

Australian Standard[®]

**Chemical analysis—Interlaboratory
test programs—For determining
precision of analytical method(s)—
Guide to the planning and conduct**

This Australian standard was prepared by a subcommittee of the Association's Committee CH/23, General Methods of Chemical Analysis. It was approved on behalf of the Council of the Standards Association of Australia on 19 April 1986 and published on 4 August 1986.

The following interests are represented on Committee CH/23:

Australian Government Analytical Laboratories
Australian Institute of Food Science and Technology
Australian Mineral Development Laboratories
Bureau of Steel Manufacturers of Australia
Commonwealth Scientific and Industrial Research Organization
Confederation of Australian Industry
National Association of Testing Authorities, Australia
National Health and Medical Research Council
Railways of Australia Committee
Royal Australian Chemical Institute
University of Sydney

The following interests also assisted in the production of this standard:

Kraft Foods Limited
Rural Water Commission of Victoria
SGS Australia Proprietary Limited
State Electricity Commission of Victoria
Tomago Aluminium Company Proprietary Limited

Review of Australian Standards. To keep abreast of progress in industry, Australian Standards are subject to periodic review and are kept up to date by the issue of amendments or new editions as necessary. It is important therefore that Standards users ensure that they are in possession of the latest edition, and any amendments thereto.

Full details of all Australian Standards and related publications will be found in the Standards Australia Catalogue of Publications; this information is supplemented each month by the magazine 'The Australian Standard', which subscribing members receive, and which gives details of new publications, new editions and amendments, and of withdrawn Standards.

Suggestions for improvements to Australian Standards, addressed to the head office of Standards Australia, are welcomed. Notification of any inaccuracy or ambiguity found in an Australian Standard should be made without delay in order that the matter may be investigated and appropriate action taken.

This Standard was issued in draft form for comment as DR 84042.

Australian Standard[®]

Chemical analysis—Interlaboratory test programs—For determining precision of analytical method(s)— Guide to the planning and conduct

First published 1986

PUBLISHED BY STANDARDS AUSTRALIA
(STANDARDS ASSOCIATION OF AUSTRALIA)
1 THE CRESCENT, HOMEBUSH, NSW 2140

ISBN 0 7262 4229 2

 **Documents**
Document Sharing Hub

PREFACE

This standard was prepared by the Association's Committee on General Methods of Chemical Analysis under the direction of the Chemical Standards Board. It has been prepared to assist persons and organizations involved in the development and testing/validation of methods of chemical analysis.

The standard provides guidance in choosing test materials, laboratories and test procedures to select and validate methods of chemical analysis and includes the subsequent determination of the precision of the test methods. It is based on ISO 5725—Precision of Test Methods—Determination of Repeatability and Reproducibility by Interlaboratory Tests, but has been expanded to include procedures which assist in the validation of test methods, and modified to aid its clarification and application.

It is suggested that the procedures described herein may, with simple but appropriate modification, be applied to many areas of physical testing also, e.g. viscosity determinations, fracture testing, density determinations.

It should be noted that proposals for expanding this standard, particularly with respect to including more details of data evaluation, have already been received by the committee and will be considered for inclusion in the second edition. It was felt that the proposals required far more deliberation and discussion than was considered acceptable for further delaying publication of this workable and much needed standard.

© Copyright — STANDARDS AUSTRALIA

Users of Standards are reminded that copyright subsists in all Standards Australia publications and software. Except where the Copyright Act allows and except where provided for below no publications or software produced by Standards Australia may be reproduced, stored in a retrieval system in any form or transmitted by any means without prior permission in writing from Standards Australia. Permission may be conditional on an appropriate royalty payment. Requests for permission and information on commercial software royalties should be directed to the head office of Standards Australia.

Standards Australia will permit up to 10 percent of the technical content pages of a Standard to be copied for use exclusively in-house by purchasers of the Standard without payment of a royalty or advice to Standards Australia.

Standards Australia will also permit the inclusion of its copyright material in computer software programs for no royalty payment provided such programs are used exclusively in-house by the creators of the programs.

Care should be taken to ensure that material used is from the current edition of the Standard and that it is updated whenever the Standard is amended or revised. The number and date of the Standard should therefore be clearly identified.

The use of material in print form or in computer software programs to be used commercially, with or without payment, or in commercial contracts is subject to the payment of a royalty. This policy may be varied by Standards Australia at any time.

