Standard Practice for
Conducting an Interlaboratory Study to Evaluate the Performance of an Analytical Method[1]

This standard is issued under the fixed designation E 1601; 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.

ε¹ NOTE—Caution notes were moved into the text editorially in November 2003.

1. Scope

1.1 This practice covers procedures and statistics for an interlaboratory study (ILS) of the performance of an analytical method. The study provides statistical values which are useful in determining if a method is satisfactory for the purposes for which it was developed. These statistical values may be incorporated in the method's precision and bias section. This practice discusses the meaning of the statistics and what users of analytical methods may learn from them.

1.2 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health practices and determine the applicability of regulatory limitations prior to use.*

2. Referenced Documents

2.1 *ASTM Standards:* [2]
E 135 Terminology Relating to Analytical Chemistry for Metals, Ores, and Related Materials
E 691 Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method
E 1169 Guide for Conducting Ruggedness Tests
E 1763 Guide for Interpretation and Use of Results from Interlaboratory Testing of Chemical Analysis Methods

3. Terminology

3.1 *Definitions*—For definitions of terms used in this practice, refer to Terminology E 135.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *interlaboratory study (ILS)*—study undertaken to ascertain if a test method is suitable for its intended use. The ILS includes preparation, testing, and evaluation phases.

3.2.2 *interlaboratory test*—measures the variability of results when a test method is applied many times in a number of laboratories.

3.2.3 *replicate results*—results obtained by applying a test method a specified number of times to a material.

3.2.4 *result*—the numerical value obtained by applying a test method once to a material.

3.2.5 *test method*—gives directions for producing a single result.

3.2.6 *test protocol*—gives instructions to each participating laboratory, detailing the way it is to conduct its part of the interlaboratory test program.

3.3 Unless the test method destroys the test portion each time it is applied, the protocol for a Plan A test specifies, if possible, replicate results on a single test portion (which may be in solution). The protocol for a Plan B test specifies the number of test portions of a material and requires duplicate results (2 only) on each portion (which may be in solution).

4. Summary of Practice

4.1 Instructions are provided for planning and conducting a cooperative evaluation of a proposed analytical method.

4.2 The following list describes the organization of this practice:

4.2.1 Sections 1-5 define the scope, significance and use, referenced documents, and terms used in this practice.

4.2.2 Section 6 helps users of analytical methods understand and use the statistics found in the Precision and Bias section of methods.

4.2.3 Sections 7 and 8 instruct the ILS coordinator and members of the task group on how to plan and conduct the experimental phase of the study.

4.2.4 Section 9 discusses the procedures for collecting, evaluating, and disseminating the data from the interlaboratory test.

4.2.5 Section 10 presents the statistical calculations.

4.2.6 Sections 11 and 12 discuss the use of statistics to evaluate a test method and the means of incorporating the ILS statistics into Precision and Bias statements.

4.2.7 The Annex A1 gives the rationale for the calculations in Section 10.

5. Significance and Use

5.1 Ideally, interlaboratory testing of a method is carried out by a randomly chosen group of laboratories that typifies the kind of laboratory that is likely to use the method. In actuality, this ideal is only approximated by the laboratories that are available and willing to undertake the test work. The coordinator of the program must ensure that every participating laboratory has appropriate facilities and personnel and performs the method exactly as written. If this goal is achieved, the statistics developed during the ILS will be adequate for determining if the method is capable of producing satisfactory precision in actual use. If the program includes certified reference materials, the test data also provide information concerning the accuracy of the method. The statistics provide a general guide to the expected performance of the method in the laboratories of those who will use it.

6. Statistical Guide for the Users of Analytical Methods Evaluated in Accordance With This Practice

6.1 *Standard Deviations*:

6.1.1 *Minimum Standard Deviation of Method, s_M*—This statistic measures the precision of test results under conditions of minimum variability. Because it is improbable that a method in ordinary use will exhibit precision this good, no predictive index is calculated for s_M. Users adept in statistics may wish to compare s_M and the short-term standard deviation of the method measured in their laboratory. For most methods, short-term variability refers to results obtained within several minutes. (**Caution**—The standard deviation of results obtained on different occasions, even in the same laboratory, probably will exceed s_M.)

6.1.2 *Between-Laboratory Standard Deviation, s_R*—This statistic is a measure of the precision expected for results obtained in different laboratories. It reflects all sources of variability that operate during the interlaboratory test (except test material inhomogeneity in tests designed to eliminate that effect). It is used to calculate the reproducibility index, R. Use s_R for evaluating the precision of methods. It represents the expected variability of results when a method is used in different laboratories.

6.1.3 *Within-Laboratory Standard Deviation, s_r*—This statistic cannot be calculated in a normal interlaboratory test. It is determined only in tests designed to measure variability within laboratories. When this statistic is given in a method, it reflects all variability that may occur from day-to-day within a laboratory (for example, from calibration, standardization, or environmental changes). It is used to calculate the repeatability index, r. The user is cautioned that additional sources of variation may affect results obtained in other laboratories.

6.2 *Predictive Indexes*—For the following indexes to apply, these conditions must be met: (*1*) the test materials must be homogeneous; (*2*) analysts must be competent and diligent; (*3*)

analytical instruments and equipment must be in good condition; and (*4*) the method must be performed exactly as written.

6.2.1 *Reproducibility Index, R*—This statistic estimates the expected range of differences in results reported from two laboratories, a range that is not exceeded in more than 5 % of such comparisons. Use R to predict how well your results should agree with those from another laboratory: First, obtain a result under the conditions stated in 6.2, then add R to, and subtract R from, this result to form a concentration confidence interval. Such an interval has a 95 % probability of including a result obtainable by the method should another laboratory analyze the same sample. For example, a result of 46.57 % was obtained. If R for the method at about 45 % is 0.543, the 95 % confidence interval for the result (that is, one expected to include the result obtained in another laboratory 19 times out of 20) extends from 46.03 to 47.11 %.

Note 1—For those not conversant with statistical concepts, it is important to realize that in most such comparisons, the differences will be much smaller than the confidence interval implies. The 50 % confidence interval is only about one third (34.6 %) as wide. Thus, the "average" interval for the above result (one expected to include the result obtained by another laboratory half the time) extends from 46.4 to 46.8 %. The obvious implication is that, although half the differences will be more than 0.2 %, half will be less than 0.2 %.

6.2.2 *Repeatability Index, r*—This statistic is given in the method only if the interlaboratory test was designed to measure s_r. It estimates the expected range of results reported in the same laboratory on different days, a range that is not exceeded in more than 5 % of such comparisons.

7. Interlaboratory Test Planning

7.1 Analytical test methods start from a perceived need to service one or more material specifications.

7.1.1 Develop a performance requirement for a method from the material specification(s). Include the following factors: expected ranges of chemical compositions of the materials to be covered (method's general scope); specified elements and their concentrations (determination concentration ranges); and the precision required.

7.1.2 Prepare a table of the elements and concentration ranges to cover the critical values in the material specifications. Use this information together with knowledge of the characteristics of the candidate analytical method to select test materials for the interlaboratory program.

7.2 *Draft Method*—The process of developing methods and testing them in a preliminary way is beyond the scope of this practice. All analytical skill and experience available to the task group must be exerted to ensure that the method will meet the project requirements in 7.1 and that it is free of technical faults. A preliminary, informal test of a method must be carried out in several laboratories before the final draft is prepared. Individuals responsible for selecting the method may find helpful information in Practice E 691 and Guide E 1169. The formal interlaboratory test must not start until the task group reaches consensus on a clearly written, explicitly stated, and unambiguously worded draft of the method in ASTM format, which has completed editorial review.

7.3 *Test Materials*—Appropriate test materials are essential for a successful ILS. The larger the number of test materials

included in the test program, the better the statistical information generated. On the other hand, the burden of running a very large number of materials may reduce the number of laboratories willing to participate. A method must cover a concentration range extending both above and below the specified value(s). If possible, provide test materials near each limit. Concentration ranges covering several orders of magnitude should be tested with three or more materials.

7.3.1 Material composition and form must be within the general scope of the method. If possible, include all material types the scope is expected to cover. Often, only limited numbers of certified reference materials are available. Use those that best meet the criteria for the test. If they do not cover all concentration levels, find or prepare other materials to fill in missing values.

7.3.2 The quantity of the material must be sufficient to distribute to all laboratories participating in the test with about 50 % held in reserve to cover unforeseen eventualities.

