Designation: ~~C802 – 09a~~ C802 – 14

Standard Practice for Conducting an Interlaboratory Test Program to Determine the Precision of Test Methods for Construction Materials[1]

This standard is issued under the fixed designation C802; 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.

1. Scope*

1.1 This practice describes techniques for planning, conducting, and analyzing the results of an interlaboratory study ~~of~~ (ILS) with the objective of developing the precision statement of a test method. It is designed to be used in conjunction with Practice C670. ~~Thus, the procedures recommended in this practice have the limited purpose of providing reliable information on which precision statements of the type described~~ The methods used in this standard are consistent with those in Practice ~~C670~~E691 ~~can be based. It is not appropriate for use in programs whose purpose is to develop a test method or to assess the relative merits of two or more test methods.~~.

1.2 This practice is not intended for use in programs whose purpose is to develop a test method or to assess the relative variability of two or more test methods. Refer to Practice C1067 for procedures to evaluate the ruggedness of a test method.

1.3 The ~~values stated in inch-pound units are to be regarded as standard. The values given in parentheses are mathematical conversions to SI units that are provided for information only and are not considered standard.~~system of units for this practice has not been specified. Dimensional quantities in the practice are presented only in examples of calculations.

1.4 *This standard does not purport to address all 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]

C109/C109M Test Method for Compressive Strength of Hydraulic Cement Mortars (Using 2-in. or [50-mm] Cube Specimens)
C136 Test Method for Sieve Analysis of Fine and Coarse Aggregates
C311/C311M Test Methods for Sampling and Testing Fly Ash or Natural Pozzolans for Use in Portland-Cement Concrete
C670 Practice for Preparing Precision and Bias Statements for Test Methods for Construction Materials
C1067 Practice for Conducting a Ruggedness Evaluation or Screening Program for Test Methods for Construction Materials
E105 Practice for Probability Sampling of Materials
E177 Practice for Use of the Terms Precision and Bias in ASTM Test Methods
E178 Practice for Dealing With Outlying Observations
E456 Terminology Relating to Quality and Statistics
E691 Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method

3. Terminology

3.1 *Definitions:*

3.1.1 For definitions of general statistical terms, refer to Terminology E456.

3.1.2 For definitions of terms associated with precision of test methods for construction materials, refer to Practice C670.

4. Significance and Use

4.1 ~~Certain criteria need to be met before undertaking~~This practice provides requirements for planning and conducting an interlaboratory study to ~~determine the precision of~~ obtain data to develop single-operator and multilaboratory precision statements

[1] This practice is under the jurisdiction of ASTM Committee C09 on Concrete and Concrete Aggregates. This practice was developed jointly by ASTM Committee C01, C09, D04, and D18, and is endorsed by all four committees.

Current edition approved ~~Dec. 15, 2009~~Dec. 15, 2014. Published ~~February 2010~~June 2015. Originally approved in 1974. Last previous edition approved in 2009 as ~~C802 – 09.~~C802 – 09A. DOI: ~~10.1520/C0802-09A.~~10.1520/C0802-14.

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

***A Summary of Changes section appears at the end of this standard**

for a test method. It ~~is not necessary that all of the following conditions described be completely fulfilled in every case; however, if some conditions are not met or are met incompletely, the program will become more complicated and require more work and expense, or may result in impaired information. The recommendations outlined in this section are intended to ensure that the test method is free of technical difficulties to the greatest extent possible before an expensive and time-consuming interlaboratory study~~includes methods to evaluate data consistency before carrying out the calculations to develop the precision statement. The procedures are compatible with those in Practice E691~~is undertaken.~~.

~~3.1.1 The first requirement is the existence of a valid and well-written test method that has been developed in one competent laboratory (or by cooperative work in a small number of laboratories), and has been subjected to a screening procedure, or to ruggedness testing as described in Practice C1067. As a result of the screening procedure and some experience with the test method in the sponsoring laboratory and one or two others, a written version of the test method has been developed (but not necessarily published as a standard method) that describes the test procedure in terms that can easily be followed in any properly equipped laboratory. Conditions that affect the test results should be identified and the proper degree of control of those conditions should be specified in the description of the test procedure (see Note 1).~~

~~NOTE 1—The desired degree of control of conditions that affect test results may not always be practically achievable, and tolerances in the test method should recognize this fact. Variations in test results due to variations in such conditions contribute to the total variation which determines the precision of the test method. If the resulting variation is so great that uncertainties in average values obtained by the test method are unacceptably high, then the test method itself is at fault, and efforts should be made to improve it or to replace it by a better one. An expensive and time-consuming interlaboratory study should not be undertaken on such a test method.~~