CONTENTS

	<i>Page</i>
1 SCOPE	4
2 APPLICATION	4
3 REFERENCED DOCUMENTS	4
4 DEFINITIONS	4
5 PRIOR DEVELOPMENT AND VALIDATION OF METHOD	5
6 SELECTION OF TEST MATERIALS	5
7 LABORATORY/PERSONNEL REQUIREMENTS	5
8 SETTING UP A TEST PROGRAM	5
9 EXECUTION OF TEST PROGRAMS	6
10 STATISTICAL ANALYSIS OF RESULTS	7
11 EVALUATION OF PRECISION DATA	8
12 EXPRESSION OF PRECISION DATA	8
 APPENDICES	
A STATISTICAL ANALYSIS AS A STEP BY STEP PROCEDURE	10
B STATISTICAL ANALYSIS FOR UNIFORM-LEVELS EXPERIMENTS	13
C STATISTICAL ANALYSIS FOR SPLIT-LEVELS EXPERIMENTS . . .	22
D CRITICAL DIFFERENCES DERIVABLE FROM REPEATABILITY AND REPRODUCIBILITY	25
E COMPUTER PROGRAM FOR CALCULATING REPEATABILITY AND REPRODUCIBILITY	27
F EXAMPLES OF INSTRUCTION/REPORTING SHEETS FOR ACQUIRING TEST DATA FROM PARTICIPATING LABORA- TORIES	38
F1 EXAMPLE 1 ANALYSIS OF HEAVY MINERAL SANDS	38
F2 EXAMPLE 2 ANALYSIS OF COAL/COKE	40
F3 EXAMPLE 3 pH OF WATER-BASED PAINTS (AS 1580.505.1)	41
F4 EXAMPLE 4 ALUMINIUM, PHOSPHORUS AND MOISTURE IN ALUMINIUM ORES	44
G ASSIGNMENT OF STATISTICAL SYMBOLS USED IN THIS STANDARD	46

STANDARDS ASSOCIATION OF AUSTRALIA

Australian Standard

for

**CHEMICAL ANALYSIS—INTERLABORATORY TEST PROGRAMS—
FOR DETERMINING PRECISION OF ANALYTICAL METHODS—
GUIDE TO THE PLANNING AND CONDUCT**

1 SCOPE. This standard sets out procedures for the planning and execution of interlaboratory test programs, including the evaluation of results, to determine the precision of methods of chemical analysis in terms of repeatability, within-laboratory reproducibility and reproducibility.

NOTE: Depending upon the plan selected for the test program (see Clause 8.1), and with due consideration of industrial practice, it may be (more) appropriate, within the context of the statistical treatment of results, that the term 'repeatability (r)' be regarded as being interchangeable with or additional to the term 'within laboratory reproducibility (R_w)' as defined in this standard.

2 APPLICATION. The procedures described are applicable to analytical methods that have been previously developed and validated in terms of concentration range, interfering elements and operating parameters to ensure that technical difficulties and bias have been minimized and methods are rugged and suitable for the purpose for which they have been designed.

The procedures are applicable to participating laboratories which are reasonably experienced in the methods to be evaluated, and to test materials which are homogeneous so that any variability in the material, as a component of the total analytical error, is minimal.

3 REFERENCED DOCUMENTS. The following standards are referred to in this standard:

- | | |
|----------|---|
| AS 2667 | Chemical Testing—Sampling—Glossary of Terms |
| AS 2706 | Numerical Values—Rounding and Interpretation of Limiting Values |
| BS 4237 | Report on Reproducibility of Methods of Chemical Analysis Used in the Iron and Steel Industry |
| BS 5497 | Precision of Test Methods
Part 1—Guide for the Determination of Repeatability and Reproducibility for a Standard Test Method |
| ISO 5725 | Precision of Test Methods—Determination of Repeatability and Reproducibility by Interlaboratory Tests. |

4 DEFINITIONS. For the purpose of this standard, the following definitions apply. For additional definitions refer to AS 2667.

4.1 Accuracy—the closeness of agreement between the true value and the mean of the results.

NOTE: The smaller the systematic component of the experimental errors which affect the results, the more accurate is the procedure.

4.2 Bias—an error which has the same value under a variety of conditions and remains constant from one

measurement to another. Such an error is not reduced by taking the average of repeated measurements.

NOTE: Bias may also be termed 'constant error' or 'systematic error'.

4.3 Cell—each combination of a laboratory and a concentration level is called a cell of the precision experiment.

4.4 Critical difference—the value at or below which the absolute difference between two test results, obtained under specified conditions, may be expected to lie with a specified probability.

NOTE: Both repeatability and reproducibility are special types of critical difference.

4.5 Critical value—the value of the test statistic (for a statistical test) above which a given percentage of values for that test statistic may be expected to lie.

4.6 Outlier—an experimental result from which a test statistic that is greater than its 1 percent critical value has been calculated.

4.7 Precision—the closeness of agreement between results obtained by applying the experimental procedure several times under prescribed conditions.

NOTE: The smaller the random part of the errors which affect the result, the more precise is the procedure.

4.8 Range—the absolute difference between the greatest and least values observed.

4.9 Repeatability (r)—the value at or below which the absolute difference between two single test results obtained with the same method on identical test material, under the same conditions (same operator, same apparatus, same laboratory, and the minimum practical time consistent with separate tests), may be expected to lie with a specified probability.

NOTE: In the absence of other specifications, the probability is 95 percent.

4.10 Replicate determinations—one or more additional determinations carried out under repeatability (or within-laboratory reproducibility) conditions on separate test portions of the same test material.

4.11 Reproducibility (R)—the value at or below which the absolute difference between two single test results obtained with the same method on identical material obtained by operators in different laboratories may be expected to lie with a specified probability.

NOTE: In the absence of other specifications, the probability is 95 percent.

4.12 Single test result—the value obtained by applying the standard test method fully once to a single test portion.

4.13 Split-levels experiment—an exercise in which each substantially different concentration level selected for examination is split into two sub-levels of slightly different concentrations. A single sample at each sub-level of the determinand is sent to each laboratory which is instructed to perform a single determination at each sub-level.