7.3.3 Materials should be homogeneous on the scale of the test portion consumed in each determination as well as among the portions sent to different laboratories. Usually certified reference materials have been tested for homogeneity, but test materials from other sources may have had only a minimal examination. The use of laboratory-scale melting and casting to produce test materials can sometimes lead to segregation of one or more components in an alloy. Unless specially gathered or prepared materials have been subjected to a thorough homogeneity test, they require the use of Test Plan B. It statistically removes the effect of moderate test material inhomogeneity from the estimates of the ILS statistics.

7.3.4 Test material sent to each laboratory must be permanently marked with its identity in such a manner that the identification is not likely to be lost or obliterated.

7.3.5 If the test program is to evaluate the accuracy of the method, at least one test material must be certified for the concentration of each element. More certified materials provide more complete information on accuracy.

7.3.6 Prepare a list of the test materials, their identifying numbers, a brief description of material type (for example, low-carbon steel), and approximate concentration of the elements to be determined. This table becomes part of the documentation sent to participating laboratories and provides information needed for the research report and the precision and bias statement.

7.4 *Number of Cooperating Laboratories*—Conventional wisdom holds that the more laboratories participating in an ILS, the better. Further, the laboratory types included in the study task group should consist of typical users' laboratories. There is wide agreement that estimates of precision based upon fewer than six laboratories become increasingly unreliable as the number decreases. A test program involving fewer than six laboratories does not comply with the requirements of this practice (Note 2). An effort should be made to enlist at least seven qualified laboratories before beginning a test program, to allow for attrition. To be qualified to participate, a laboratory must have proper equipment and personnel with sufficient training and experience to enable them to perform the method exactly as it is written.

Note 2—If all reasonable effort fails to recruit at least six cooperating laboratories, up to two of the recruited laboratories may each volunteer to submit two independent sets of test data as an expedient to provide a total of at least six sets of data. Minimum requirements for independence are that two typical analysts, who do not consult with each other about the method, perform the test protocol on different days. They should use separate equipment if possible and must not share calibration solutions or calibration curves.

8. Conducting the Interlaboratory Test

8.1 *Program Coordinator*—One individual (presumably the task group chairman) will coordinate the entire ILS. A prospective ILS program coordinator will find helpful information on conducting the program in Practice E 691. One way to organize the work to provide close control while moving the program steadily to its conclusion is as follows:

8.1.1 Prepare a draft of the method to be tested.

8.1.2 Recruit a task group of participating laboratories.

8.1.3 Select a set of test materials and assemble them into kits, one for each laboratory.

8.1.4 Write the test protocol to instruct each laboratory how to run the test.

8.1.5 Prepare a report form.

8.1.6 Establish a realistic time schedule for each part of the test program.

8.1.7 Assemble and deliver to each participating laboratory everything needed to run the test: the draft method; the test materials and a document which describes them; the test protocol; the report forms; and a cover letter which includes the deadline for return of results; and the name, address, telephone, and FAX numbers of the person who will handle problems and receive the completed report forms.

8.1.8 Expedite the laboratory testing. Follow up to be sure that the laboratories receive the test materials and understand what is expected of them. Encourage laboratories to complete the work.

8.1.9 Inspect results on each report form as it is received. Resolve omissions and apparent clerical errors at once. Obtain missing values. If obviously erroneous data is submitted, find out the cause, if possible, and help the laboratory eliminate the problem. Encourage the laboratory to submit a replacement set of data, if circumstances permit. (The final decision about replacing data will be made by the task group later, after the testing is complete.)

8.1.10 Perform a preliminary statistical analysis. Summarize the comments from laboratories to explain questionable results. Present this information to the task group.

8.1.11 As approved by the task group, prepare the final statistical evaluation and the research report. Obtain the task group's approval for the completed study.

8.1.12 Modify the scope of the method, if necessary, and prepare the precision and bias statement. Submit the completed method to the technical subcommittee chairman for editorial review, followed by subcommittee ballot.

8.2 *Task Group*—The task group usually consists of one representative from each participating laboratory. The laboratory representative's name, address, and telephone and FAX numbers should be given to the task group chairman when a laboratory agrees to participate.

8.2.1 The laboratory representative shall be fully cognizant of the laboratory's capabilities and be in a position to ensure the following:

8.2.1.1 The laboratory is capable of performing the method properly,

8.2.1.2 Appropriate personnel are assigned to perform the work and the method is followed exactly as written,

8.2.1.3 Test materials are handled properly,

8.2.1.4 The test protocol is complied with in all details,

8.2.1.5 The results are recorded accurately on the report form, and

8.2.1.6 The laboratory adheres to the program time schedule.

8.2.2 As a member of the task group, the laboratory representative must be familiar enough with the analytical techniques used in the method to be able to understand the significance of the test statistics and render considered judgment on how well the method's performance meets the original analytical requirements.

8.3 *Test Protocol*—Preparation of the test protocol is the responsibility of the coordinator. The protocol gives instructions to the participating laboratories such as the following:

8.3.1 *Test Pattern*—Practice E 691 requires estimates of the performance of a method under two extreme conditions of variability, minimum variability, and variability among different laboratories. Minimum variability requires that replicate results be obtained with as little elapsed time as possible. For a material of proven homogeneity, specify Test Plan A: three or more sequential replicate results on one portion of the material (Note 3). Direct each laboratory to analyze test materials in random order, but to complete readings for the replicate results (number specified in the protocol) on one test material before proceeding to another. For a test material of unknown homogeneity, specify Test Plan B (Note 4): sequential duplicate results on at least three portions of the material. Direct each laboratory to obtain the readings for duplicate results on one test portion, followed by the specified number of other portions of the same material before proceeding to another material. Give explicit instructions to the analyst for each test material, especially if the study uses Test Plan A for some materials and Test Plan B for others.

NOTE 3—In some methods, the test portion is completely consumed in obtaining one result. In these cases, select the sequential test portions to minimize variation in composition, if possible. Any variation that does occur will increase the method's minimum standard deviation.

NOTE 4—Test Plan B is effective only when duplicate results can be taken on a relatively homogeneous test portion. Ideal methods for this approach are those in which replicate test portions can be put into solution and duplicate results obtained on each solution. If determinations are made directly on solid specimens, Test Plan B should be attempted only if each laboratory can be provided with at least three portions of the test material and there is reason to expect that duplicate results on each portion will show less variability than results obtained from different portions.

8.3.2 A third test pattern may be used if the task group wishes to measure the within-laboratory standard deviation, s_r, and calculate the repeatability index, r. Obtain sequential duplicate results on a test material of proven homogeneity on each of at least three days. Direct each laboratory to obtain duplicate results on one test portion of a material on the specified number of (not necessarily sequential) days. Several conditions must be explicitly spelled out in the protocol, as follows:

8.3.2.1 For methods in which samples are dissolved, prepare a single test solution each day. For solid specimens, prepare them each day in the manner specified by the method.

8.3.2.2 Each day the method must be performed in its entirety, including instrument setup, preparation of the calibration solutions and calibration, and other steps necessary for each day's work in accordance with the method. If the method includes standardization, it must be performed before each day's work whether or not need for it is indicated.

8.3.2.3 Determine the duplicate results on a single test solution. For solid samples, determine the duplicate results with as little disturbance of the specimen as the method permits.

8.3.3 The test protocol includes other details of the way the laboratory is expected to conduct its part of the ILS (see Note 5).

NOTE 5—The following is an illustrative rather than exhaustive example: (*1*) Specify the number of significant digits with which results are to be recorded for each concentration level. (*2*) Show how to fill out the report forms. (*3*) Emphasize the importance of keeping written observations that might reveal the cause of unexpected results. (*4*) Emphasize the necessity for immediate communication with the coordinator when a problem is encountered. (*5*) Ask for information that might prove useful in the task group's evaluation of the test data, such as a description of test equipment.

8.4 *Report Forms*—Provide official report forms to each laboratory. Data forms should be convenient to fill out and simple to use when transcribing the data for statistical analysis. Provide spaces for the laboratory to identify itself and the date the test was performed.

