homogeneous as possible (Terminology E 456 without the notes).

3.1.8 *reproducibility*—the closeness of agreement between test results obtained under reproducibility conditions (that is, conditions under which test results are obtained with the same test method on identical material in different laboratories (Terminology E 456 without the notes).

3.1.9 *test*—technical operation that consists of determination of one or more characteristics of a given product, process or service according to a specified procedure (ISO/IEC Guide 2).

3.1.10 *trueness*—the closeness of agreement between the population mean of the measurements or test results and an accepted reference value (Terminology E 456 without the note).

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *accepted reference value*—a value that serves as an agreed-upon reference for comparison and which is derived as: (*1*) a theoretical or established value, based on scientific principles, (*2*) an assigned value, based on experimental work of some national or international organization, and (*3*) a consensus value, based on collaborative experimental work under the auspices of a scientific or engineering group.

3.2.2 *Discussion*—When the accepted reference value is the theoretical value, it is sometimes referred to as the "true" value. (This is a small variation from the definition in Terminology E 456.)

3.2.3 *assigned value*—estimate of the true value used in the assessment of proficiency (also referred to as assigned reference value).

[2] *Annual Book of ASTM Standards*, Vol 14.02.

[3] Available from American National Standards Institute, 11 W. 42nd St., 13th Floor, New York, NY 10036.

3.2.4 *coordinator*—person or body that coordinates all the activities associated with a proficiency program.

3.2.5 *internal quality control (IQC)*—the set of procedures undertaken by a laboratory for continuous monitoring of operations and results in order to decide whether the results are reliable enough to be released; IQC primarily monitors the batch-to-batch accuracy of results on quality control materials, and precision on independent replicate analyses of test materials.

3.2.6 *outlier*—an observation that appears to deviate markedly from the other observations of the sample (also referred to as extreme result, outlying or doubtful observation, or aberrant value) (see Practice E 178).

3.2.7 *quality assurance system*—the sum total of a laboratory's activities aimed at achieving the required standard of analysis.

3.2.8 *reference laboratory*—laboratory that establishes the accepted reference value or assigned value.

3.2.9 *test item*—material(s) or artifact(s) presented to the participating laboratory for the purpose of proficiency testing.

3.2.10 *testing laboratory*—laboratory that performs tests (including calibration) (also referred to as "participating laboratory" or just "laboratory").

4. Significance and Use

4.1 The previous edition of this guide (E 1301 – 89) covered the development and operation of laboratory proficiency testing programs with limited, if any, emphasis on the use of the outcomes of proficiency testing by accreditation bodies.

4.2 This revised version is now intended to provide guidance in three areas:

4.2.1 The introduction to this guide distinguishes between use of interlaboratory tests for proficiency testing and for other purposes.

4.2.2 Part A of this guide provides guidance on the development and operation of interlaboratory tests for use in proficiency testing programs.

4.2.3 Part B of this guide provides guidance on the selection and use of proficiency testing programs by laboratory accreditation bodies.

4.3 Annex A1 through Annex A2 provide statistical guidance on treatment of data from proficiency testing programs and a guide to the documentation of the quality assurance system of proficiency testing programs.

4.4 While the emphasis of Part A of the guide is on operation of interlaboratory tests for proficiency testing, most of the principles and guidance given are applicable to operation of interlaboratory tests for other purposes.

4.5 While many laboratory accreditation bodies operate their own proficiency testing programs, a significant number also use proficiency testing programs or other forms of interlaboratory tests operated by other bodies. The purpose of Part B of this guide is to provide harmonized principles for selection of suitable interlaboratory test programs for use as proficiency testing programs by laboratory accreditation bodies.

4.6 Part B of this guide is intended:

4.6.1 To establish principles for the selection of proficiency testing programs for use in laboratory accreditation programs; and

4.6.2 To assist in harmonizing the use of results of proficiency testing programs by laboratory accreditation bodies.

4.7 As results from proficiency testing programs may be used in accreditation decisions, it is important that both the accrediting bodies and participating laboratories have confidence in the design and operation of the programs.

4.8 It is also important for participating laboratories and laboratory accreditation assessors to have a clear understanding of the accrediting bodies' policies for participation in such programs; the criteria they use for judging successful performance in proficiency testing programs; and their policies and procedures for following up any unsatisfactory results from a proficiency test.

4.9 It should be recognized that laboratory accrediting bodies and their assessors may take into account the suitability of test data produced from other activities apart from proficiency testing programs. This includes results of laboratories' own internal quality control procedures with control samples, comparison with split-sample data from other laboratories, performance of one-time audit tests with certified reference materials, and so on. The use of data from these sources by laboratory accrediting bodies is not covered by this guide. However, the principles set out in this guide, regarding follow-up of unsatisfactory performance, could also apply to these activities.

Part A: DEVELOPMENT AND OPERATION OF PROFICIENCY TESTING PROGRAMS

5. Types of Proficiency Testing

5.1 Proficiency testing techniques vary depending on the nature of the item or material under test, the test method in use and the number of testing laboratories participating. They possess the common feature of comparison of test results obtained by one testing laboratory with those obtained by one or more other testing laboratories. In some programs, one of the participating laboratories may have a controlling, coordinating, or reference function. Paragraphs 5.2-5.4 describe the major types of proficiency testing programs.

5.2 *Measurement Comparison Programs*— Measurement comparison programs involve the item (measurement artifact) to be tested or calibrated being circulated successively from one participating laboratory to the next. Features of such programs usually are:

5.2.1 The item will often be periodically returned to a central laboratory acting as the reference laboratory for calibration, testing or inspection before being passed on to the next successive participating laboratory in order to determine whether any changes have taken place to the item or its assigned reference values.