NOTE: Appendix C provides elaboration of this definition.

4.14 Straggler—an experimental result from which a test statistic that lies between the 5 percent and 1 percent critical values has been calculated.

4.15 Uniform-levels experiment—an experiment in which single samples, each representing a substantially different concentration level of the determinand, are sent to different laboratories which are instructed to perform replicate determinations at each concentration level.

4.16 Within-laboratory reproducibility (R_w)—the value at or below which the absolute differences between two single test results obtained on portions taken from the same test sample, at different times with or without different operators in the same laboratory using the same method, may be expected to lie within a specified probability.

5 PRIOR DEVELOPMENT AND VALIDATION OF METHOD. The method to be subjected to the inter-laboratory test program should be the most suitable for determination of the analyte under study in the particular matrix being considered. The method may be chosen from methods already in existence or may have to be developed from first principles.

Once the method has been chosen and suitably documented it should be validated for the specified concentration range, preferably by more than one operator, and tested to ensure that bias and interference problems have been minimized for the test material (refer also to Clause 2).

The validation procedure should include the following:

- (a) Aspects of analyte recovery over the whole of the specified concentration range.
- (b) Identification, correction and/or elimination of interference effects.
- (c) Assessment of the accuracy of the method using appropriate certified reference materials, where available.

Finally, the method should be so written that it is unambiguous and can be easily interpreted and correctly used.

6 SELECTION OF TEST MATERIALS. The following criteria shall be taken into consideration when test materials are chosen:

- (a) The appropriateness of the materials for the analytical method being tested, e.g. natural iron ores for the determination of the manganese content in iron ores.
- (b) The availability of sufficient material to provide adequate reserves for subsequent additional testing and for possible use as reference materials.
- (c) The homogeneity and form of the test material. The materials shall be as homogeneous as possible and of suitable form, for example with respect to shape, mass and/or particle size. The sampling procedure

for the method shall ensure that the level of homogeneity is maintained for the mass of test portion taken (see also Clause 9.2(g)).

NOTE: Every care should be taken in producing and distributing samples to ensure homogeneity is maintained.

- (d) Where appropriate, the provision of a sufficient variety of test materials to cover the scope, differing matrix types and concentration range(s) to which the method is applicable.
- (e) Where possible, the inclusion of one or more unidentified certified reference materials in the test materials to provide information on accuracy as well as precision.
- (f) The stability or the need to stabilize the samples with respect to the property of interest for the duration of the test program (see also Note to Clause 9.2(g)).

7 LABORATORY/PERSONNEL REQUIREMENTS.

7.1 Qualifications and experience of personnel.

Laboratories/operators selected for the test programs shall be appropriately qualified and shall have experience with the particular analytical technique and test material. Each laboratory shall have adequate facilities, equipment and reagents specified for carrying out the method. Where applicable, the participants shall carry out calibration and performance checks of equipment immediately prior to commencing the program.

7.2 Laboratory—adherence to test schedule. Participating laboratories shall be required to indicate their ability to complete the tests as prescribed within the specified time.

8 SETTING UP A TEST PROGRAM.

8.1 Choice of experiment. Before setting up a test program the organizer(s) (refer Clause 9.1) shall evaluate the proposed method to determine the following:

- (a) Whether a uniform-levels or a split-levels experiment is to be carried out. A split-levels experiment is a statistically efficient method of determining precision and avoids the possibility that, unlike a uniform-levels experiment, the operator is influenced by the result of his first test. It is therefore recommended that whenever possible the split-levels experiment be used in preference to the uniform-levels experiment. However, it may be difficult to find sufficient suitable samples for carrying out a split-levels experiment, especially in the analysis of naturally occurring materials, and in those cases a uniform-levels experiment is necessary.
- (b) The levels of within-laboratory precision to be determined. If repeatability is required, the program for a uniform-levels experiment shall be carried out within a 'short period of time' i.e. concurrently. If within-laboratory reproducibility is required, replicate determinations must be carried out at different times e.g. different days, and with the appropriate recalibration.

8.2 Number of laboratories. Laboratories which are to participate should be selected at random from a number of suitably experienced laboratories. As many laboratories as possible should be involved with a recommended minimum of five laboratories.

If five laboratories cannot be recruited, the test program may be carried out using a reduced number of laboratories. However, this should be regarded as a *serious limitation* in the interlaboratory program and the precision data so obtained should be viewed with caution.

Where only a small number of laboratories is available, it may be desirable, at the discretion of the convenor, to obtain additional data by conducting a further test program (see also Clause 8.4.1).

8.3 Number of concentration levels. The number of concentration levels to be used should cover the range of concentrations for which the method is applicable, bearing in mind the cost of performing tests. A minimum of three concentration levels should be used, preferably at evenly-spaced intervals.

If it is known that the precision depends upon the concentration level, at least four levels are needed to establish adequately the relationship between these quantities. However, if the concentration range to which the method is applicable is large, additional levels will be required.

8.4 Number of replicate determinations.

8.4.1 Uniform-levels experiment. A minimum of two determinations shall be carried out at each level by each laboratory. It should be noted, especially when the number of participating laboratories is small, that a larger number of replicates may be advisable for a uniform-levels experiment. However, increasing the number of replicates only increases the information available on the within-laboratory component of the error, and does not significantly affect the information available on the between-laboratory component of the error which is usually the larger component.