9. Evaluating Data

9.1 The task group must ensure that data are handled properly both in the laboratory and during statistical analysis. Laboratory representatives should be cautioned against submitting "selected" data. For example, a laboratory might be tempted to take extra readings and submit only those that agree well with each other. Such practices or other deviations from the test protocol must not be tolerated because they destroy the integrity of the test design and make correct interpretation of the test results impossible. No result may be rejected just because it does not look good or exceeds a statistical rejection limit, but only for assignable cause. Assignable cause is evidence that the method was not performed as written or that standard laboratory practice was not followed. This may involve human error or equipment malfunction, or both. In this event, the laboratory should correct the problem and, if possible, rerun the test or the portion of the test affected by it. However, laboratory personnel must not make changes in the method. Problems that are perceived as stemming from the method must be discussed with the coordinator. Any unauthorized deviation from the written method, no matter how trivial it may seem to the analyst, may render the laboratory's results unusable.

9.2 When test data are received from a laboratory, the coordinator immediately reviews it for consistency and adherence to the test protocol.

9.2.1 The coordinator discusses questionable values with the laboratory representative and clarifies the reasons for rerun data (if any). He transfers the original data to test material tables, marking any values that were questioned or warranted a rerun and recording substitute values (if any) as footnotes. The reasons for proposed deletions or substitutions are documented, observations on the method reported by the laboratories are summarized, and a preliminary statistical evaluation to flag inconsistent data by the h and k statistics is performed. The coordinator questions laboratories that submitted flagged data to see if assignable causes can be found.

9.3 When all data have been received and the tables and comments have been assembled, the coordinator presents this information to the task group at a meeting. The task group must decide whether or not the evidence supplied by the contributing laboratory supports rejecting questionable data. When rerun data are presented, it should also consider whether or not the integrity of the test is jeopardized by substitution of the rerun data for the rejected data. If a misunderstanding of the method contributed to a problem, the task group may wish to edit the language of the method (Note 6) to ensure that it will not continue to trouble future users.

NOTE 6—An editorial change to a method, proposed after testing is completed, must be examined carefully to ensure that it does not make or imply a change in the technical substance of the method nor that such a change can be inferred from the edited wording.

9.4 The coordinator performs a final statistical analysis using the data authorized by the task group in the previous step and prepares the research report and the precision and bias section of the method. If the method meets the original project requirements, the task group authorizes its chairman to submit the method to the technical subcommittee chairman for final editorial review and subcommittee ballot. If the task group decides that the method does not meet the requirements, it should examine the test data (with the help of a person who is both adept at using statistics and experienced in analytical chemistry) in order to change the method to improve its performance. Proposed changes to the method should be tested by a small group of laboratories before attempting a full-scale retest. Because such changes affect the technical substance of the method, the revised method must undergo another ILS.

10. Calculation

10.1 The ILS test program measures the variability of the test method in typical laboratories. The between-laboratory standard deviation, s_R, and reproducibility index, R, are calculated for this purpose. If the calculated values of these statistics are to reflect the expected future performance of the method, the test data should not contain extraneous results. The h and k statistics are provided to aid the task group in its search for extraneous data, but the task group is cautioned that statistics alone cannot provide sufficient cause for excluding data. For the relatively small data set produced in a typical ILS using this practice, a result is truly extraneous only if it is caused by errors in chemical manipulations, improper operation of equip-

ment, or failure to follow generally accepted procedures or specific instructions of the method. The task group must use principles of chemistry and physics as well as its analytical experience to show that flagged data are inconsistent with reasonable interpretation and execution of the instructions provided in the method and test protocol. Failing that, the task group must retain the data.

10.2 The equations are arranged for manual calculation of the statistics, but the coordinator is encouraged to use a computer version to save time and avoid errors. A separate statistical analysis is performed for each test material.

10.3 The data for an ILS run according to Test Plan A are shown in Table 1. Each column represents a test material with each laboratory's replicate results in rows.

10.4 *Test Plan A Calculations*—The results of the statistical calculations on the data in Table 1 are displayed in Table 2. (In these equations, *x* represents the replicate results reported by a laboratory, *n* equals the number of replicate results per laboratory, and *p* equals the number of laboratories which provided the data used for this material.)

10.4.1 For each laboratory, calculate the mean, standard deviation, and the square of the standard deviation:

$$\bar{x} = (\Sigma(x)/n$$
$$s = \sqrt{\Sigma(x - \bar{x})^2/(n - 1)}; \qquad \text{and } s^2$$

10.4.2 Calculate the overall mean result for the material:

$$\bar{\bar{x}} = (\Sigma\bar{x})/p$$

TABLE 1 Nickel ILS Data (% Nickel)

Laboratory Number	Test Materials				
	A	B	C	D	E
1	0.0053	0.053	0.122	0.217	1.08
	0.0053	0.052	0.120	0.215	1.07
	0.0054	0.053	0.120	0.215	1.07
2	0.0057	0.052	0.124	0.207	1.07
	0.0077	0.054	0.124	0.204	1.06
	0.0059	0.053	0.119	0.195	1.05
3	0.0060	0.053	0.120	0.221	1.08
	0.0057	0.055	0.113	0.213	1.05
	0.0060	0.053	0.119	0.220	1.07
4	0.0058	0.057	0.121	0.219	1.06
	0.0053	0.056	0.123	0.225	1.08
	0.0065	0.058	0.130	0.230	1.14
5	0.0058	0.054	0.125	0.220	1.06
	0.0050	0.054	0.123	0.220	1.06
	0.0057	0.053	0.126	0.219	1.08
6	0.0060	0.054	0.120	0.215	1.05
	0.0059	0.054	0.115	0.215	1.05
	0.0060	0.054	0.120	0.210	1.05
7	0.0055	0.056	0.120	0.221	1.05
	0.0060	0.057	0.125	0.221	1.07
	0.0050	0.057	0.125	0.215	1.05
8	0.0069	0.058	0.118	0.218	1.07
	0.0069	0.058	0.121	0.216	1.06
	0.0063	0.057	0.118	0.217	1.08
9	0.0066	0.056	0.117	0.213	1.10
	0.0060	0.057	0.130	0.220	1.05
	0.0062	0.054	0.123	0.225	1.05
10	0.0058	0.055	0.122	0.221	1.08
	0.0056	0.053	0.124	0.223	1.06
	0.0055	0.055	0.120	0.220	1.08
11	0.0049	0.055	0.127	0.220	1.03
	0.0043	0.057	0.132	0.216	1.06
	0.0053	0.054	0.125	0.214	1.05

TABLE 2 Statistical Calculations for Nickel Material E (NBS 82a, 1.07 % Nickel)

Laboratory Number	Test Results, x			$\bar{x}$	s	d	s^2	d^2	h	k
	1	2	3							
1	1.08	1.07	1.07	1.0733	0.0058	0.0076	0.00003329	0.00005746	0.59	0.32
2	1.07	1.06	1.05	1.0600	0.0100	−0.0058	0.00010000	0.00003318	−0.45	0.55
3	1.08	1.05	1.07	1.0667	0.0153	0.0009	0.00023348	0.00000083	0.07	0.84
4	1.06	1.08	1.14	1.0933	0.0416	−0.0276	0.00173306	0.00076066	2.16	2.28
5	1.06	1.06	1.08	1.0667	0.0116	0.0009	0.00013340	0.00000083	0.07	0.63
6	1.05	1.05	1.05	1.0500	0.0000	−0.0158	0.00000000	0.00024838	−1.24	0.00
7	1.05	1.07	1.05	1.0567	0.0116	−0.0091	0.00013340	0.00008263	−0.71	0.63
8	1.07	1.06	1.08	1.0700	0.0100	0.0042	0.00010000	0.00001798	0.33	0.55
9	1.10	1.05	1.05	1.0667	0.0289	0.0009	0.00083348	0.00000083	0.07	1.58
10	1.08	1.06	1.08	1.0733	0.0116	0.0076	0.00013340	0.00005625	0.59	0.63
11	1.03	1.06	1.05	1.0467	0.0153	−0.0191	0.00023348	0.00036443	−1.50	0.84
				$x = 1.0658$			$\Sigma(s^2) = 0.00366699$			

$n = 3, p = 11$

$\Sigma(d^2) = 0.00162346$

$s_{\bar{x}} = \sqrt{0.00162346/10} = 0.01274; s_M = \sqrt{0.00366699/11} = 0.01826;$

$s_t = \sqrt{0.000162346 + (0.000333363)(2/3)} = 0.01961; s_R = 0.01961;$

$R = (2.8)(0.01961) = 0.0594; R_{rel} = (100)(0.0594)/1.0658 = 5.15 \%.$

ILS Statistics Summary:

Material Mean Concentration: $x = 1.066$

Minimum Standard Deviation of the Method: $s_M = 0.0183$

Reproducibility Standard Deviation: $s_R = 0.0196$

Reproducibility Index: $R = 0.0549; R_{rel} = 5.15 \%$

10.4.3 For each laboratory, calculate its laboratory difference and the square of the difference:

$$d = \bar{x} - \bar{\bar{x}}; \quad \text{and} \quad d^2$$

10.4.4 Calculate the standard deviation of laboratory differences:

$$s_{\bar{x}} = \sqrt{\Sigma(d^2)/(p-1)}$$

10.4.5 Calculate the method's minimum standard deviation:

$$s_M = \sqrt{\Sigma(s^2)/p}$$

10.4.6 Calculate a trial value for the reproducibility standard deviation:

$$s_t = \sqrt{(s_{\bar{x}})^2 + [(s_M)^2(n-1)/n]}$$

10.4.7 Select the final value for the reproducibility standard deviation:

$$s_R = \text{the larger of } s_t \text{ or } s_M$$

10.4.8 Calculate the reproducibility index and percent relative reproducibility index:

$$R = 2.8(s_R); \quad \text{and} \quad R_{rel} = 100R/\bar{\bar{x}}$$

10.4.9 For each laboratory, calculate its between-laboratory consistency statistic:

$$h = d/s_{\bar{x}}$$

10.4.10 For each laboratory, calculate its within-laboratory consistency statistic:

$$k = s/s_M$$

10.5 *Test Plan B Calculations*—Data for a single material obtained in accordance with Test Plan B is shown in Table 3.

TABLE 3 Iron Material 1A Data, ppm Iron

Laboratory Number	Replicate	Test Results		Replicate Mean, X	D[A]	D^2
		x_1	x_2			
1	1	348	345	346.5	3	9
	2	343	339	341.0	4	16
	3	332	327	329.5	5	25
2	1	347	356	351.5	9	81
	2	333	340	336.5	7	49
	3	363	357	360.0	6	36
3	1	325	317	321.0	8	64
	2	313	310	311.5	3	9
	3	330	320	325.0	10	100
4	1	326	322	324.0	4	16
	2	322	329	325.5	7	49
	3	325	337	331.0	12	144
5	1	338	336	337.0	2	4
	2	335	331	333.0	4	16
	3	325	343	334.0	18	324
6	1	339	335	337.0	4	16
	2	333	335	334.0	2	4
	3	338	340	339.0	2	4
7	1	356	346	351.0	10	100
	2	336	331	333.5	5	25
	3	343	346	344.5	3	9
$n = 3, p = 7$						$\Sigma(D^2) = 1100$

$s_M = \sqrt{1100/(2)(3)(7)} = 5.118$

[A] A The difference between duplicate test results is D.

It is arranged like Table 1, except that space is provided for duplicate results on each replicate portion analyzed by a laboratory. Other test materials in the iron method test are not shown. The results of the statistical calculations start in the last two columns of Table 3 and continue in Table 4. For a test including data for day-to-day within-laboratory variability (replicates analyzed in duplicate on different days in the same laboratory), proceed in accordance with 10.6. For a test including data for material variability (replicates are separate portions analyzed on the one day), proceed in accordance with 10.7.

NOTE 7—In the following equations, x_1 and x_2 represent the duplicate

TABLE 4 Statistical Calculations for Iron Material 1A

Laboratory Number	Replicate Means, X			Laboratory mean, $\bar{x}$	s	d	s^2	d^2	h	k
	1	2	3							
1	364.5	341.0	329.5	339.00	8.675	3.476	75.255625	12.082576	0.35	1.20
2	351.5	336.5	360.0	349.33	11.899	13.810	141.586201	190.716100	1.38	1.64
3	321.0	311.5	325.0	319.17	6.934	−16.357	48.080356	267.551449	−1.63	0.96
4	324.0	325.5	331.0	326.83	3.686	−8.690	13.586596	75.516100	−0.87	0.51
5	337.0	333.0	334.0	334.67	2.082	−0.857	4.334724	0.734449	−0.09	0.29
6	337.0	334.0	339.0	336.67	2.517	1.143	6.335289	1.306449	0.11	0.35
7	351.0	335.5	344.5	343.00	8.846	7.476	78.251716	55.890576	0.75	1.22
	$n = 3, p = 7$			$\bar{x}/Av = 335.5238$			$\Sigma(s^2) = 367.430507$	$\Sigma(d^2) = 603.797699$		

$s_M^2 = 26.190476$ (from Table 3); $s_X^2 = 367.430507/7 = 52.490072$; $s_{\bar{x}}^2 = 603.797699/6 = 100.632950$;
$s_M = 5.118$; Proceed to either (1) or (2) (but not both), depending on the provisions of the test protocol:

(1) Statistics for Day-to-Day ILS:

$$s_r^2 = s_X^2 + \frac{1}{2}s_M^2 = 52.490072 + 26.190476/2 = 65.58531$$
$$s_r = 8.098$$
$$s_R^2 = s_{\bar{x}}^2 + \frac{n-1}{n}s_X^2 + \frac{1}{2}s_M^2$$
$$= 100.632950 + \frac{2}{3}52.490072 + \frac{1}{2}26.190476$$
$$= 148.721569$$
$$S_R = 12.195$$
$r = 2.8 \times 8.098 = 22.67$; $R = 2.8 \times 12.195 = 34.15$
$R_{rel} = 100 \times 34.15/335.52 = 10.18\%$

(2) Statistics for ILS to Eliminate Material Variability Effect:

$$s_H^2 = s_X^2 - \frac{1}{2}s_M^2 = 52.490071 - \frac{1}{2}26.190476 = 39.394834$$
$$s_t^2 = s_{\bar{x}}^2 - \frac{1}{n}s_X^2 + \frac{1}{2}s_M^2$$
$$= 100.632950 - \frac{1}{3}52.490072 + \frac{1}{2}26.190476 = 96.231497$$
$s_R = \sqrt{s_t^2} = 9.810$; $R = 2.8 \times 9.810 = 27.47$
$R_{rel} = 100 \times 27.47/335.52 = 8.19\%$
$$F_H = \frac{s_M^2 + 2s_H^2}{s_M^2} = (26.190476 + 2 \times 39.394834)/26.190476$$
$$= 4.01, \text{ with}$$
$f_1 = 2 \times 7 = 14$ and $f_2 = 3 \times 7 = 21$ degrees of freedom

---

results from one replicate in one laboratory, X represents their mean, n equals the number of replicates per laboratory, and p equals the number of laboratories providing data used in the calculations for one material.

10.6 *Test Plan B—Day-to-Day Variability (see* Note 7)—The replicates are portions of the test material that are analyzed in duplicate on each of several days in each laboratory (see 8.3.2).

10.6.1 For each test portion, calculate the mean of the duplicate results, their difference, and the square of the difference:

$$X = (x_1 + x_2)/2$$
$$D = x_1 - x_2; \text{ and } D^2$$

10.6.2 Calculate the method's minimum standard deviation:

$$s_M = \sqrt{\Sigma D^2/2pn}$$

10.6.3 For each laboratory, calculate the laboratory mean, the standard deviation of the replicate means, and the square of the standard deviation:

$$\bar{x} = \Sigma X/n$$
$$s = \sqrt{\Sigma(X - \bar{x})^2/(n - 1)}; \text{ and } s^2$$

10.6.4 Calculate the overall mean result for the material:

$$\bar{\bar{x}} = \Sigma\bar{x}/p$$

10.6.5 For each laboratory, calculate its laboratory difference and the square of the difference:

$$d = \bar{x} - \bar{\bar{x}}; \text{ and } d^2$$

10.6.6 Calculate the pooled standard deviation of the replicate means and its square:

$$s_x = \sqrt{\Sigma s^2/p}; \text{ and } s_x^2$$

10.6.7 Calculate the standard deviation of the laboratory means and its square:

$$s_{\bar{x}} = \sqrt{\Sigma d^2/(p - 1)}; \text{ and } s_{\bar{x}}^2$$

10.6.8 Calculate the repeatability standard deviation:

$$s_{t1} = \sqrt{s_x^2 + \frac{1}{2}s_M^2}$$

$s_r = $ the larger of s_{t1} or s_M

10.6.9 Calculate the reproducibility standard deviation:

$$s_{t2} = \sqrt{s_{\bar{x}}^2 + \left(\frac{n-1}{n}\right) s_x^2 + \frac{1}{2} s_M^2}$$

$$s_R = \text{the larger of } s_{t2} \text{ or } s_r$$

10.6.10 Calculate the repeatability index, the reproducibility index and percent relative reproducibility index:

$$r = 2.8(s_r); R = 2.8(s_R); \quad \text{and} \quad R_{\text{rel}} = 100R/\bar{\bar{x}}$$