~~3.1.2 Any apparatus required for performing the test should be appropriately designed and available at reasonable cost.~~

~~3.1.3 Personnel in participating laboratories should have enough experience with the test method so that they are competent to run the test. The importance of this requirement will vary with the complexity of the method and the degree to which it departs from familiar procedures.~~

~~3.1.4 Preliminary knowledge should exist about how changes in materials and conditions affect the test results. There should be a reasonable degree of certainty that the within-laboratory variances are the same in different laboratories, and that troublesome interactions do not exist. These conditions are investigated in the analysis of the data of an interlaboratory study, and are discussed further in 8.2.2, 8.2.3, and Appendix X1.~~

~~3.1.5 Facilities and procedures for procurement, preparation, and distribution of samples must be available and should be as simple and free of difficulties as practicable.~~

~~3.1.6 Selection of samples must be done by a randomization process, and one person who is familiar with randomization procedures should be responsible for seeing that the procedure is carried out. Refer to Recommended Practice E105.~~

~~3.1.7 Adequate numbers of participating laboratories, operators, and materials must be available. Requirements in these areas are specified in Sections 4 and 5.~~

~~3.1.8 The entire interlaboratory test program should be developed from the beginning with the help and advice of persons familiar with statistical procedures and with the materials involved (see Note 2). The same persons who design the experiment should also carry out, or at least have control over, the process of analysis of the data.~~

~~NOTE 2—It may not always be possible to obtain people who are familiar with the materials involved who have a sufficient knowledge of the proper statistical techniques and their proper use. In this case, the committee should obtain the services of a competent statistician who has experience in practical work with data from materials testing, and provide him with an opportunity for learning something about the particular materials and test method involved. Planners of an interlaboratory study should also be warned to avoid the pitfall of assuming that the use of statistical analysis software programs necessarily results in special expertise in the handling of data or the interpretation of results.~~

4.2 The ILS data obtained from this practice are intended for developing the precision values for writing single-operator and multilaboratory precision statements in accordance with Practice C670.

4.3 Appendix X1 provides an example to illustrate the calculations to obtain the precision values of the test method from the ILS data. This may be used to check a user-developed electronic spreadsheet for carrying out the calculations.

4.4 ~~It~~Appendix X2 ~~is important to bear in mind that estimates of the precision~~ discusses the additional calculations required for an interlaboratory study of a test method ~~are always based on a particular set of data obtained at a particular time and they need to be kept up-to-date. As materials, apparatus, and conditions change, and operators change or gain more experience, the characteristic precision of the results obtained may change, especially if the test method is new. In some cases, it may even be desirable to conduct more tests at a later date (though not necessarily a repetition of the complete interlaboratory study) in order to provide a check on estimates previously obtained and either verify them or introduce revisions.~~that includes making test specimens as part of the procedure. In this case, batch-to-batch variability needs to be considered.

4.5 Appendix X3 discusses the use of analysis of variance as an alternative approach to obtain the precision values from the ILS data.

5. General Requirements

5.1 Certain criteria need to be met before undertaking an interlaboratory study to determine the precision of a test method. If some conditions are not met or are met incompletely, the program will become more complicated to administer and require more

work and expense, or may result in impaired information. The requirements outlined in this section are intended to ensure that the test method is free of technical difficulties to the greatest extent possible before an expensive and time-consuming interlaboratory study is undertaken.

5.2 The first requirement is the existence of a valid and well-written test method that has been developed in one laboratory and has been subjected to ruggedness evaluation of the testing procedure and conditions as described in Practice C1067. As a result of the screening procedure and some experience with the test method in the sponsoring laboratory and one or two others, a written version of the test method has been developed (but not necessarily published as a standard) that describes the test procedure in terms that can be followed by a competent operator in any properly equipped laboratory. Critical conditions that affect the test results need to be identified and the proper and realistic degree of control of those conditions have to be specified in the description of the test procedure.

5.2.1 The tolerances established for various conditions in a test method provide reasonable ranges for these conditions and recognize that precise values with small tolerances may not be achievable in practice. Variations in test results due to variations in such conditions contribute to the total variation, which determines the precision of the test method. If the resulting variation is so great that uncertainties in average values obtained by the test method are unacceptably high, the test method itself is at fault and it will need to be improved or replaced by a better one. An expensive and time-consuming interlaboratory study is not recommended for such a test method.