The number of replicate determinations for a uniform-levels experiment shall be as follows:

- (a) For more than 10 laboratories—2 determinations per level per laboratory.
- (b) 8 to 10 laboratories—3 determinations per level per laboratory.
- (c) 5 to 7 laboratories—4 determinations per level per laboratory.

8.4.2 Split-levels experiment. Only one determination shall be made in each participating laboratory at each of two sub-levels for each concentration level.

8.5 Number of operators per laboratory. If it is considered relevant to the program, this should be qualified.

8.6 Combining information. The minimum number of laboratories, levels and replicate determinations given above (Clause 8.2 to Clause 8.4) represent, in many cases, a considerable cost in time and effort and may relate to the size of interlaboratory trials that are practically feasible. The accuracy of the estimates of variance components, from which repeatability and reproducibility are calculated, is, however, critically dependent upon these numbers. Thus for crucial properties, when a large trial is not feasible, it is necessary to conduct a series of trials, to stagger the costs and effort over several years, leading to a gradual increase in the information on variances and an updating of the repeatability/reproducibility estimates. In such cases the initial estimates should be regarded as interim values.

9 EXECUTION OF TEST PROGRAMS.

9.1 Personnel and responsibilities.

9.1.1 Composition of the working group. The working group organizing the test program and assessing results should ideally comprise a convenor, a statistician and one or more technologists who are closely involved in the use of the method under examination. Their responsibilities shall be as follows.

9.1.2 The convenor. The convenor shall be responsible for:

- (a) Organizing preparation and distribution of test samples (see also Notes to Clause 9.2(j)).
- (b) Distribution of—
 - (i) the detailed method;
 - (ii) instructions for carrying out the test; and
 - (iii) format for reporting results.
- (c) Setting the closing date for return of results.

9.1.3 The statistician. The statistician shall be responsible for:

- (a) Ensuring that the test program is designed to obtain maximum information.
- (b) Evaluating the results.
- (c) Preparing a report.

NOTE: If a statistician is not available, the convenor should be responsible for ensuring that the statistical principles laid down in this standard are followed and the advice of a statistician is obtained, if deemed necessary.

9.1.4 The technologist(s). The technologist(s) shall be responsible for providing specific guidance on matters listed in Clause 9.2.

9.2 Preliminary considerations. The following matters should be discussed by the working group:

- (a) Whether the method has been satisfactorily validated in accordance with Clause 5 prior to being submitted to the test program.
- (b) The range of concentrations encountered in practice.
- (c) Whether a uniform-levels or split-levels experiment is to be conducted.
- (d) Whether the repeatability (r) or within-laboratory reproducibility (R_w) is required in addition to the reproducibility (R).
- (e) The number of concentration levels that should be used in the experiment.
- (f) Materials suitable for representing these concentration levels.
- (g) The need for any special homogenization of test materials prior to subsampling and distribution for testing.

NOTE: When the material to be tested is not homogeneous, it is important to prepare the samples in the manner prescribed by the method, preferably starting with one batch of commercial material for each level. Provision should be made to ensure that the amount of material available is sufficient to cover the experiment and for a reserve stock. For the samples at each concentration level, a number of separate containers should be used where there is any danger of the material deteriorating once the container has been opened, e.g. as with hygroscopic or oxidizable material, or material containing volatile components. In the case of unstable materials, special instructions on storage and treatment should be prescribed.

In general, it is recommended that the materials used in a precision experiment and the range of materials to which r (R_w) and R therefore apply, be clearly specified.

- (h) The number of replicates (n) to be specified and the amount of test sample to be sent to each laboratory (see also Clause 8.4.).
- (j) The need to send additional test sample to the laboratories for conducting familiarization tests prior to the performance of the official tests.

NOTES:

1. An operator will not, as a rule, achieve normal precision when he carries out a test for the first time or after a long interval. In such a case, the operators should be instructed to carry out a few unofficial trials on additional test samples, preferably including certified reference materials if available, in order to gain experience before the official precision experiment is undertaken. Whether this is needed should be decided by the working group or by the laboratory supervisor; test material for such familiarization tests should be supplied by the convenor.
2. Familiarization tests may be incorporated in a preliminary interlaboratory test program. This can lead to refinement and clarification of the prescribed method prior to conducting the final interlaboratory program for obtaining the precision data.

- (k) The number of laboratories required to participate in the experiment (see Clause 8.2.).
- (l) The manner of approach to prospective laboratories and the requirements they should satisfy (see Clauses 7 and 8.2.).
- (m) The instructions to be issued to the laboratory supervisors concerning the execution of the tests, and the number of figures to which the test results should be reported (see Clause 9.5.).
- (n) Any additional information to be requested besides the numerical test results (see Clause 9.5.).

9.3 Preparation for the test. Details of the method, the test conditions and the reporting procedure proposed for the interlaboratory test program should be distributed to all participating laboratories in advance of the official test, to enable the participants to make any comments necessary to assist in the correct interpretation of the proposed program. If, prior to making any comment, a laboratory considers it necessary to familiarize itself with the method by making preliminary tests, it should use different samples to those which are to be subjected to the interlaboratory test (refer also to Clause 9.2(j)).