10.6.11 For each laboratory, calculate its between-laboratory consistency statistic:

$$h = d/s_{\bar{x}}$$

10.6.12 For each laboratory, calculate its within-laboratory consistency statistic:

$$k = s/s_X$$

10.7 *Test Plan B—Material Variability (see* Note 7)— Separate replicate portions of a test material are analyzed in duplicate on one day in each laboratory (see 8.3.1)

10.7.1 For each replicate, calculate the mean of the duplicate results, their difference, and the square of the difference:

$$X = (x_1 + x_2)/2$$

$$D = x_1 - x_2; \quad \text{and} \quad D^2$$

10.7.2 Calculate the method's minimum standard deviation:

$$s_M = \sqrt{\Sigma D^2/2np}$$

10.7.3 For each laboratory, calculate the laboratory mean, the standard deviation of the replicate means, and the square of the standard deviation:

$$\bar{x} = (\Sigma X/n)$$

$$s = \sqrt{\Sigma(X - \bar{x})^2/(n-1)}; \quad \text{and} \quad s^2$$

10.7.4 Calculate the overall mean result for the material:

$$\bar{\bar{x}} = \Sigma\bar{x}/p$$

10.7.5 For each laboratory, calculate its laboratory difference and the square of the difference:

$$d = \bar{x} - \bar{\bar{x}}; \quad \text{and} \quad d^2$$

10.7.6 Calculate the pooled standard deviation of the replicate means and its square:

$$s_x = \sqrt{\Sigma s^2/p}; \quad \text{and} \quad s_x^2$$

10.7.7 Calculate the standard deviation of the laboratory means and its square:

$$s_{\bar{x}}\sqrt{\Sigma d^2/(p-1)}; \quad \text{and} \quad s_{\bar{x}}^2$$

10.7.8 Calculate the variance of the material homogeneity effect:

If $\left(s_X^2 - \frac{1}{2}s_M^2\right)$ is negative or zero,

set $s_H^2 = 0$; otherwise,

$$s_H^2 = s_X^2 - \frac{1}{2}s_M^2$$

10.7.9 Calculate the reproducibility standard deviation:

$$s_{t3} = \sqrt{s_{\bar{x}}^2 - \frac{1}{n}s_X^2 + s_M^2}$$

$$s_R = \text{the larger of } s_{t3} \text{ or } s_M$$

10.7.10 Calculate the reproducibility index and percent relative reproducibility index:

$$R = 2.8(s_R); \quad R_{\text{rel}} = 100R/\bar{\bar{x}} =$$

10.7.11 For each laboratory, calculate its between-laboratory consistency statistic:

$$h = d/s_{\bar{x}}$$

10.7.12 For each laboratory, calculate its within-laboratory consistency statistic:

$$k = s/s_X$$

10.7.13 *Optional (see Note 8)*—Calculate the material homogeneity *F*-statistic and its numerator and denominator degrees of freedom:

$$F_H = (s_M^2 + 2s_H^2)/s_M^2$$

$$f_1 = p(n-1)$$

$$f_2 = pn$$

NOTE 8—Those adept at statistics may wish to calculate the homogeneity *F*-statistic to test the hypothesis that the test material is homogeneous.

11. Using Statistics in Task Group Decisions

11.1 *Preliminary Screening of Test Data for Consistency*— Most outright mistakes (of the types where equipment fails during the test, a wrong reagent is used, or a test solution is spilled) are caught immediately in the laboratory and are corrected before the test data are submitted. In the same category are misunderstandings about the calculations, transcription errors, and so forth, which often produce such gross distortion of the data that the coordinator can see them at a glance and ask for immediate clarification from the laboratory. Other errors may produce more subtle changes. The pattern of the affected results may not be obvious within the random variation of the rest of the test data. The *h* and *k* statistics help the task group locate such data in its search for assignable causes.

11.2 *h and k Tables*—Place the *h* and *k* statistics in tables, arranged by test material (columns) and laboratory (rows) as in Tables 5 and 6. Some trends are more easily recognized if the materials are arranged by increasing concentration from the first to last column. Consult Table 7 to find the critical value (CV) for each statistic: CV depends upon the number of laboratories actually contributing data to the statistics in the

TABLE 5 Nickel—*h* Statistic[A]

Note—Between-laboratory consistency statistic.

Laboratory Number	Test Material				
	A	B	C	D	E
1	−0.90	−1.31	−0.47	−0.22	0.59
2	1.17	−1.11	0.06	x−2.58x[A,B]	−0.45
3	0.17	−0.72	−1.53	0.18	0.07
4	0.10	1.25	0.80	1.33	2.16[A]
5	−0.59	−0.72	0.80	0.47	0.07
6	0.29	−0.52	−1.21	−0.63	−1.24
7	−0.59	1.05	0.37	0.35	−0.71
8	1.67	1.64	−1.00	0.01	0.33
9	0.85	0.46	0.37	0.41	0.07
10	−0.34	−0.32	−0.05	0.75	0.59
11	−1.84	0.27	1.85	−0.05	−1.50
CV	±2.34	±2.34	±2.34	±2.34	±2.34

[A] Underlined values exceed approximately 87 % of CV.
[B] Values flagged with x___x exceed CV.

TABLE 6 Nickel—*k* Statistic[A]

Note—Within-laboratory consistency statistic.

Laboratory Number	Test Material				
	A	B	C	D	E
1	0.12	0.59	0.34	0.30	0.32
2	x2.29x[AB]	1.02	0.85	1.64	0.55
3	0.36	1.17	1.11	1.15	0.84
4	1.25	1.02	1.39	1.45	x2.28x[A,B]
5	0.91	0.59	0.45	0.15	0.63
6	0.12	0	0.85	0.76	0
7	1.04	0.59	0.85	0.91	0.63
8	0.72	0.59	0.51	0.26	0.55
9	0.64	1.55	1.91	1.58	1.58
10	0.32	1.17	0.59	0.40	0.63
11	1.05	1.55	1.06	0.80	0.84
CV	2.13	2.13	2.13	2.13	2.13

[A] Underlined values exceed approximately 87 % of CV.
[B] Values flagged with x___x exceed CV.

TABLE 7 Critical Values of *h* and *k* at the 0.5 % Significance Level

Critical Value of h	p[A]	Critical Values of k Number of Replicates, n								
		2	3	4	5	6	7	8	9	10
1.15	3	1.72	1.67	1.61	1.56	1.52	1.49	1.47	1.44	1.42
1.49	4	1.95	1.82	1.73	1.66	1.60	1.56	1.53	1.50	1.47
1.74	5	2.11	1.92	1.79	1.71	1.65	1.60	1.56	1.53	1.50
1.92	6	2.22	1.98	1.84	1.75	1.68	1.63	1.59	1.55	1.52
2.05	7	2.30	2.03	1.87	1.77	1.70	1.65	1.60	1.57	1.54
2.15	8	2.36	2.06	1.90	1.79	1.72	1.66	1.62	1.58	1.55
2.23	9	2.41	2.09	1.92	1.81	1.73	1.67	1.62	1.59	1.56
2.29	10	2.45	2.11	1.93	1.82	1.74	1.68	1.63	1.59	1.56
2.34	11	2.49	2.13	1.94	1.83	1.75	1.69	1.64	1.60	1.57
2.38	12	2.51	2.14	1.96	1.84	1.76	1.69	1.64	1.60	1.57
2.41	13	2.54	2.15	1.96	1.84	1.76	1.70	1.65	1.61	1.58
2.44	14	2.56	2.16	1.97	1.85	1.77	1.70	1.65	1.61	1.58
2.47	15	2.57	2.17	1.98	1.86	1.77	1.71	1.66	1.62	1.58
2.49	16	2.59	2.18	1.98	1.86	1.77	1.71	1.66	1.62	1.58
2.51	17	2.60	2.19	1.99	1.86	1.78	1.71	1.66	1.62	1.59
2.53	18	2.61	2.20	1.99	1.87	1.78	1.72	1.66	1.62	1.59
2.54	19	2.62	2.20	2.00	1.87	1.78	1.72	1.67	1.62	1.59
2.56	20	2.63	2.21	2.00	1.87	1.79	1.72	1.67	1.63	1.59
2.57	21	2.64	2.21	2.00	1.88	1.79	1.72	1.67	1.63	1.59
2.58	22	2.65	2.21	2.01	1.88	1.79	1.72	1.67	1.63	1.59
2.59	23	2.66	2.22	2.01	1.88	1.79	1.72	1.67	1.63	1.59
2.60	24	2.66	2.22	2.01	1.88	1.79	1.73	1.67	1.63	1.60
2.61	25	2.67	2.23	2.01	1.88	1.79	1.73	1.67	1.63	1.60
2.62	26	2.67	2.23	2.02	1.89	1.80	1.73	1.68	1.63	1.60
2.62	27	2.68	2.23	2.02	1.89	1.80	1.73	1.68	1.63	1.60
2.63	28	2.68	2.23	2.02	1.89	1.80	1.73	1.68	1.63	1.60
2.64	29	2.69	2.24	2.02	1.89	1.80	1.73	1.68	1.64	1.60
2.64	30	2.69	2.24	2.02	1.89	1.80	1.73	1.68	1.64	1.60