5.2.2 Programs involving sequential participation take time (in some cases years) to complete. This causes a number of difficulties such as ensuring the stability of the item, the strict monitoring of its circulation and the time allowed for testing by individual participants, and the need to supply feedback on individual performance to laboratories during the program

rather than waiting until it finishes. In addition, it may be difficult to compare results on a group basis as there may be relatively few laboratories whose measurement capabilities closely match each other.

5.2.3 The individual measurement results are compared with the reference values established by the reference laboratory. The coordinator may have to take into account the claimed measurement uncertainty of each participating laboratory.

5.2.4 Examples of items (measurement artifacts) used in this type of proficiency testing include reference standards, such as resistors, gauges and instruments.

5.3 *Interlaboratory Testing Programs*— Interlaboratory testing programs involve randomly selected subsamples from a source of material being distributed simultaneously to participating testing laboratories for concurrent testing. Usual features of such programs include:

5.3.1 Subsamples provided to each participant must be sufficiently homogeneous so that any results later identified as extreme cannot be attributed to significant sample variability.

5.3.2 Testing results are returned to the coordinator and analyzed against an assigned value (best estimate of the "true value") to give an indication of the performance of the individual laboratories and the group as a whole.

5.3.3 This is the type commonly used by accreditation bodies and other organizations when they conduct programs in the testing field as opposed to the measurement/calibration field.

5.3.4 Examples of samples used in this type of proficiency testing include food, bodily fluids, water, soils, and other environmental material. In some cases, separate portions of previously established certified reference materials are circulated.

5.4 *Split Sample Testing Programs*— These programs involve samples of a product or a material being divided into two or more parts with each participating laboratory testing one part of each sample.

5.4.1 These programs differ from the type of proficiency testing described in 5.3, as there is usually limited control of, or preliminary data on, the homogeneity of the sample being divided.

5.4.2 This technique is sometimes used by clients of laboratory services, including regulatory authorities.

5.4.3 This type of program often needs retention of sufficient material for possible further analysis conducted by additional laboratories in order to resolve any perceived differences among the limited number of laboratories initially involved.

5.4.4 There may be statistical limitations on analysis of data provided by the laboratories, due to the small number involved, often only two laboratories.

6. Organization and Design

6.1 *Framework*:

6.1.1 The design stage of any proficiency program requires the input of technical experts, statistician(s), and a program coordinator to ensure its success and smooth operation.

6.1.2 The coordinator in consultation with these other personnel should develop a program appropriate to the particular proficiency test. A plan should be agreed upon and documented (see Annex A2) before the start of the program and typically should include the following information:

6.1.2.1 The name and the address of the organization conducting the proficiency program;

6.1.2.2 The name and address of the coordinator and other personnel involved in the design and operation of the proficiency program;

6.1.2.3 The nature and the purpose of the proficiency program;

6.1.2.4 A procedure for the manner in which the participants are selected;

6.1.2.5 The name and address of the laboratory or laboratories performing various parts of the program, such as sampling, sample processing, homogeneity testing and stability testing, and also a list of the potential participating laboratories;

6.1.2.6 The nature of the test item(s) and test(s) selected, as well as a short description of the considerations underlying these choices;

6.1.2.7 A description of the manner in which the test items are obtained, processed, checked and transported;

6.1.2.8 A description of the information that is supplied to participants in the prenotification phase and of the time schedule for the various phases of the proficiency testing (refer to Section 8);

6.1.2.9 The expected initial and final date of the proficiency program including the date(s) for the testing to be carried out by the participants;

6.1.2.10 Details of methods or procedures which participants should use to perform the tests;

6.1.2.11 The basis for the selected statistical model and any outlier tests to be used;

6.1.2.12 The techniques for evaluating laboratory performance;

6.1.2.13 A description of the extent to which the test results, and the conclusions that will be based on the outcome of the proficiency test, are to be made public.

6.2 *Personnel*:

6.2.1 The personnel involved in providing the program should have adequate qualifications and experience in the design, implementation, and reporting of interlaboratory tests and include appropriate technical, statistical and administrative skills.

6.2.2 The operation of particular interlaboratory test comparisons may also require the guidance of persons with detailed technical knowledge and experience of the test methods and procedures involved. To this end, the coordinator may need to enlist an advisory panel of at least two persons drawn from, for example, professional bodies, contract laboratory (if any), program participants, accrediting bodies, or end-users of the data. The involvement of the advisory panel should be active and ongoing.

6.2.3 The functions of this advisory panel include:

6.2.3.1 The development and review of procedures for the planning, execution, analysis, and reporting of the proficiency testing program;

6.2.3.2 The identification and evaluation of interlaboratory test comparisons organized by other bodies;

6.2.3.3 The evaluation of proficiency test results regarding the performance of participating laboratories; and

6.2.3.4 Providing advice to a body assessing the technical competence of testing or calibration laboratories on the use of proficiency testing as an element of its laboratory evaluations.

6.3 *Equipment*—The use of a computer-based system, while not essential, is strongly recommended. Whatever facilities are used, they should be adequate to conduct all the necessary data entry and statistical analyses and provide timely and valid results. Procedures for checking data entry should be implemented and all relevant software should be verified, supported and backed up. The storage and security of data files is another important equipment consideration.

6.4 *Statistical Design*:

6.4.1 The statistical model and data treatment techniques that are to be used should be documented together with a short description of the background process and criteria used to select a specific model. Details of common statistical models and treatment of proficiency testing data appear in Annex A1.

6.4.2 Appropriate statistical design of a proficiency testing program is essential and careful consideration should be given to the following matters and their interaction:

6.4.2.1 The inherent repeatability and reproducibility of the test(s) involved;

6.4.2.2 The smallest differences to be detected between participating laboratories at a desired confidence level;

6.4.2.3 The number of participating laboratories;

6.4.2.4 The number of samples to be tested and the number of repeat tests or measurements to be carried out on each sample;

6.4.2.5 The procedures to be used to establish the assigned value; and

6.4.2.6 Procedures to be used to eliminate outlier results.

6.4.3 In the absence of reliable information concerning the inherent repeatability and reproducibility of the test(s) involved, it may be necessary in some cases to organize a pilot interlaboratory test comparison, round robin or collaborative trial to obtain it.