5.2.2 Apparatus required for performing the test must be defined clearly and must be available or able to be produced. If alternative apparatus is permitted, criteria need to be provided on the performance requirements of the apparatus, such as by specifying acceptable limits of measurements on standard reference materials.

5.3 Personnel in laboratories participating in the ILS should have adequate experience with routine laboratory procedures so that they are competent to run the test. The importance of this requirement will vary with the complexity of the method and the degree to which it departs from familiar procedures.

5.4 It is helpful to have preliminary knowledge about how changes in materials and conditions affect the test results. There should be a reasonable degree of certainty that the single-operator variances are the same in different laboratories, and that troublesome interactions do not exist. These conditions are investigated in the initial analysis of the data of an interlaboratory study, and are discussed further in 10.4.

5.5 Facilities and procedures for procurement, preparation, and distribution of samples or test specimens must be available.

5.6 Selection of samples or test specimens must be done by a randomization process, and one person who is familiar with randomization procedures needs to be responsible for seeing that an appropriate randomization technique is used. Refer to Practice E105.

5.7 The precision of the test method should be evaluated on different materials with a range of the characteristic being measured that encompasses the typical use of the method in practice. (See 7.1 and 7.2.)

5.8 Adequate numbers of participating laboratories, operators, and materials must be available. Requirements in these areas are specified in Sections 6 and 7.

5.9 The entire interlaboratory test program should be developed from the beginning with the help and advice of persons familiar with statistical procedures and with the materials involved. The ASTM International Interlaboratory Study Program can support subcommittees in the development of precision statements by assisting in the design of an interlaboratory study, distribution of specimens or samples, data analysis, and preparation of a draft research report. Additional information about the ASTM ILS program can be obtained from the ASTM Website.

5.9.1 It may not always be possible to obtain people who are familiar with the materials involved and who have a sufficient knowledge of the proper statistical techniques and their proper use. In this case, the subcommittee should obtain the services of a statistician who has experience in practical work with data from materials testing, and provide that person with an opportunity for learning something about the particular materials and test method involved. Planners of an interlaboratory study need to avoid the pitfall of assuming that the use of statistical analysis software programs necessarily results in special expertise in manipulating the data or interpreting the results.

5.10 It is important to bear in mind that estimates of the precision of a test method are always based on a particular set of data obtained at a particular time and precision values need to be kept up-to-date. As materials, apparatus, and conditions change, and operators change or gain more experience, the characteristic precision of the results obtained may change, especially if the test method is new. In some cases, it may be desirable to conduct more tests at a later date (though not necessarily a repetition of the complete interlaboratory study) in order to provide a check on estimates previously obtained and either verify them or introduce revisions. When a subcommittee revises a test method, it should consider whether the proposed changes might affect the precision of the method. If there is a possibility that precision will be affected, limited interlaboratory testing is recommended to evaluate whether the existing precision statement is still applicable or if a new ILS needs to be organized to better reflect the precision of the revised method.

6. Laboratories

6.1 ~~The problem of obtaining competent~~ Obtaining participating laboratories for an interlaboratory study is often one of the most difficult ~~ones~~problems connected with the process. The number of laboratories available is seldom as extensive as one would like, and if the test method is new, complicated, or expensive and time-consuming to run, the problem is further complicated. ~~The problem usually becomes one of finding and obtaining the cooperation of enough qualified laboratories to obtain meaningful estimates of precision, rather than that of selection among a group of available laboratories. If there is great difficulty in obtaining a sufficient number of competent and willing laboratories, then the possibility exists that the test method should not be subjected to a formal interlaboratory study.~~

6.2 For the purposes of programs using this ~~recommended~~ practice, it is recommended that at least ten ~~participating~~competent laboratories be included (**1, 2**).[3] In cases where it is impossible to obtain ten laboratories, the effect of an increased number may be obtained by repeating the program with the same group of laboratories six months later. ~~Usually, results obtained from the same laboratory after a time lapse of approximately six months display most of the characteristics of results from a different laboratory, especially if a different operator and apparatus can be used. If this procedure is followed, it is necessary to be sure that the same materials are used, and that their characteristics have not changed in the interim.~~If this procedure is followed, it is necessary to be sure that the same materials are used, and that their characteristics have not changed in the interim. This approach, however, may not provide a proper measure of the between-laboratory component of variance, unless different operators or equipment, or both, are used for the repeat testing. In any case, six is the absolute minimum number of laboratories for evaluating the precision of a test method. This means that at least seven to eight laboratories should be in the ILS study in case problems are encountered with the data provided by a participating laboratory.