9.4 Test instructions. Cooperation in an interlaboratory trial demands that discipline be observed by all participants so that the operators are thoroughly familiar with the details of the test procedures. Test instructions should include the following details:

- (a) The fully-described test method.
- (b) The nominal concentration level of each test sample.
- (c) Handling or storage instructions for the sample(s) (see NOTE to Clause 9.2(g)). These should be sent with the sample(s) before the test program gets underway.
- (d) A list of the samples to be tested and the order in which the samples are to be tested.

NOTE: A random order of testing, different for each laboratory as determined by the working group convenor, is recommended where possible.

- (e) Number of determinations per sample to be carried out by each operator.

NOTE: Conducting more tests than is officially prescribed, aside from familiarization tests (see Note, Clause 9.2(i)) should be strongly discouraged. This eliminates the temptation to the operator, of discarding the 'worst' data without the knowledge of the working group organizing the test program.

- (f) Date of commencement of test program, particularly when stability of the test sample(s) is a problem, and if it is desired that participants conduct the test at the same time or at the same elapsed time since the original sampling.

- (g) A report sheet or table (see Appendix F for examples) with instructions for reporting the results to ensure that the desired data and only the desired data is collected. This should include the name and address of the authority to whom the data is to be sent.

NOTES:

1. Advice on rounding of numbers is given in AS 2706.
2. Original observations and calculations should also be obtained if required for later scrutiny (see also Clause 9.5(b)).

- (h) A request for a description of any unusual occurrences or difficulties associated with the interpretation or implementation of the test method to supplement the test report.
- (j) A request for other relevant information obtained during the test, e.g. value of blank test, absorbance values obtained in spectrometric methods.
- (k) The date by which results and comments are required.

9.5 Reporting of results. The supervisor of each laboratory should complete the report form sheet. This report should contain the following particulars:

- (a) The test results, taking particular care to avoid computational, transcription and typing errors.
- (b) Where required, the original observations and/or calculations from which the final results were derived.
- (c) Any comments by the operator as well as information about irregularities or disturbances that may have occurred during the test (see also Clause 9.4(h)).
- (d) The date(s) on which the samples were received.
- (e) The date(s) and time(s) at which they were officially tested.
- (f) Information about the equipment used, when this is considered relevant (see Clause 9.4(h)).
- (g) Any other relevant information (see Clause 9.4(j)).

10 STATISTICAL ANALYSIS OF RESULTS.

10.1 General. Data generated from an interlaboratory test program should be treated according to the procedures detailed in Appendices A, B, C, D and E.

10.2 Personnel to analyze results. The analysis of test results is normally the task of the statistician who has taken part in planning the experiment (see Clause 9.1.3). Where a statistician is not available, the convenor of the working group should either consult a statistician or become sufficiently familiar with the principles involved to analyze the data satisfactorily.

10.3 Cells. In the ideal case, each cell will contain n results, that can all be used for computing the repeatability, the within-laboratory reproducibility, and the reproducibility. This ideal situation is not, however, always attained in practice. Departures may occur due to excess data, missing data, and outliers.

10.4 Excess data (see also Note to Clause 9.4(e)). All data shall be reported. The submission of excess data should be discouraged in the instructions to participants. Although a laboratory sometimes carries out more than the number of replicates officially prescribed, such excess data should only be submitted if there has been identifiable error in some of the determinations, in which case the laboratory supervisor should identify the erroneous results and give reasons for excess data. If all data is equally valid, it should all be used in the final calculation.

10.5 Missing data. Sometimes it may happen that some test results may be missing, for example, due to the loss of a sample, or a mistake in performing the test. The procedure recommended in this standard is that completely empty cells can be ignored, while partly empty cells can be taken into account. The reasons for missing test results should be included in the laboratory supervisor's report.

If one of the two test results in a cell of a split-levels experiment is missing, the result available must be discarded and the cell treated as an empty one.

10.6 Identification of outliers.

10.6.1 Cochran's and Dixon's tests. This standard uses Cochran's maximum variance test and Dixon's outlier test (both outlined in Appendix B) in combination with the following criteria for identifying outliers:

- (a) $P \geq 5\%$; that is, Cochran's or Dixon's test statistic is less than or equal to its 5% critical value: the item tested is accepted as correct.
- (b) $5\% \geq P \geq 1\%$; that is, the test statistic is greater than its 5% critical values but less than or equal to its 1% critical value: the item tested is called a *straggler* and is marked with a single asterisk.
- (c) $P < 1\%$; that is, the test statistic is greater than its 1% critical value: the item is called an outlier and is marked with a double asterisk.

P is the probability of the observed value of the test statistic.

10.6.2 Reasons for outliers and/or stragglers. Where stragglers and outliers occur (see Appendix A), attempts should be made to determine their cause, which may have been procedural error(s) in the conduct of the analysis, computational error(s), clerical error(s), analysis of a wrong sample, etc. When a reasonable explanation can be construed (and preferably confirmed by additional enquiries) the item is either considered as an outlier that does not belong to the experiment and is rejected, or it is corrected in keeping with the explanation obtained.

When any stragglers and/or outliers remain that have not been explained or rejected as belonging to an outlier laboratory, the stragglers are retained as correct items, and the outliers are rejected.