6.5 *Test Items*:

6.5.1 Preparation of test items may either be contracted out or undertaken by the coordinator. The laboratory preparing the test item should have demonstrable competence in the area of testing or calibration being examined.

6.5.2 Any conditions relating to the test items that may affect the integrity of the interlaboratory comparison, such as homogeneity, stability, possible damage in transit and effects of ambient conditions, should be considered (see 6.6).

6.5.3 The test items or materials to be distributed in the program should generally be similar in type to those routinely tested.

6.5.4 The number of test items or materials to be distributed per round may depend on whether or not there is a requirement to cover a range of property characteristics.

6.5.5 The assigned value should not be disclosed to the participants until after the results have been collated.

6.6 *Sample Management*:

6.6.1 Procedures for sampling, randomizing, transporting, receiving, identifying, labelling, storing, and handling of test items or materials should be documented.

6.6.2 When used, bulk material prepared for a proficiency test must be sufficiently homogeneous for each test parameter so that all laboratories will receive test samples that do not differ significantly in the parameters to be measured. The coordinator should clearly state the procedure used to establish the homogeneity of the test material. Ideally, homogeneity testing will be carried out before distribution of the test items to the participating laboratories.

6.6.3 Where possible, the coordinator should also provide evidence that the test materials or artifacts are sufficiently stable to ensure that they will not undergo any significant change throughout the conduct of the proficiency test. Therefore, prior to distribution the stability of the test material may need to be determined by carrying out measurements after it has been stored for an appropriate period of time. The storage conditions, especially time and temperature, used in the stability trials should represent those conditions likely to be encountered for the full period of the proficiency test. Stability trials therefore take account of the transport of the test samples to participating laboratories as well as the conditions encountered in the laboratory environment. The various parameters to be measured should show no significant changes during the stability tests, the magnitude of a significant change being assessed from the knowledge of the variance expected for replicate analyses of the bulk material. When unstable parameters need to be tested, it may be necessary for the coordinator to prescribe a date by which the tests must be completed.

6.6.4 Coordinators should consider any hazards that the test materials might pose and take appropriate action to advise any party who might be at risk of the potential hazard involved, such as test material distributors, and staff of testing laboratories.

6.7 *Choice of Method/Procedure*:

6.7.1 The method/procedure used by participants may be prescribed by the coordinator, or the participants may be allowed to use the method of their choice. Methods and procedures used will normally have been previously validated by an appropriate means, such as an interlaboratory study. Often national or internationally accepted standard methods will be prescribed. As a general principle, however, procedures used by laboratories participating in proficiency testing programs should reflect those used in their routine analytical work.

6.7.2 Where a calibration procedure is used, the assigned value will often be a reference value obtained from measurements obtained by a high-echelon calibration laboratory, often a national standards laboratory, using a well-defined and accepted procedure. It is desirable that participating laboratories use the same or a similar procedure, but this will not always be practicable for calibration laboratories.

7. Operation and Reporting

7.1 The day-to-day operation of a program should be the responsibility of a coordinator. The following policies and procedures should be documented (see Annex A2):

7.2 *Instructions*:

7.2.1 Detailed instructions covering all aspects of the program that must be adhered to by the participating laboratories should be provided.

7.2.2 The instruction should include details concerning factors that may influence the testing of the supplied proficiency test items or materials. These factors may include operators, nature of items or materials, equipment status, selection of test procedure and timing of testing.

7.2.3 The instructions should advise that proficiency test items be handled as similarly to the routine tests as possible.

7.2.4 The instructions should prohibit special handling of graded test items including extra replicates, handling by a special person or method, and collaboration with other laboratories.

7.2.5 Specific instructions on the recording and reporting of test or calibration results should also be supplied, including units, number of significant figures, reporting basis, and so on.

7.3 *Packaging and Transportation*—The coordinator of the program must consider several matters regarding the distribution of the test or measurement item. The packaging must be adequate to protect the stability and characteristics of the test items. There may be certain restrictions on transportation such as dangerous goods regulations or customs requirements. Long distance transportation adds to the difficulties. In some cases, the laboratories themselves must also take responsibility for the transport of the items, particularly in sequential measurement comparison programs.

7.4 *Data Analysis and Records*:

7.4.1 The data received from the participating laboratories should be processed and analyzed and then reported back as soon as practicable. It is essential that procedures are put in place to check the validity of data entry and transfers and subsequent statistical analysis (refer to 6.3). Data capture sheets, computer back-up files or print-outs, graphs, and so on, should be retained for a reasonable period of time.

7.4.2 Data analysis should generate summary measures and performance statistics and associated information that are consistent with the statistical model and the goals of the program. Statistically-designed outlier-exclusion routines may be used to prevent extreme values from influencing the summary statistics. Suggestions for statistical techniques for evaluation are given in Annex A1.

7.4.3 The coordinator should have criteria for ungradable test results. The results for a test item may be such that the item should not be used to evaluate performance. This could be due to specimen instability or inhomogeneity, or errors in reporting.

7.5 *Program Report*:

7.5.1 The content of program reports may vary depending on the purpose of a particular program, but such a report should be clear and comprehensive and should include data on the distribution of results from the laboratories together with an indication of each individual participant's performance and all statistics on which the performance of participants was evaluated (see 7.7).