6.3 In general, it is recommended that any laboratory that is considered qualified to run the test in routine testing situations should be permitted and encouraged to participate. "Qualified" implies proper laboratory facilities and testing equipment, competent operators familiar with ~~the test method, a reputation for~~routine laboratory techniques, a history of reliable testing work, and sufficient time and interest to do a good job. It does not mean, however, that only a select group of laboratories that are considered to be those best qualified for the interlaboratory study should be picked. Precision estimates for inclusion in a test method must be obtained under conditions and through the efforts of laboratories and personnel that are representative of the situations in which the test method will be used in practice (**32**). If a laboratory ~~has~~satisfies all the other requirements, but its personnel has had insufficient experience with the method, the operators in that laboratory should be given an opportunity to familiarize themselves with the method and to practice its application before the interlaboratory study starts.

7. Materials

7.1 *Number*—The number ~~and type~~ of materials to be included in an interlaboratory study will depend on the following:

7.1.1 The range of the values of the property ~~being measured on a given material~~that may be measured in practice and how the precision varies over that ~~range,~~range;

7.1.2 The ~~number~~types of ~~different~~ materials to which the test method is to be ~~applied,~~applied;

7.1.3 The difficulty and expense involved in obtaining, processing, and distributing ~~samples,~~samples or specimens;

7.1.4 The difficulty of, length of time required for, and expense of performing the ~~tests,~~tests; and

7.1.5 The uncertainty of prior information on any of these points. For example, if it is already known that the precision is relatively constant or proportional to the average level over the range of values of interest, a smaller number of materials will be needed than if it is known that the precision changes erratically at different levels. A preliminary pilot or screening program may help to settle some of these questions, and may often result in the saving of considerable time and expense in the full interlaboratory study (**41**).

7.2 In general, ~~a minimum of three materials should be considered acceptable.~~at least three materials or three different average values of the measured test characteristic is considered acceptable. The materials need to be selected to obtain as broad a range of the test characteristic as is practicable. If the test method is used to determine properties that are used for acceptance testing in a specification, it is particularly important that materials be included in the ILS whose properties are near the specification limits.

7.3 *Specimen Distribution*—The ILS is based on the assumption that all tests are performed on specimens that are as similar as is possible. Generally, two approaches are used for making and distributing the specimens or materials for the ILS.

7.3.1 For a test method that does not involve production of the test specimens as part of the method, specimens are produced at one location from a homogenous sample and then distributed to the participating laboratories. The specimens need to be assigned to the participating laboratories on a random basis. If the characteristic to be measured changes with age, specific instructions on test age need to be provided.

7.3.2 For a test method that involves fabrication of test specimens as part of the method, the raw materials for making the test specimens are shipped to the participating laboratories. In this case, samples of the constituent materials are taken from homogenous blends of the materials. The samples are selected on a random basis for shipment to the participating laboratories. Facilities are needed that have the proper equipment for blending the materials.

[3] The boldface numbers in parentheses refer to the list of references at the end of this practice.

7.3.3 In some cases, it is not possible to distribute materials to participating laboratory because of the nature of the material or effects of transportation or age. This may require operators from participating laboratories to convene at one location to test the materials. This procedure is used commonly in developing precision statements for fresh concrete test methods.

8. Estimates of Precision

8.1 In accordance with ~~Recommended~~ Practice C670, the procedure described in this practice is designed to provide data to develop two basic estimates of the precision of a test method: ~~((~~(a)~~a)~~ single-operator precision, and ~~((~~(b)~~b) multilaboratory precision. If other estimates of precision are desired, other references should be consulted (see Practice~~ multilaboratory precision E177~~)~~ (**51**). (See Note 1.)

8.2 *Single-operator precision* provides ~~an estimate of the~~ estimates of the inherent variability of the test method and the maximum difference that may be expected between ~~duplicate~~replicate measurements made on the same material in the same laboratory by the same operator using the same apparatus within a time span of a few days. The words "may be expected" mean that there is 5 % likelihood that the difference will exceed the stated maximum difference, even if testing conforms to the test method. In Practice C670, the maximum acceptable difference is referred to as the "difference limit (d2s)" or "difference limit (d2s%)" if the coefficient of variation is the appropriate measure of precision.