[A] p = number of laboratories.

11.2.1.3 A laboratory's *h*-values exhibit a preponderance of one sign at low concentrations but the opposite at high concentrations, or a consistent trend to larger or smaller values as the concentration increases. That laboratory may have a problem with the slope of its calibration curve.

11.2.2 *k Statistic*—The *k*-statistic is a measure of the variability of a laboratory's replicate results compared with the common variability of all other laboratories for a given material. If all laboratories have similar variability, the *k*-values are randomly small and large, but none should exceed CV. The task group should investigate if any of these conditions exist:

11.2.2.1 An individual *k*-value is flagged as larger than CV. One or more of the replicate results reported by that laboratory on that material may have been incorrectly transcribed or perhaps were influenced by a condition in the laboratory environment that did not affect the other results.

11.2.2.2 A laboratory has several *k*-values flagged, especially if others approach CV. Some condition of that laboratory's environment (which includes instruments and personnel) may not have been as well controlled as in other laboratories.

11.2.2.3 A laboratory exhibits only unusually small *k*-values, especially if many are zero. The laboratory may have an instrument that is insensitive in its response, an insensitive range of readings may have been used, or the analyst may have rounded readings to produce results with artificially small variability.

11.3 *Interpretation of Statistical Values*—When the consistency statistics exceed their critical values, it merely suggests that a problem might exist. The task group, with the help of the appropriate laboratory personnel, has the responsibility of

column. CV for *k* also depends upon the number of replicates reported by each laboratory. Label the line following the last laboratory, "CV," and enter the appropriate value at the bottom of each column. In Table 5, eleven laboratories provided data for each material. CV for *h*, found in Table 7 on the line for *p* = 11, is 2.34. In Table 6, the eleven laboratories each reported three results. CV for *k*, found in Table 7 on the line for *p* = 11 and in the column for *n* = 3, is 2.13. Mark for subsequent investigation each column entry that equals or exceeds the CV of that column.

11.2.1 *h Statistic*—The *h* statistic is a measure of how close the laboratory's mean is to the grand mean of all laboratories for a given material. If the laboratory's mean is higher, *h* is positive; if it is lower, *h* is negative. Each laboratory should have approximately equal numbers of positive and negative values, and none should be larger in absolute value than the CV. The task group should investigate if any of the these conditions exist:

11.2.1.1 An individual *h*-value is flagged as larger than CV. Something may have happened to affect the mean result for that material in that laboratory.

11.2.1.2 A laboratory's *h*-values are the same sign for most materials. That laboratory may have a problem that caused a bias. It is of more concern if one or more materials exceed CV.

determining if a specific problem was likely to have occurred and, if it did, whether to replace the defective data (if substitute values can be obtained), discard it, or retain it. Tables 5 and 6 display the *h* and *k* statistics for the nickel data shown in Table 1. These data were collected long ago and it is now impossible to follow up on the questionable results. For purposes of this discussion, we are assuming a scenario to illustrate how a task group might handle them.

11.3.1 The coordinator noted the following items in Tables 5 and 6:

Item 1—Material D, Laboratory 2: h = −2.58 exceeds CV.

Item 2—Material E, Laboratory 4: h = 2.16 nearly exceeds CV and k = 2.28 exceeds CV.

Item 3—Material A, Laboratory 2: k = 2.29 exceeds CV.

Item 4—Material C, Laboratory 9: k = 1.91 nearly exceeds CV.

Item 5—Laboratories 1, 5, 6, and 8 all had a preponderance of small k-values.

11.3.2 The coordinator contacted representatives of Laboratories 2, 4, and 9 to determine if causes could be found for each suspected problem. The information was evaluated and presented as a report to the task group:

11.3.2.1 Laboratory 2 found that the second reading on Test Material A was actually 0.0057 rather than the 0.0077 reported (miscopied from the notebook). The analyst performing the test had noticed that the Test Material D solution had "bumped a bit" on the hot plate, but, because he believed the coverglass had retained the sample, the results were reported without comment.

11.3.2.2 Laboratories 4 and 9 could find no reason to question the data they submitted. When asked about the apparently high value of 1.14 reported on Test Material E, the representative from Laboratory 4 said that it was not unusual to find one such disagreement among so many replicates. The analyst from Laboratory 9 noticed no problems during the test, believing that 0.117, 0.130, and 0.123 represented reasonable agreement for Test Material C.

11.3.3 The following actions were recommended to the task group:

11.3.3.1 Eliminate the data for Test Material D from Laboratory 2. The analyst had not followed good analytical practice by losing the sample. Laboratory 2 was unable to provide a replacement data set.

11.3.3.2 Retain the data for Test Material E from Laboratory 4 because no cause could be found for the high result of 1.14. The coordinator agreed with the laboratory representative that the result could have been caused by random variation.

11.3.3.3 Substitute the correct value 0.0057 for the erroneous value 0.0077 for Test Material A from Laboratory 2.

11.3.3.4 Retain the data for Test Material C from Laboratory 9 because the agreement did appear to be reasonable in the absence of an observed problem.

11.3.3.5 The data reported by some laboratories seemed to be unusually precise. The task group had the option of rerunning the entire test, but the coordinator recommended accepting the results because the reproducibility was not likely to be affected.

11.3.4 Tables 8 and 9 display *h* and *k* statistics for the revised data. The task group accepted the revised data after a discussion on how to obtain more typical results in future ILS programs (Note 9). Although the *h* and *k* statistics still suggest some laboratory bias or calibration slope effects, the task group could find no reason to believe that the laboratories had failed to use accepted laboratory practices or had failed to carry out the method as written (with the exceptions already addressed). The revised test data were used to calculate the test statistics summarized in Table 10. The task group considered the reproducibility standard deviation and index at each concentration level. While these statistics did not quite achieve the precision hoped for at the inception of the program, the task group felt that the test method would meet the practical needs of the industry and approved the test method as ready for subcommittee ballot.

NOTE 9—Prospective coordinators will recognize that a discussion designed to improve an ILS will be most effective if it precedes the laboratory testing phase. For an ILS, the reported results must include what the analyst might consider "extra" digits. These are the part of the result that provides the variability information and must be reported. While participants should be urged to take care to obtain reliable values, they must be discouraged from deliberately gathering data in a more precise or accurate way during an interlaboratory test than they would use in normal activities. The prudent ILS coordinator, in his protocol and in his pretest discussions with the task group, will stress these points.

11.4 *Plan B Test*—For an ILS conducted in accordance with one of the Test Plan B protocols, the data and statistical calculations follow the patterns and equations of the example shown in Tables 3 and 4. The task group makes a decision before the laboratory phase of the ILS begins to select Test Plan B to test either day-to-day repeatability or, if the test method is amenable to this option (Note 4), to test reproducibility free from the effects of suspected test material inhomogeneity. This practice will not allow a task group to estimate both kinds of statistics in a single ILS. Although the same pattern of results is obtained from both test protocols, a repeatability-oriented ILS gives meaningless statistics if analyzed in accordance with the equations for eliminating heterogeneity effects, while data obtained for the purpose of eliminating the adverse effects of inhomogeneity will not correctly estimate repeatability. (The data set in Table 3 is analyzed both

TABLE 8 Nickel—*h* Statistic

NOTE—Between-laboratory consistency statistic.