10.6.3 Outlier laboratories. When several unexplained stragglers and/or outliers occur at different concentration levels within the same laboratory, that laboratory may be considered as an outlier, having too high a within-laboratory variance, or too large a systematic error in the level of its test results, or both. It may then be necessary to discard all the data from such a laboratory (see Clause 10.6.2). The primary decision should be the

responsibility of the statistician, but it is important that the working group be notified of all such laboratories so that further action may be determined.

10.7 Calculation of repeatability and reproducibility. In accordance with the planning considerations given in Clause 8.1, the calculation of the repeatability, within-laboratory reproducibility and reproducibility is carried out according to either Appendices B or C for each concentration level using the data remaining after correction or elimination of the stragglers and/or outliers.

NOTE: A computer program* (see Appendix E) has been developed for the calculation of repeatability, within-laboratory reproducibility and reproducibility. The program is written in FORTRAN but could be modified to any other computer language, e.g. BASIC.

11 EVALUATION OF PRECISION DATA. After a test method has been subjected to an interlaboratory test program to determine its precision, the values achieved for repeatability, within-laboratory reproducibility and reproducibility should be evaluated critically by both the working group conducting the test program, and the committee which originally drafted the test method.

Aspects which should be considered include the following:

- (a) Whether the values for repeatability, within-laboratory reproducibility and reproducibility fall within the ranges originally required of the method.
- (b) Comparison of the precision data with those published for other available methods used to determine the same parameter.
- (c) The acceptability of the values for repeatability, within-laboratory reproducibility and reproducibility for commercial, regulatory, or other purposes, where the method under test may be nominated as the reference method.
- (d) Whether the test program resulted in identification of outlier results and/or laboratories, and whether their sources were investigated and identified.
- (e) Identification by the participants of any significant interpretation, or implementation problems associated with the method.
- (f) Whether qualifications should be placed upon the precision data to be published with the test method, e.g. the test program may have involved a limited group of sample types (or matrices), although it may be intended to cover a broader range of materials.

The outcome of discussions pertaining to the above, and other relevant matters may indicate a need to redraft the test method, or in some cases, may suggest that it be discarded and preference given to an alternative procedure. If significant modifications are made to the test method, it is essential that the redrafted method be subjected to a supplementary interlaboratory test program, and precision data determined for the new method.

12 EXPRESSION OF PRECISION DATA.

12.1 General. The results of a test program to determine the repeatability and reproducibility of an analytical method may be presented in that method either in tabular format or as a function of the concentration level.

* State electricity Commission of Victoria, 'A Program for Calculating the Repeatability and Reproducibility for a Standard Test Method from the Results of Interlaboratory Tests', December 1981, Report No GO/81/34. (A modified version of the computer program in BS 5497, Part 1.)

12.2 Tabular format. When the results of a test program are presented in a tabular format they may be presented in relation to either—

- (a) individual concentration levels (see Table 1); or
- (b) a concentration range (see Table 2).

12.3 Functional relationships. When four or more concentration levels have been used during the test program, the results may be able to be expressed as a function of the concentration level. Frequently a linear relationship can be used, e.g.

$$\begin{aligned} r &= a + bm \\ R_w &= a' + b'm \\ R &= A + Bm \end{aligned}$$

where

$a, a', A, b, b', B =$ constants

$m =$ the mean concentration of the element being determined.

A logarithmic relationship (e.g. $\log r = a + b \log m$) may also be considered.

12.4 Statement on repeatability, within-laboratory reproducibility and reproducibility. A statement linking the precision to the difference between two results and to the 95 percent probability level should also be inserted in the analytical method. A suggested wording is as follows:

‘The absolute difference between two single results found on an identical test material by one operator using the same apparatus within the minimum practical time consistent with separate tests will equal or

exceed the repeatability (r) on average not more than once in 20 cases in the normal and correct operation of the method.

The absolute difference between two single and independent results, found by different operators in the same laboratory or by the same operator and at a different time in the same laboratory on identical test material, will exceed the within-laboratory reproducibility (R_w), on average, not more than once in 20 cases in the normal and correct conduct of the method.

The absolute difference between two single and independent results found by two operators working in different laboratories on identical test material will exceed the reproducibility (R) on average not more than once in 20 cases in the normal and correct operation of the method.’

12.5 Suspect results. A statement may also be added indicating that both results shall be considered suspect if the repeatability, within-laboratory reproducibility or reproducibility, as appropriate, is exceeded. Statements regarding subsequent actions, e.g. repetition of the test, may also be included in the section on precision.

12.6 Reporting of experiment. In general, a brief mention of the precision experiment should be added to the precision section, possibly as a footnote. A suggested wording is as follows:

‘The precision data were determined from an experiment conducted in (year) involving (p) laboratories and (q) concentration levels and (n) replicates.’

TABLE 1
EXPRESSION OF THE REPEATABILITY, WITHIN-LABORATORY REPRODUCIBILITY AND REPRODUCIBILITY AS A FUNCTION OF INDIVIDUAL CONCENTRATION LEVEL

Sample identification	Mean concentration, percent	Repeatability, r or within-laboratory reproducibility, R_w	Reproducibility R
A B C D E			

TABLE 2
EXPRESSION OF THE REPEATABILITY AND REPRODUCIBILITY AS A FUNCTION OF THE CONCENTRATION RANGE

Concentration range percent	Repeatability, r or within-laboratory reproducibility, R_w	Reproducibility R
.....		