8.3 *Multilaboratory precision* provides ~~an estimate of the~~ estimates of the variability among laboratories and the maximum difference that may be expected between measurements made on the same material in two different laboratories.

8.4 If estimates of precision due to other factors are required for the test method, the ILS needs to be planned to provide data to develop the appropriate statistics for the systems of causes being considered and the appropriate combination of modifiers given in Practice E177 should be used to describe those statistics. The advice of a statistical consultant is recommended for these cases.

NOTE 1—Appendix X2, for example, explains how to analyze ILS data for developing the *single-operator, multi-batch precision* of a test method that involves making the test specimens as part of the procedure. Another example is developing the *single-operator, multi-day precision*, which would involve the additional variability due to testing on different days.

9. Collection of Data

9.1 In order to minimize the problems concerned with analysis of data, a definite form and instructions for obtaining and recording the data ~~should~~ have to be developed and distributed to all participating laboratories.

9.2 *Directions to Laboratories*—The directions to the laboratories should deal mainly with reporting of data. No special instructions for performing the tests that differ from those given in the ~~description of the~~ test method should be included. The laboratories ~~should~~must be ~~cautioned~~instructed to conduct tests and report results exactly as specified in the test method, with the one exception as noted in ~~7.2.2~~9.2.2. Often data are disseminated in digital form, but laboratories need to maintain hard-copies of their data to provide documentation in the event of digital data corruption.

9.2.1 *Averaging Test ~~Results—~~Determinations*—Laboratories should particularly be cautioned against practices such as running a number of tests and selecting the "best" ones or reporting the average of several determinations, except as such averaging is specified in the test method. For example, Test Method C109/C109M specifies three or more test specimens, and requires that the strength of all acceptable test specimens made from the same ~~sample~~batch and tested at the same period shall be averaged and reported. In this case, the directions for the interlaboratory ~~test should~~study must specify the number of individual determinations to be obtained and reported. ~~Whenever~~If a test result is defined, either in the test method or in the instructions to laboratories participating in an interlaboratory test program, as the average of a particular number of determinations, the individual determinations ~~should~~ shall always be reported, in addition to the averages. ~~When two or more measurements are averaged to obtain a test result, the data from the interlaboratory test program may be used to develop an estimate of the precision of these individual measurements. See 3.3.3 of Practice C670.~~

9.2.2 *Rounding of Data:*

9.2.2.1 Generally, laboratories ~~should be required~~ need to report all figures obtained in making the measurements, rather than rounding the results before recording them. In some cases, this may result in recording of more digits than is customary or even more than the test method calls for in the section on ~~Reporting (see~~ Reporting. ~~X1.3.1).~~ This is necessary because the variation from which information about the precision of the test method comes is contained in the least significant digits, which are often discarded in reporting the results of routine testing (**63**). For example, Test Method C136 calls for reporting ~~of percentages~~ the percentage retained on a sieve to the nearest whole number. This is adequate for the usual reporting purposes, but for purposes of determining precision, at least one decimal place is needed. It is better to require the reporting of too many decimal places than too few, ~~since~~because a decision about rounding all data can be made when the analysis is done. If too few places are reported, however, valuable information may be irretrievably lost, and the result might well be the impairment of the entire program.

9.2.2.2 ~~In cases where~~ If a test ~~result~~determination is the result of a calculation based on two or more measured quantities, the basic measurements should be used in the calculations without any rounding. The planners of the interlaboratory program will then have to determine how many places need to be reported in order to retain the essential information for determining variability. Sometimes it is advisable to ask the laboratories to report the basic quantities measured instead of, or in addition to, the final calculated result. This enables the final result to be checked, or changes in decisions about the test results to be made, when the data are analyzed. ~~This procedure is especially appropriate if the results are to be analyzed by computer, and the program can be~~

~~utilized to perform the basic calculations and analyze the calculated results.~~An example would be a strength test for which the measured specimen dimensions along with the ultimate load should be reported so that the reported strength can be verified.