7.5.2 The following information should normally be included in reports of proficiency programs.

7.5.2.1 Name and address of organization providing the program;

7.5.2.2 Names and affiliations of persons involved in the design and conduct of the program (see 6.1 and 6.2);

7.5.2.3 Date of issue of report;

7.5.2.4 Report number and clear identification of program;

7.5.2.5 Clear description of items or materials used including details of sample preparation and homogeneity testing;

7.5.2.6 Laboratory identity codes and test results;

7.5.2.7 Statistical data and summaries including assigned value and range of acceptable results, for either quantitative or categorical values;

7.5.2.8 Test methods/procedures used by each laboratory, when different methods are permitted;

7.5.2.9 Comments on overall laboratory performance by the coordinator and technical advisers (see 7.6 and 7.7);

7.5.2.10 Procedures used to design and implement the program; and

7.5.2.11 Procedures used to statistically analyze the data with any pertinent references.

7.5.3 Reports should be made available as quickly as possible after the return of results to the coordinator. Although ideally all original data supplied should be reported to participants, it may not be possible to achieve this in some very extensive programs. Participants should, however, receive at least the results of all laboratories in graphical form, for example, histograms.

7.5.4 The level of understanding of statistical analysis among participating laboratories will vary. If applicable, the results of proficiency testing, at least initially, should therefore be analyzed to show the results of each participating laboratory clearly in relation to those of other participants. The use of histograms and charts may assist. More details of the types of statistical procedures that can be employed appear in Annex A1.

7.6 *Assessment of Performance*:

7.6.1 The coordinator should retain control over the assessment of performance in order to help maintain the integrity of the program.

7.6.2 The coordinator may enlist the assistance of technical advisers to provide expert commentary on performance with respect to:

7.6.2.1 Overall performance versus prior expectations, taking uncertainties into account;

7.6.2.2 Variation within and between laboratories, and comparisons with any previous programs or published precision data;

7.6.2.3 Variation between methods, procedures, or test equipment, when applicable;

7.6.2.4 Possible sources of error related to extreme results and suggestions for improving performance; and

7.6.2.5 Any other suggestions, recommendations or general comments.

7.6.3 It may be necessary to provide individual summary sheets for participants after a particular program and these may include updated summaries of performance of individual laboratories over various rounds of an ongoing program. Such summaries can be further analyzed and trends highlighted as required.

7.6.4 A variety of techniques exist to assess performance of participants, both for one-time programs and also after consecutive rounds of regular on-going programs. Some examples of such procedures are given in Annex A1.

7.7 *Feedback/Corrective Action*:

7.7.1 Participants should be provided with detailed information on joining a proficiency testing program. Subsequent communication with participants can be by letter, newsletter, reports, periodic open meetings, or some combination of these. Participants should be advised promptly of any changes in program design or operation.

7.7.2 The coordinator should have resources in place to assist participants with error detection and ways to determine whether corrective actions have been effective. This could include consulting advice, references to appropriate experts or journals, and previously used proficiency test items (unless not stable).

7.7.3 Participants who consider that their performance assessment is in error should be able to refer the matter to the coordinator.

7.7.4 Feedback from laboratories should be encouraged, so that participants actively contribute to the development of the program.

7.7.5 The procedures associated with the corrective action undertaken by participants, particularly in relation to feedback to accreditation bodies, is addressed in Part B of this guide.

8. Confidentiality/Ethical Considerations

8.1 *Confidentiality of Records*—The preservation of anonymity amongst participating laboratories is normally integral to the design and execution of proficiency programs. Laboratories should not feel threatened by the chance that a poor result may reflect on their overall reputation. This is distinct from a laboratory being prepared to accept the consequences of a poor performance, which may mean extensive follow-up and corrective action or, in some instances, loss of accreditation. The identity of participants should only be known to the absolute minimum number of people involved in coordinating a program, and this should extend to any subsequent remedial advice or action applied to a laboratory exhibiting poor performance. In some circumstances, a coordinator may be required to provide performance records upon request. In other circumstances, a coordinator may be required to report poor performance to a particular authority, but participants must be made aware in advance of this possibility. In other cases, a client or prospective client may approach the laboratory and ask them to provide details of their performance in a particular program.

8.2 *Collusion and Falsification of Results*:

8.2.1 Although proficiency testing programs are intended primarily to help participants, some participants may try to provide a falsely optimistic impression of their capabilities. For example, collusion may take place between laboratories, so that truly independent data are not submitted. Laboratories may also give a false impression of their performance if they routinely carry out single analyses, but report the mean of replicate determinations on the proficiency test items or conduct additional replicates to those specified for a particular program. Proficiency testing programs should, where practicable, be designed to hinder the possibility of collusion and falsification. For example, alternative samples could be distributed within one round, with no identifiable reuse of the materials in succeeding rounds. Also, instructions to participants should make it clear that collusion is contrary to professional, scientific conduct and serves only to nullify the benefits of proficiency testing to customers, accrediting bodies and analysts alike.

8.2.2 Although all reasonable measures should be taken by a coordinator to prevent collusion, it must be appreciated that it is the responsibility of the participating laboratories to avoid it.

Part B: SELECTION AND USE OF PROFICIENCY TESTING PROGRAMS BY LABORATORY ACCREDITATION BODIES

9. Background

9.1 The objective of laboratory accreditation is to provide a formal, independent recognition that a laboratory is competent to perform specific tests, measurements or calibrations. The procedures used to determine competence include assessment of laboratories' specific capabilities by independent technical assessors who evaluate both technical competence and the compliance of the laboratories with appropriate management and quality systems criteria such as those described in Guide E 548 (ISO/IEC Guide 25).

9.2 Most laboratory accreditation bodies complement their on-site assessments with various forms of practical testing to judge whether a laboratory's data are comparable to either reference data or to data provided by a laboratory or laboratories already determined to be competent in the relevant tests or measurements.

9.3 Some of the practical testing (also called audit testing) may be of a one-time nature involving a single laboratory, such as through submission of a certified reference material or a reference calibration artifact to a single laboratory. More comprehensive forms of practical testing, which involve interlaboratory comparisons between groups of two or more laboratories, are defined as proficiency testing.