NOTE: The values to be shown for r , R_w and R relate to the upper level of each concentration range.

APPENDIX A

STATISTICAL ANALYSIS AS A STEP BY STEP PROCEDURE

INTRODUCTORY NOTE: In this Appendix, the ‘within laboratory reproducibility (R_w)’ as defined by Clause 4.16 has been treated as ‘repeatability (r)’ as defined by Clause 4.9 (see also Note to Clause 1).

The following is a detailed description of the procedure for the collection, examination and treatment of data generated by a test program:

- (a) Collect all available test results in one table e.g. Table A1.

NOTES:

1. It is recommended that this table be arranged into p rows, indexed $i = 1, \dots, p$, representing the p laboratories that have contributed data, and q columns, indexed $j = 1, 2, \dots, q$, representing the q concentration levels in increasing order.
2. In a uniform-levels experiment (see Appendix B) it is not necessary to distinguish between the test results within a cell of Table A1; they may be put in any desired order.
3. In a split-levels experiment (see Appendix C), clearly state which of the two test results belongs to sub-level A and which to sub-level B; enter the results in that specific order.

TABLE A1
ORIGINAL TEST RESULTS

Uniform-levels experiment							Split-levels experiment								
Laboratory \ Level	1	2		j		q	Laboratory \ Level	1		2		j		q	
								A	B	A	B	A	B	A	B
1							1								
2							2								
i				y_{ij1} $\vdots$ y_{ijk}			i					y_{ijA} y_{ijB}			
p							p								

Cell (i, j) contains n_{ij} results u_{ijk} ($k = 1, 2, \dots, n_{ij}$)

TABLE A2
CELL AVERAGES

Laboratory \ Level	1	2		j		q
1						
2						
i				$\bar{y}_{ij}$		
p						

$\bar{y}_{ij}$ = cell average

TABLE A3
MEASURES OF CELL SPREAD

Uniform-levels experiment						Split-levels experiment					
Laboratory \ Level	1	2	j		q	Laboratory \ Level	1	2		j	q
1						1					
2						2					
i			s_{ij} OR w_{ij}			i				d_{ij}	
p						p					

s_{ij} = cell standard deviation, or if $n = 2$ for all the cells, use
 w_{ij} = cell range

$d_{ij} = y_{ijA} - y_{ijB}$

- (b) Inspect Table A1 for any obvious irregularities. Investigate, and if necessary, discard any obviously erroneous data and report this to the working group.

NOTE: It is sometimes immediately evident that the test results of a particular laboratory or in a particular cell lie at a level inconsistent with the other data. Discard such obviously discordant data immediately, but report the fact to the working group for their further consideration.

- (c) From Table A1, corrected according to step (b) above when needed, calculate Table A2 containing the cell averages (see Clause B1) and Table A3 containing measures of within-cell spread.

NOTE: When a cell in Table A1 of a uniform-levels experiment contains only a single test result, adopt one of the options of Clause B5.1. Discard any single test result in a cell of a split-levels experiment (see Clause C1.2).

- (d) Inspect Tables A2 and A3, level by level, for possible stragglers and/or outliers (see Clauses B2 and B3 and Clause C2).
- (e) Apply the statistical tests of Clauses B2 and B3 to all suspect items and mark the stragglers with a single asterisk and the outliers with a double asterisk. If there are no stragglers or outliers, go to step (h).
- (f) Investigate whether there is, or may be, technical explanation for the stragglers and/or outliers and, if possible, verify such explanations.
- (g) Correct or discard, as required, those stragglers and/or outliers that have been satisfactorily explained, and apply corresponding corrections to the other tables.

If there are no unexplained stragglers or outliers left, go to step (h).

NOTES:

1. A large number of stragglers and/or outliers may indicate a pronounced variance in homogeneity or pronounced differences between laboratories, and thereby cast doubt on the suitability of the test method. This again should be reported to the working group.
 2. If the evidence against the suspected outlying laboratories is considered strong enough to justify the rejection of some or all data from these laboratories, discard the requisite data and report to the working group. The decision to reject some or all data from a particular laboratory is the responsibility of the statistician carrying out the analysis, but the working group needs to consider the action further (see Clause 9).
 3. If any stragglers and/or outliers remain that have not been explained or attributed to an outlying laboratory, discard the outliers but retain the stragglers.
 4. If in the previous steps any cell in Table A2 has been rejected, the corresponding cell in Table A3 shall also be rejected, and vice versa.
- (h) From the entries that have been accepted as correct in Tables A2 and A3, calculate for each level separately—
- (i) the mean level (m_j),
 - (ii) the repeatability (r_j) and
 - (iii) the reproducibility (R_j)

by the procedures of Clause B5 or Clause C4.

If only a single level has been used, or if it is decided that repeatability and reproducibility should be given for each level separately go to step (k).

- (j) Plot r_j and R_j against m_j and judge from these plots whether any of them depend on m . This decision cannot be made solely on statistical grounds. It will be influenced by experience and by technical information about the test method.