Laboratory Number	Test Material				
	A	B	C	D	E
1	−0.85	−1.31	−0.47	−0.89	0.59
2	0.03[A]	−1.11	0.06	...[A]	−0.45
3	0.30	−0.72	−1.53	−0.15	0.07
4	0.23	1.25	0.80	1.97	2.16[B]
5	−0.51	−0.72	0.80	0.38	0.07
6	0.44	−0.52	−1.21	−1.63	−1.24
7	−0.51	1.05	0.37	0.17	−0.71
8	1.93	1.64	−1.00	−0.47	0.33
9	1.05	0.46	0.37	0.28	0.07
10	−0.24	−0.32	−0.05	0.91	0.59
11	−1.87	0.27	1.85	−0.57	−1.50
CV	±2.34	±2.34	±2.34	±2.29	±2.34

[A] Underlined values exceed approximately 87 % of CV.
[B] Data revised or deleted.

TABLE 9 Nickel—*k* Statistic

NOTE—Within-laboratory consistency statistic.

Laboratory Number	Test Material				
	A	B	C	D	E
1	0.17	0.59	0.34	0.33	0.32
2	0.33[A]	1.02	0.85	...[A]	0.55
3	0.50	1.17	1.11	1.26	0.84
4	1.72	1.02	1.39	1.59	x2.28x[B,C]
5	1.25	0.59	0.45	0.17	0.63
6	0.17	0	0.85	0.83	0
7	1.43	0.59	0.85	1.00	0.63
8	0.99	0.59	0.51	0.29	0.55
9	0.87	1.55	1.91[B]	1.74	1.58
10	0.44	1.17	0.59	0.44	0.63
11	1.44	1.55	1.06	0.88	0.84
CV	2.13	2.13	2.13	3.11	2.13

[A] Data revised or deleted.
[B] Underlined values exceed approximately 87 % of CV.
[C] Values flagged with x___ x exceed CV.

TABLE 10 Nickel—Statistical Summary

Test Material	Number of Laboratories	Mean, x	s_M	s_R	R	R_{rel}, %
A	11	0.00575	0.000349	0.000567	0.0016	27.6
B	11	0.0549	0.000985	0.00188	0.0053	9.6
C	11	0.122	0.00341	0.00421	0.0118	9.6
D	10	0.219	0.00347	0.00423	0.0118	5.4
E	11	1.066	0.0183	0.0196	0.0549	5.2

ways in order to emphasize the difference between the calculations appropriate for each experimental design.)

11.4.1 *Interpretation of Day-to-Day Statistics*—For this type of ILS, the protocol specifies duplicate results from a test material on each of three or more days in each laboratory. If the method specifies a calibration each time the method is used, a complete calibration shall be performed each day. If the method specifies standardization, it must be performed each day. If the method specifies standardization, it must be performed each day without exception before the test results are obtained. Under these conditions, a Plan B Test protocol will produce data that is likely to include the most important sources of within-laboratory day-to-day variability. This test design estimates a repeatability standard deviation (day-to-day within-laboratory) as well as the minimum standard deviation of the method and reproducibility standard deviation obtained in the Test Plan A design. Follow the interpretative procedures outlined in 11.1-11.3.

11.4.2 *Interpretation of Statistics to Exclude Material Variability*—For this type of ILS, the protocol specifies that duplicate results be obtained from each of three or more replicate portions of a test material in an uninterrupted analytical session in each laboratory. The task group should expect that variability between duplicates will be less than variation between replicate material portions; for example, if the duplicates are aliquots from a test sample solution, they will exhibit nearly perfect homogeneity in comparison with separate solutions prepared from replicate sample portions. Follow the interpretive procedures outlined in 11.1-11.3. The homogeneity effect statistics, s_H and F_H, relate only to the test material, not the method, and need not concern the task group.

12. Preparation of Research Report, Precision and Bias Statement, and Adjustment of the Method's Scope Limits

12.1 *Research Report*—The research report provides a permanent record of the data of the task group that is kept on file at ASTM Headquarters for future reference. The following should be considered the minimum contents of a research report:

12.1.1 The full title of the method;

12.1.2 The names and affiliations of the ILS coordinator and the representatives of the participating laboratories;

12.1.3 The test materials, their identification code and material type, source from which obtained, and the critical concentration values (if an accepted reference material).

12.1.4 The test pattern (from the test protocol), that is, how laboratories handled each portion of the test materials to obtain the results reported in the data tables.

12.1.5 The table of test data as reported by the participating laboratories. Include in the body of the table any substituted or corrected values. Use ellipses for rejected data. Footnote each such entry with a brief description of the action taken and the reason for the action.

12.1.6 Include a table of the ILS test statistics to be used in the test method's Precision and Bias section. For example, the statistics shown in Table 10 may suffice. Include other method parameters, such as the upper and lower concentration limits for the test method scope. If calculations are made in accordance with a standard practice, it is only necessary to identify which practice was followed. If other statistical relationships are used, these should be explained in detail.

12.1.7 Include the research report when the test method is submitted to the technical subcommittee chairman for editorial review and subcommittee ballot.

12.2 *Precision and Bias*—For methods that use this practice, the mandatory Precision and Bias section will contain the information shown in the example in Table 11. Other information may be included as appropriate.

12.2.1 *Precision*—Use the following format:

TABLE 11 Statistical Information—Nickel

Test Material	Number of Laboratories	Nickel Found,%	Min SD (s_M, E 1601)	Reproducibility SD (s_R, E 1601)	Reproducibility Index (R, E 1601)	R_{rel}, %
A	11	0.0058	0.00035	0.00057	0.0016	27.6
B	11	0.0549	0.00098	0.00188	0.0053	9.6
C	11	0.122	0.0034	0.0042	0.012	9.6
D	10	0.219	0.0035	0.0042	0.012	5.4
E	11	1.066	0.0183	0.0196	0.055	5.2

	Certified Nickel, %	Number	Source	Description
A	0.005	SRM 10g	NBS	carbon steel
B	0.056	SRM 152a	NBS	carbon steel
C	0.120	SRM 7g	NBS	cast iron, high phosphorus
D	0.217	SRM 106b	NBS	Nitralloy G
E	1.07	SRM 82a	NBS	cast iron

Eleven laboratories cooperated in testing this method and obtained the precision information summarized in Table 11. Supporting data have been filed at ASTM Headquarters. Request RR:E01-XXXX [where XXXX is the Research Report number assigned by ASTM for this set of data].

12.2.2 *Bias*:

12.2.2.1 If certified reference materials have been tested, use this format:

The accuracy of this method has been deemed satisfactory based upon the bias data in Table 11. Users are encouraged to use these or similar reference materials to verity that the method, is performing accurately in their laboratories.

12.2.2.2 If certified reference materials have not been tested, use this format:

No information on the accuracy of this method is known, because at the time it was tested, no accepted reference materials were available. Users are encouraged to employ suitable reference materials, if available, to verify the accuracy of the method in their laboratories.

12.3 *Method's Scope Limits*:

12.3.1 *Lower Limit (L)*—The lower limit is the concentration in a material below which a method may not be used to report quantitative values. If the method is to be used near the lower end of its effective concentration range, calculate *L*:

$$L = 100R/e_{max}$$

where:

R = reproducibility index of the lowest test material, and

e_{max} = maximum acceptable percent relative error (Note 10).

Experience has shown, for the 95 % confidence level at which R is calculated, that a value of 50 % for e_{max} yields results useful for determining residual levels of trace elements (Note 11). For such methods, the calculation reduces to $L = 2R$. (For the nickel example, $L = 2(0.0016) = 0.003$ %.)

NOTE 10—It is important that at least one of the test materials in the ILS be near or below the lowest concentration level sought. At these low concentrations there is no generally valid relationship for extrapolating standard deviations to lower concentrations, so this practice takes the conservative approach of calculating the lower scope limit from the standard deviation of the lowest test material(s). If the concentration of the lowest test material is considerably higher than the level of interest, the calculated lower limit will probably be higher than it would be if estimated from a test material of optimum concentration. Although unfortunate, this error is preferable to claiming an unsubstantiated extrapolated value.