9.4 Proficiency testing programs may be operated either by laboratory accreditation bodies or by other organizations. As the results of laboratories' performance in proficiency testing programs are used in judging their technical competence, and thus in the decisions to grant or maintain their accreditation, it is critical that the proficiency testing programs used by accreditation bodies be operated effectively and fairly.

10. Selection of Proficiency Testing Programs

10.1 To assist in the evaluation of competence of laboratories for laboratory accreditation purposes, accreditation bodies should use proficiency testing programs complying with the guidelines described in Part A of this guide.

10.2 Proficiency testing programs, whether conducted by an accrediting body or not, should be periodically audited and reviewed for compliance with Part A of this guide. Records of these audits and reviews should be maintained. If a proficiency testing program used by a laboratory accreditation body is operated by another organization, the laboratory accreditation

body should seek documentary evidence that the subcontracted program(s) comply with Part A of this guide before recognizing the program.

10.3 The accreditation body should require from the program coordinator documented evidence of active involvement by the advisory panel.

10.4 In selecting a proficiency testing program, the following factors should be considered by the laboratory accreditation body:

10.4.1 The tests, measurements, or calibrations involved should match the types of tests, measurements, or calibrations performed by the accredited laboratories proposed for participation.

NOTE 1—Some proficiency testing programs may offer tests that are not an exact match for the tests performed by an accredited laboratory (for example, the use of a different national standard for the same determination) but it may still be technically justified to include the laboratories in the program if the treatment of the data allows for consideration of any significant differences in test methodology or other factors.

10.4.2 With the agreement of their accredited laboratories, the accreditation body should have access to accredited participants' results, together with details of the program's design, instructions to participants, statistical treatment of data and the final report from each selected proficiency test.

10.4.3 The proficiency test program should have participants using a representative variety of methods. The accreditation body should consider the scope of test methods covered by the program and the variety of methods commonly used.

10.4.4 The frequency at which the program is run.

10.4.5 The suitability of the organizational logistics for the program such as timing, location, test item stability considerations, distribution arrangements, etc., relevant to the group of accredited laboratories proposed for the program.

10.4.6 The availability of acceptance criteria for the participating laboratories (that is, for judging successful performance in the proficiency test).

10.4.7 The costs of the selected programs.

10.4.8 The policy on maintaining the confidentiality of the identity of the participants and their results.

10.4.9 The timeliness of the program's reporting of results.

10.4.10 Confidence in the suitability of test materials, measurement artifacts, etc., used by the program for characteristics such as homogeneity, stability, and, where appropriate, traceability to national or international standards.

10.5 The selection of a specific proficiency testing program by a laboratory accreditation body should be authorized by, and monitored by, suitably qualified personnel of the accreditation body.

11. Policies on Participation in Proficiency Testing Programs

11.1 Laboratory accreditation bodies should document their policies for participation in proficiency testing programs by accredited and applicant laboratories. Such documented policies should be publicly available to laboratories and other interested parties.

11.2 Issues that should be addressed in participation policies include:

11.2.1 Whether participation is mandatory or voluntary for specific proficiency testing programs.

NOTE 2—In some cases, laboratory accreditation bodies may have policies that require mandatory participation in a minimum number of approved proficiency testing programs and accept voluntary participation in any additional programs that may be available.

11.2.2 The frequency at which laboratories are expected or invited to participate in proficiency testing programs.

11.2.3 The criteria used by the laboratory accreditation body to evaluate successful or unsatisfactory performance in a specific program.

NOTE 3—The designs of proficiency testing programs vary depending on the technologies involved and the acceptance criteria may also vary from program to program. In many cases, acceptance data will be derived from the results obtained during conduct of a specific program and thus will not be available to laboratories in advance. In such cases, laboratory accreditation bodies should provide participating laboratories with details of the principles on which acceptance criteria will be based.

11.2.4 Whether laboratories may be required to participate in follow-up programs if performance is judged to be unsatisfactory in a specific program.

11.2.5 How the results of proficiency testing will be used in accreditation decisions.

11.2.6 Details of the laboratory accreditation body's policy on preserving participants' confidentiality.

12. Use of Results by Laboratory Accreditation Bodies

12.1 The results from proficiency testing programs are useful for both participating laboratories and accreditation bodies. There are, however, limitations on the use of such results to determine competence. Successful performance in a specific program may represent evidence of competence for that exercise but may not reflect ongoing competence. Similarly, unsuccessful performance in a specific program may reflect a random departure from a laboratory's normal state of competence. It is for these reasons that proficiency testing alone is not used by laboratory accreditation bodies in their accreditation processes.

12.2 If a laboratory submits a result or results that fall outside acceptance criteria for a specific program, a laboratory accreditation body should have procedures for acting on such results.

12.3 The procedures for acting on results should include early reporting to the laboratory of its results with an invitation for the laboratory to investigate and comment on its performance.

NOTE 4—Some proficiency testing programs take considerable time to complete, particularly where participants are sequentially provided with the same artifact to test, measure or calibrate. In such cases, it is desirable that the laboratories be provided with interim reports on their performance, and particularly if their reported results are unsatisfactory. This will allow investigation and any subsequent corrective action to be taken quickly without awaiting publication of a final report from the program.

12.4 For laboratories submitting unsatisfactory results the laboratory accreditation body should have policies to:

12.4.1 Have the laboratory investigate and comment on its performance within an agreed time-frame.

12.4.2 Where necessary, have the laboratory undertake any subsequent proficiency test that may be available to confirm that any corrective actions taken by the laboratory are effective.

12.4.3 Where necessary, have on-site assessment of the laboratory by appropriate technical assessors to confirm that corrective actions are effective.

12.5 The laboratory accreditation body should advise participating laboratories of the possible outcomes from unsatisfactory performance in a proficiency testing program. These may range from continuing accreditation subject to successful attention to corrective actions within agreed time frames; temporary suspension of accreditation for the relevant tests, subject to corrective action, through to withdrawal of accreditation for the relevant tests. Normally, the options selected by a laboratory accreditation body will depend on the history of performance of the laboratory over time and from the most recent on-site assessments.