NOTES:

1. The number of levels used is often inadequate for determining whether a relationship exists between either r and m and/or R and m . It is doubtful whether a statistically valid relationship can be determined when the number of levels is less than four.
2. It cannot be assumed that a regular functional relationship usually exists between either r and m and/or R and m particularly where material heterogeneity forms an inseparable part of the variability of the test results—a relationship may only apply if the heterogeneity is a regular function of m . This is uncertain with solid materials of different composition and obtained from different production processes.

If it is judged that r and/or R is dependent on m then, in consultation with the technical experts, the statistician should decide which functional relationship best fits the data. This matter is not considered further in this standard, but has been well documented in ISO 5725. However, it should be noted that if a straight line relationship between r and/or R and m is to be fitted, the technique of weighted linear regression should be used because, statistically, the standard error of r is proportional to the value of r .

If it is judged that r is constant over levels then a combined estimate may be calculated from the separate level values (r_j). To allow for different amounts of replication, define—

$$d_j = \sum_i (n_{ij} - 1)$$

where

n_{ij} = number of replicates for the i th laboratory on the j th sample of material.

Then the combined estimate of r is

$$r = (\sum_i d_j r_j^2 / \sum_i d_j)^{1/2}$$

where

$j = 1, 2, \dots, q$.

When n_{ij} is a constant value, the formula becomes

$$r = (\sum_j r_j^2 / q)^{1/2}$$

A similar method for the combination of R -values is not available.

If it is judged that R and r are independent of m then methods are available in other literature, for the combination of data, using a two-way analysis of variance. This aspect is not dealt with in the present standard but will be considered for inclusion when this standard is due for revision.

- (k) When the final values of r and R have been established, verify that they correspond to a 95% probability, by means of the data from which they were calculated.

For example, if 10 operators each made 5 measurements on a test material, then there are 100 possible absolute differences between measurements, 10 for each operator, which correspond to the definition of repeatability. No more than 5 of these absolute differences should be greater than the calculated repeatability.

Verification may not be needed, but it may serve as a check on the correctness of the calculations for r and R .

APPENDIX B

STATISTICAL ANALYSIS FOR UNIFORM-LEVELS
EXPERIMENTS

INTRODUCTORY NOTE: In this Appendix, the within laboratory reproducibility (R_w) as defined by Clause 4.16 has been treated as repeatability (r) as defined by Clause 4.9 (see also Note to Clause 1).

B1 NOTATION, DEFINITIONS AND BASIC FORMULAS.

B1.1 Introduction. To calculate precision as prescribed in this standard, it is necessary to calculate the repeatability variance and the between-laboratory reproducibility variance. For each level the average of the individual laboratory standard deviations provides information on the repeatability variance, while the variation in the means from individual laboratories provides information on the between-laboratory reproducibility variance.

As stated in Clause 10.3 the ideal case is when p laboratories each perform n replicates at each of the q levels. However, because of the existence of excess results (see Clause 10.4), missing results (see Clause 10.5) outlier results (see Clause 10.6), or outlier laboratories (see Clause 10.6.3), this ideal situation is not always attained. In general for each level j —

p_j is the number of laboratories

n_{ij} is the number of results for laboratory i

y_{ijk} is any one of these results ($k = 1, 2, \dots, n_{ij}$)

To simplify the notation the subscript j will be dropped and therefore the procedures given in Clauses B1.2 to B4.4 apply to data from the same level. These procedures should be repeated for each of the individual levels.

B1.2 Original test results (Table A1). Table A1, consists of all the test results, with each column containing all the data from one level, and each row containing all the data from one laboratory. For each level—

n_i is the number of results for laboratory i

y_{ik} is any one of these results ($k = 1, 2, \dots, n_i$)

B1.3 Laboratory means (Table A2). For each level, the laboratory means (see Table A2) are derived from Table A1, as follows:

$$\bar{y}_i = \frac{1}{n_i} \sum_{k=1}^{n_i} y_{ik}$$

The averages should be given with one more significant figure than the test results in Table A1.

B1.4 Within-laboratory standard deviations (Table A3). For each level, the within-laboratory standard deviations (see Table A3) are derived from Tables A1 and A2, as follows:

$$s_i = \sqrt{\frac{1}{n_i - 1} \sum_{k=1}^{n_i} (y_{ik} - \bar{y}_i)^2}$$

$$= \sqrt{\frac{1}{n_i - 1} \left[\sum_{k=1}^{n_i} y_{ik}^2 - \frac{1}{n_i} \left(\sum_{k=1}^{n_i} y_{ik} \right)^2 \right]}$$

The standard deviations should be expressed with one more significant figure(s) than the test results in Table A1.

NOTES:

- For the particular case where all $n_i = n = 2$, it is more convenient to work with the laboratory range (w_i)—

$$w_i = |y_{i1} - y_{i2}| = s_i \sqrt{2}$$

The range is always greater than or equal to zero.

- The two formulas for s_i may be equal algebraically, but arithmetically this may not be so when real numbers are substituted e.g. where individual measurements are large and their differences small (ie. 1 000 000.01, 1 000 000.02, 1 000 000.04). However, the first form, using $\sum (y_{ik} - \bar{y}_i)^2$, can be depended upon at all times.