NOTE 11—Under no circumstance may e_{max} be larger than 50 %. Use smaller values of e_{max} for applications requiring greater precision.

12.3.2 *Upper Limit (U)*—The upper limit is the concentration in a material above which use of the method is not recommended. Set the upper limit to a value that the task group believes is warranted by the ILS test results. A reasonable extrapolation above the highest test material concentration is sometimes permissible, although the task group is cautioned not to extend method limits to concentrations with which no one has actual experience.

12.3.3 In the scope of the test method, set the lower end of the method's concentration range to any desired value equal to or greater than *L*. Set the upper end of the method's concentration range equal to or less than *U*.

13. Keywords

13.1 bias; interlaboratory test; precision; statistics

ANNEXES

(Mandatory Information)

A1. BACKGROUND INFORMATION

A1.1 The statistical basis for this practice can be found in Practice E 691.

A1.2 *Test Plan A*—This basic ILS design assumes (in addition to the other assumptions common to all analysis of variance) that the test material is homogeneous in composition or, if the composition does vary, that it is satisfactory to include that variability in the estimate of the error SD (method's minimum SD). Test Plan A follows the test protocol and statistical analysis recommended in Practice E 691. The user of this practice should look there for the theoretical justification of the basic aspects of this practice.

A1.3 *Test Plan B*—Task groups developing methods of chemical analysis have encountered two situations not covered by Test Plan A. They may wish to estimate a standard deviation relating to results obtained in the same laboratory on separate occasions, or they may need to include in the ILS test materials

they suspect are less homogeneous than the majority of other materials in the study. Both versions are represented in Table A1.1.

A1.3.1 *Repeatability*—In the chemical analysis laboratory, the term "repeatability" has traditionally been associated with very long-term variability within a laboratory. A good approximation of this long term test can be obtained if the participating laboratories perform duplicate determinations under conditions of minimum variability on three or more days, repeating each day all aspects of the method most affecting the precision and accuracy of the results. Consequently, this type of ILS is quite expensive, requiring nearly three times the effort in each laboratory as a Plan A experiment. The repeatability index, *r*, predicts the range between two results obtained on the same material on any two days in the same laboratory. The highest level represents the variability due to the *p* laboratories, the next lower level the variability due to the *n* replicates (days)

within each laboratory, and the lowest level is the variability of duplicate results (variance of the minimum standard deviation, s_M,) nested within laboratories and days. The repeatability standard deviation, s_r, is the square root of the sum of the replicate and error variances, while the reproducibility standard deviation, S_R, is the square root of the sum of the variances of all three sources.

A1.3.2 *Imperfectly Homogeneous Test Materials*—If the task group conducting the ILS is not assured of the homogeneity of a test material and does not want to include that material's variability in the method's statistics, the third variable of Test Plan B may be used for the material homogeneity effect and the statistical calculations modified to eliminate that source from their estimate of the method's reproducibility. This kind of experiment is possible only for methods in which each laboratory can perform the duplicate determina-

tions on the replicates under conditions of minimum variability. The test is performed only one day in each laboratory and the additional work of the extra determination per replicate is minimal. Task groups may find this alternative ILS test design useful and not costly. Each laboratory reports duplicate results from each of at least three replicates under conditions of minimum variability (for example, from aliquot portions of dissolved replicate samples). The highest level represents the variability due to the p laboratories, the next lower level represents the variability due to the n replicates within laboratories, and the lowest level is the variability of duplicate results (variance of the minimum standard deviation, s_M,) nested within labortories and replicates. The reproducibility standard deviation, s_R, is calculated as the square root of the sum of the laboratory and error variances (omitting the contribution of the material's inhomogeneity).

A2. STATISTICAL THEORY

A2.1 *Model*—As with Test Plan A, Test Plan B provides data that may be analyzed in accordance with a completely randomized model. Level 1 corresponds to the effect of laboratories, an effect that sums to zero over all laboratories and exhibits a variability measured by σ_{lab}. Level 2 corresponds to the effect of replication within each laboratory, an effect that sums to zero over all test portions and exhibits a variability measured by σ_{repl}. Level 3 corresponds to the residual error for results produced by the method. Error is assumed to be randomly distributed over all laboratories and test portions, sums to zero, and is measured by σ_e. In this practice, the minimum error of the method is estimated from duplicate results on each replicate.

A2.2 *ANOVA Table*—Practice E 691 does not follow the traditional calculation scheme. This practice follows the same approach used in Practice E 691. Table A1.1 displays the analysis of variance relationships for the Plan B design. The derivation of the calculations used in Section 10 of the practice is based upon Table A1.1.

A2.3 *Derivations*—The standard deviations are obtained by setting the expected mean squares (EMS) equal to the corresponding experimental mean squares (MS). Type B experiments generate three kinds of differences used to measure the variability contribution of each of the three levels included in the experiment: differences between duplicate results on a test portion: $D = (x_1 + x_2)$; differences between a replicate average and the laboratory's average: $d_1 = (X - \bar{x})$; and differences between a laboratory average and the average of all laboratories: $d_2 = (\bar{x} - \bar{\bar{x}})$.

The corresponding pooled variances are as follows:

$$s_D^2 = \Sigma D^2 / 2pn$$
$$s_X^2 = \Sigma d_1^2 / p(n-1)$$
$$s_{\bar{x}}^2 = \Sigma d_2^2 / (p-1)$$

A2.3.1 The variances of the three effects from Table A1.1:

$$\sigma_e^2 = \Sigma D^2 / 2pn = s_D^2$$
$$\sigma_{repl}^2 = \frac{1}{2}(2\Sigma d_1^2 / p(n-1) - \sigma_e^2) = s_X^2 - \frac{1}{2}s_D^2$$
$$\sigma_{lab}^2 = \frac{1}{2n}(2n\Sigma d_2^2 / (p-1) - 2\sigma_{repl}^2 - \sigma_e^2) = s_{\bar{x}}^2 - \frac{1}{2}s_X$$

A2.3.2 The minimum standard deviation of the method is $s_M = \sqrt{s_D^2}$.

A2.3.3 For ILS conducted in accordance with 8.3.2 to measure the day-to-day within-laboratory variability (repeatability standard deviation,) the required standard deviations are:

$$s_r = \sqrt{\sigma_{repl}^2 + \sigma_e^2} = \sqrt{s_X^2 - \frac{1}{2}s_D^2 + s_D^2} = \sqrt{s_X^2 + \frac{1}{2}s_D^2}$$

$$s_R = \sqrt{\sigma_{lab}^2 + \sigma_{repl}^2 + \sigma_e^2} = \sqrt{s_{\bar{x}}^2 - \frac{1}{n}s_X^2 - \frac{1}{2}s_D^2 + s_D^2}$$

$$= \sqrt{s_{\bar{x}}^2 + \frac{n-1}{n}s_X^2 + \frac{1}{2}s_D^2}$$

A2.3.4 For ILS conducted in accordance with 8.3.1 for materials of unknown homogeneity to eliminate the effects of material inhomogeneity, the homogeneity effect variance and the standard deviation for the reproducibility are:

$$s_H^2 = \sigma_{repl}^2 = s_X^2 - \frac{1}{2}s_D^2$$

$$s_R = \sqrt{\sigma_{lab}^2 + \sigma_e^2} = \sqrt{s_{\bar{x}}^2 - \frac{1}{n}s_X^2 + s_D^2}$$

TABLE A1.1 ANOVA Table^A

Source	Definition	MS	EMS
Laboratories	$SS_1 = 2n\Sigma(\bar{x} - \bar{\bar{x}})^2$	$SS_1/(p-1)$	$\sigma_e^2 + 2\sigma_{repl}^2 + 2n\sigma_{lab}^2$
Replicates	$SS_r = 2\Sigma(X - \bar{x})^2$	$SS_r/p(n-1)$	$\sigma_e^2 + 2\sigma_{repl}^2$
Error	$SS_e = \Sigma(D)^2/2$	$SS_e/pn(2-1)$	σ_e^2

^{A} The variable symbols are defined in Section 10. SS = sum of squares, MS = mean squares, and EMS = expected mean squares.

The homogeneity *F*-ratio is the ratio of the replicate EMS to the error EMS:

$$F_H = (s_M{}^2 + 2s_H{}^2)/s_M{}^2,$$

which follows the *F* distribution with $p(n - 1)$ and pn degrees of freedom.