12.6 The laboratory accreditation body should have procedures to ensure that the records of performance of laboratories in proficiency testing programs are maintained (in accreditation files or records) for the participating laboratories and are made available to technical assessors for on-site assessments.

13. Action and Feedback by Laboratories

13.1 Laboratory accreditation bodies should have policies for feedback from accredited laboratories of action taken from results of proficiency testing programs, particularly for unsatisfactory performance.

13.2 Accredited laboratories should be required to maintain their own records of performance in proficiency testing, including the outcomes of investigations of any unsatisfactory results and any subsequent corrective actions.

14. Keywords

14.1 interlaboratory comparison; laboratory; laboratory accreditation; proficiency testing

ANNEXES

(Mandatory Information)

A1. STATISTICAL METHODS FOR EVALUATING PROFICIENCY TESTING DATA

A1.1 Proficiency test results come in many forms, spanning the range of data types and underlying statistical distributions. The statistical techniques need to be appropriate for each situation, and so are too varied to prescribe. There are, however, three basic steps common to all proficiency tests, when participant results are evaluated:

(*1*) Determine the assigned value, or correct response.

(*2*) Comparison of results: calculate the performance statistic.

(*3*) Determine performance.

And in many cases:

(*4*) Monitoring performance over time.

A1.2 Monitoring performance across proficiency test (“pt”) events shows the participant the effect of any changes they have made, and shows the variability of “pt” scores. Where appropriate, the performance measures used for evaluation should be used for monitoring performance.

A1.3 This annex gives general criteria for statistical techniques that can be applied as needed to guide specific applications.

A1.4 Different statistical techniques may be appropriate for situations where interlaboratory agreement is stable and proficiency testing is in place. With new interlaboratory comparison programs, agreement is often poor, due to new questions, new forms, the artificial test item, genuinely poor agreement of methods, or variable laboratory procedures. Analysts may have to use robust measures of relative performance, such as percentiles, until agreement improves.

A1.5 The annex does not consider statistical techniques for analytical studies other than proficiency test data. These studies would be to accomplish the other uses of interlaboratory comparison data, as listed in the introduction.

A1.6 All statistical analyses using proficiency test data should be appropriate for the design of the proficiency test, which may have any or all of the following limitations:

A1.6.1 Single test results, limiting the ability to estimate random error.

A1.6.2 Test item inhomogeneity.

A1.6.3 Differences between proficiency testing and usual lab practice:

A1.6.3.1 Test on an artificial test item.

A1.6.3.2 Coordinator’s processing instructions and forms.

A1.7 *Determine the Assigned Value, or Correct Response:*

A1.7.1 The participant results should be compared to the value or answer that best demonstrates competence with the method. Typically, this is a single number or classification, but can be a range of numbers or set of classifications.

A1.7.2 Assigned values should be chosen to evaluate participants fairly, yet to encourage interlaboratory and intermethod agreement. This is accomplished through selection of common comparison groups wherever possible, and the use of common assigned values.

A1.7.3 There is a hierarchy of preference for the assigned values, relating to their accuracy:

A1.7.3.1 Known values, with results determined by either:

(1) Expert consensus (for categorical responses); or

(2) Known by test item formulation (for example, manufacture or dilution).

A1.7.3.2 Certified Reference Material, or concentration given by definitive method analysis (with known standard error).

A1.7.3.3 Estimated by Reference Laboratory Consensus:

(1) Mean, may be weighted in some situations;

(2) Standard error, based on design and number of laboratories;

(3) Consensus response category.

A1.7.3.4 Estimated by Participant Consensus:

(1) "Average" for an appropriate comparison group:

(*a*) Mean, may be weighted or transformed (for example, geometric mean),

(*b*) Median, preferred for small groups or skewed distributions,

(*c*) Mode, may be preferred for some ordinal measures, or instances where there are few response classes;

(2) Variability for appropriate comparison group:

(*a*) Standard deviation,

(*b*) Coefficient of variation (CV),

(*c*) Percentiles (for example, interquartile range); or

(3) Consensus of a pre-determined majority percentage.

A1.7.4 *Considerations*:

A1.7.4.1 Outlier techniques are needed when participant results are used to determine assigned values and for summary statistics. Outliers are identified so that the summary statistics (mean and variance) are not influenced by extreme and inappropriate results. In smaller, new, or isolated proficiency tests, all possible outliers should be examined using techniques in Practice E 178. In larger or routine programs, it may be possible to have automated outlier screens.

A1.7.4.2 If assigned values are determined by reference or participant consensus, the coordinator must have an on-going process to test the accuracy of the assigned values. Accuracy needs to be tested relative to a value from level 1 or 2, whenever reference methods are available.

A1.7.4.3 The coordinator should have criteria for the error of an assigned value, based on the value's uncertainty (SE) and known biases. Values with wide uncertainty limits may not be suitable. For example, the ISO/REMCO N280 1993[4] document recommends that uncertainty in reference values be no more than 0.3 times sigma, which is 1/10 of the evaluation interval.

A1.7.4.4 The coordinator should have a protocol for determining that proficiency test items react the same as typical laboratory specimens. These methods would be evaluated with assigned values appropriate for their method. One method to do this is given in the NCCLS EP14-P document.[5]

A1.8 *Comparison of Results: Performance Statistics:*

A1.8.1 *Performance on Single Test Results*:

A1.8.1.1 The proficiency test results often need to be transformed into a performance statistic, to aid interpretation and to allow comparison with defined goals.

A1.8.1.2 The objective is to measure the deviation from the assigned value in a manner that allows comparison with performance criteria. Techniques may range from no processing required to complex statistical transformations. However, there should be as little alteration as possible to the participant's result.

A1.8.1.3 Commonly used statistics are listed below:

(1) Difference, or error $D = (x - X)$;

(2) Percent difference $(D/X) \times 100$ %;

(3) Percentile or rank of D within a comparison group;

(4) Transformed difference, such as $\log(1 + |D|)$ or $(D) \times 2$;

(5) Difference adjusted for random error (D/s):

With s = group standard deviation–(Z scores);
s = estimate of error for the assigned value;
s = combined estimate for uncertainty in the assigned value and in the participant results (En scores).

A1.8.1.4 *Considerations*:

(1) The simple difference between the participant result and the assigned value may be adequate to determine performance, and is most easily understood by participants.

(2) The percent difference adjusts for concentration, and is well understood by participants.

(3) Percentile or rank is useful for highly dispersed or skewed results, ordinal responses, or when there are a limited number of different responses.

(4) Transformed results may be preferred, depending on the nature of the test. For example, dilution-based results are a form of geometric scale, transformable by logarithms. For other tests, squared error may be the appropriate statistic to use.

(5) If statistical criteria (for example, Z-scores), are used, the estimates of variability must be reliable, that is, based on enough observations to allow outlier detection and have low standard error. For example, at least, $n = 20$ is often recommended for reliable measures of the standard deviation.[6] The data also must be screened for significant outliers prior to calculation.

A1.8.2 *Combined Performance Scores*:

A1.8.2.1 Performance may be evaluated at higher levels, for example, overall competency on a particular test or family of tests based on more than one result. Scores might be for a single proficiency test event, or for consecutive test events.

(1) Composite score for the same test:

(*a*) Scores combined for different test items, same test event and across different proficiency test events.[7]

Number of satisfactory results
Average absolute Z score
Average absolute error (in units or percent)
Summed absolute error (or squared error)
Other transformations

(2) Composite score for different tests:

(*a*) Scores combined within the same proficiency test event and across different test events.

[4] "International Harmonized Protocol for Proficiency Testing of (Chemical) Analytical Laboratories," *Journal of AOAC International*, Vol 76, No. 4, 1993.

[5] "Evaluation of Matrix Effects: Proposed Guideline," NCCLS Document EP14-P, National Committee for Clinical Standards, Villanova, PA, 1994.

[6] Youden, W. J., *Statistical Techniques for Collaborative Tests*, Association of Official Analytical Chemists, Washington, DC, 1967, pp. 29–32.

[7] Tholen, D. W., "A Statistical Procedure for Measuring and Evaluating Performance in Interlaboratory Comparison Programs," *Archives of Pathology and Laboratory Medicine*, Vol 112, 1988, pp. 462–470.

Number (or %) of satisfactory results
Average absolute *Z* score
Average absolute error relative to the evaluation limits[7]

(3) Composite score for general laboratory performance:

(*a*) Number (or %) of satisfactory responses

(*b*) Number (or %) of tests performed competently

A1.8.2.2 *Considerations*:

(1) Performance measures should have preferred statistical attributes, such as being unbiased, efficient, and low variability.

(2) Scores may be transformed (if necessary) so that they all follow the same assumed distribution (for example, Normal (0.1) for *Z* scores).

(3) There should be a check for extreme values which could heavily influence a quantitative composite score.

(4) In some laboratories there may be several levels of composite scoring decisions other than the single accreditation decision. Performance statistics may need to be refined at each level, as the scores become more general.

A1.9 *Determine Performance:*

A1.9.1 Results that have the same answer as the assigned value should be given the best evaluation achievable (such as "satisfactory"). If there is any difference in the answers, expert judgment is needed to determine the importance of the difference.

A1.9.2 Evaluation levels can be binary (Satisfactory/Unsatisfactory), or more than two levels (Good/Acceptable/Unacceptable).

A1.9.3 In decreasing preference, criteria for performance evaluation should be established by:

A1.9.3.1 Expert consensus, where the advisory group, or other qualified experts, determine the answers to be considered satisfactory.

Typical for tests with categorical answers
"State of the Art" expectations

A1.9.3.2 User needs for test performance, such as:

Accuracy needs for the test application in the same units as the performance statistic, such as ±0.1 g for untransformed error, ±10 % for percentage error, and ±1 dilution for dilution tests (geometric).
Extent of identification for participant's level of expertise, correct to genus or to species?
Method performance specifications for bias, repeatability and reproducibility.

A1.9.3.3 Statistical determination for scores where error was corrected for variance. Criteria must be appropriate for each score.

Possible criteria for *Z* scores: | *Z* |Cr 2 = satisfactory, 2 < | *Z* | < 3 = questionable, and | *Z* | > 3 = unsatisfactory.
For measurement comparison programs: *En* < 1 = satisfactory, and *En* > = 1 = unsatisfactory.

A1.9.3.4 Consensus of participants:

The range of scores or answers used by some percentage of all participants, or from a predefined reference group.
Central percentage (80 %, 90 %, or 95 %) satisfactory
One sided percentage (lowest 90 %) satisfactory

A1.9.4 *Considerations:*

A1.9.4.1 With criteria from A1.9.3 and A1.9.3.2, it is possible for all participants to have satisfactory scores, or all be unsatisfactory. With most statistical measures and with consensus criteria, some participants will have unsatisfactory scores, but most will be satisfactory.

A1.9.4.2 Criteria determined by user needs are generally easiest for participants to understand and support.

A1.9.4.3 Criteria determined by *a priori* statistical assumptions allow traceability to the statistical model, which may incorporate goals for Type 1 and Type 2 error. (A Type 1 error is an incorrect decision to reject acceptable data. A Type 2 error is an incorrect decision to accept unacceptable data.)

A1.9.4.4 Using consensus or percentile ranges to determine performance is less desirable because of arbitrary assignment of unsatisfactory scores.

A1.9.4.5 Graphs should be used whenever possible to show performance:

Showing distributions of participant values.
Relationship between responses on multiple test items.
Comparative distributions for different methods.

A1.10 *Monitoring Performance Over Time:*

A1.10.1 Accreditation decisions may require an overall summary of laboratory performance. Ideally, these decisions will consider general patterns of performance, using different tests, test families, and different testing events. Proficiency test failures may be an indicator that the accreditation status needs review; continued success in proficiency testing may indicate that competency is stable, and review is not warranted.

A1.10.2 The proficiency test program may therefore include techniques to monitor performance over time. In addition to the above accreditation decisions, the monitors should allow participants to see the variability in their measures, whether there are general trends or consistencies, and where the measures vary randomly. They also can see the effects of any changes they make in their methods, such as recalibration.

A1.10.3 Graphical methods should be used to facilitate interpretation by a wider variety of readers. Tabular data results allow more detailed review. Statistics used to evaluate performance should be used for these presentations.

A1.10.4 Ideal monitors are difficult if not impossible to develop. To do so requires a sound theory of the distribution of performance within and between participants. Criteria for the measures should be based on an independent competency assessment, such as on-site assessment or audit findings.

A2. QUALITY ASSURANCE OF PROFICIENCY TESTING PROGRAMS

A2.1 A documented quality assurance system should be established and maintained. It should outline the policies and procedures that exist to ensure the quality of the interlaboratory testing service provided. The coordinator's organization should conform to the requirements of quality assurance and technical competence based on the appropriate parts of the ANSI/ISO/ASQC Q9000 series and Guide E 548 (or ISO/IEC Guide 25), or both, as demonstrated by registration or accreditation, or both, by one or more recognized bodies.

A2.2 Suggested topics in a quality manual for organization of proficiency testing programs are:

A2.2.1 Quality policy;
A2.2.2 Organization of coordinator;
A2.2.3 Personnel, including responsibilities;
A2.2.4 Document and data control;
A2.2.5 Audit and review procedures;
A2.2.6 Aims, scope, statistical design and format (including frequency) of proficiency testing programs;
A2.2.7 Procedures covering:
A2.2.7.1 Sample preparation;
A2.2.7.2 Testing of test item homogeneity;
A2.2.7.3 Equipment;
A2.2.7.4 Suppliers;
A2.2.7.5 Logistics (for example, sample distribution);
A2.2.7.6 Analysis of data;
A2.2.7.7 Methods used to assign values; and
A2.2.7.8 Classification of ungradable test items;
A2.2.8 Preparation and issuing of reports;
A2.2.9 Action and feedback by participants, when required;
A2.2.10 Documentation of records for each program;
A2.2.11 Complaint handling procedures;
A2.2.12 Policies on confidentiality and ethical considerations;
A2.2.13 Computing information, including maintenance of hardware and software;
A2.2.14 Safety and other environmental factors;
A2.2.15 Subcontracting;
A2.2.16 Fees for participation;
A2.2.17 Scope of availability of program to others; and
A2.2.18 General policies on participation and on use of results from programs.

APPENDIX

(Nonmandatory Information)

X1. BIBLIOGRAPHY

American Society of Crime Lab Directors (ASCLD)/LAB, *Proficiency Test Monitoring* , 1994.

Manual on Presentation of Data and Control Chart Analysis, ASTM STP 15D, 1976.

ASTM E 177, Practice for Use of the Terms Precision and Bias in ASTM Test Methods.

Burnett, R. W., "Accurate Estimation of Standard Deviations for Quantitative Methods in Clinical Chemistry," *Clinical Chemistry* Vol 12, No. 13, 1975, pp. 1935–1938.

Dixon, W. J., "Processing Data for Outliers," *Biometrics*, BIOMA, March 1953, Vol, No. 1, pp. 74–89.

Esche, C. A., Groff, J. H., Schlecht, P. C., and Schulman, S. A., "Laboratory Evaluations and Performance Reports for the Proficiency Analytical Testing (PAT) and Environmental Lead Proficiency Analytical Testing (ELPAT) Program," DHHS (NIOSH) Publication No. 94–128, 1994.

Gladhill, R., NISTIR 5428, *Program Handbook: Requirements for Obtaining NIST Approval/Recognition of a Laboratory Accreditation Body Under P. L. 101–592, The Fastener Quality Act*, Appendix D, Proficiency Testing Providers, August 1994.

Grubbs, F. E., "Procedures for Detecting Outlying Observations in Samples," *Technometrics*, TCMTA, Vol 11, No. 4, February 1969, pp. 1–21.

Horlick, J., "Interlaboratory and Intralaboratory Proficiency Testing," *Accreditation Practices for Inspections, Tests, and Laboratories*, Harvey E. Schock, Editor, *ASTM STP 1057*, 1989.

ILAC Committee #3, Working Group 2 Report on Proposed Redraft of ISO/IEC Guide 43, "Proficiency Testing by Interlaboratory Comparisons," July 1994.

ISO REMCO N286, Procedures for Statistical Evaluation of Data Resulting from International Tests, 1993.

Kafadar, K., "A Biweight Approach to the One-Sample Problem," *Journal of the American Statistical Association* , Vol 77, No. 378, June 1982.

Mandel, John, *The Statistical Analysis of Experimental Data*, Interscience-Wiley Publishers, New York, NY, 1964 (out of print); corrected and reprinted by Dover Publishers, New York, NY, 1984.

Mandel, John, and Lashof, T. W., "The Nature of Repeatability and Reproducibility," *Journal of Quality Technology*, Vol 19, No. 1, January 1987, pp. 29–36.

Skendzel, L. P., and Youden, W. J., "Systematic versus Random Error in Laboratory Surveys," *American Journal of Clinical Pathology* , Vol 54, No. 3, Part II, September 1970.

WECC 15, Western European Calibration Cooperation—Interlaboratory Comparisons, 1992.

WELAC WGD #, WELAC Criteria for Proficiency Testing in Accreditation, 1993.