9.3 *The Data Sheet*—This practice is based on the following assumptions: *p* laboratories each have made *n* replicate ~~measurements~~determinations on each of *q* materials ~~(see Ref (74)).~~. Table 1 and Table 2 are ~~sample~~ examples of data sheets for an individual laboratory and for a summary of data for the entire ~~program for a program~~ ILS program with: *p* ~~= ten~~ = ten laboratories, *n* ~~= four replicates, (test results on each material in each laboratory), and,~~ = four replicate determinations, and *q* ~~= five~~ = five materials. These data sheets suggest the ~~forms~~format to be used ~~when~~if an individual ~~measurement~~determination constitutes ~~the basic~~ a test result. ~~In cases where individual measurements~~If individual determinations are averaged or otherwise subjected to calculation to produce a test result, the ~~form~~format of the individual laboratory data sheet may be altered or a secondary sheet provided to permit recording ~~of~~ the fundamental measurements and the test results.

9.4 *Number of ~~Replicates:~~ Replicates*—Even if the test method calls for a single determination as a test result, replicate determinations are required in the ILS in order to obtain information for calculating the single-operator precision.

9.4.1 The number of replicate determinations to be made on each material in each laboratory depends largely on the number of laboratories participating, on the homogeneity of the material, and on the expense, difficulty, and time involved in increasing the number of determinations. ~~It should be recognized that in~~ In order to obtain the necessary information to write a meaningful precision statement, it is often necessary to use more replicates in the interlaboratory study than is required for routine use of the test method. An increase in the number of replicates improves the estimates of ~~within-laboratory~~single-operator precision but has no effect on between-laboratory precision (**82**). It is recommended that, for 10 to 15 participating laboratories, at least three replicates ~~should be required. In cases where~~ are required. If it is not possible to obtain 10 participating laboratories, the number of replicates, *n*, should be ~~equal to or greater than~~ at least (30/*p*) ~~+ 1. For more than 15 laboratories, the number of replicates may be reduced to two. (If~~) + 1. If 30 is not a multiple of *p*, 30/*p* is rounded to the next higher ~~integer.)~~ integer. For more than 15 laboratories, the number of replicates may be reduced to two. This will give an adequate estimate of ~~within-laboratory~~single-operator precision, but information about ~~between-laboratory~~multilaboratory precision is not as good as desired with fewer than 10 laboratories.

9.4.2 The variation among replicate ~~measurements~~determinations is supposed to be representative of the irreducible error variance characteristic of the test method. In some cases, it is possible to take supposedly replicate measurements in such a manner that there is little or no opportunity for chance ~~variation,~~variation; and the measurements are in effect simply repetitions of the same measurement. For example, in making a chemical analysis by atomic absorption or some other kind of automatic measuring device, laboratories have been known to take three readings of the meter on the same sample in quick succession. The three readings so taken were almost identical, but were still reported as replicate readings. In cases such as this, three separate readings with different portions of the sample ~~should be tested possibly on different days,~~or with separate specimens should be obtained, with the same operator and apparatus ~~in order~~ to provide meaningful replicate measurements.

9.5 *Outliers*—Section 10.4 describes a procedure for identifying test determinations that have unexpected variations from those obtained by other participants in the ILS. In general, the practice of discarding individual test ~~results that~~determinations, which appear to differ by suspiciously large amounts from the others, ~~should~~ is to be avoided unless there is clear evidence that there was some physical reason to consider the ~~result faulty. It is recommended that no purely statistical criterion be used for the purpose. In particular, laboratories should be asked~~ determination faulty. Discarding results with unexpected deviations but without a proper basis or an assignable cause for the deviations may result in unrealistic precision values that may not be relevant to the test method. On the other hand, retaining invalid results with unexpectedly large variations may result in precision values that tolerate less careful testing. Laboratories must be instructed to report all results in their proper place and include notes describing the conditions surrounding those results that are suspected of being faulty. Sometimes if a test really went wrong, a laboratory should discard the results and repeat the test. Tests ~~should~~are not to be repeated, however, just because the results don't look good. Further ~~discussion of the problems of~~ guidance on dealing with outliers is given in ~~Appendix X2,~~ Practice E178~~,~~ and in Refs (**92**)~~,~~ ~~**and**~~ (~~**10**~~**5**).

9.6 *Missing Data*—~~Sometimes individual items of data~~ The method of data analysis used in this practices assumes a balanced set of data, which means that each laboratory provides the required number of determinations for all materials. Sometimes individual determinations are missing from the summary because they were discarded, failed to be supplied by a laboratory, or for other reasons. In general, if the number of missing data items from all laboratories constitutes no more than ~~1 %~~ about 3 % of the total number of items, the analysis may be conducted as though the missing items were present. For example, if one ~~result~~ out of

TABLE 1 Data Sheet for an Interlaboratory Test Program for an ASTM Test Method

Laboratory: XX

Replicate	Material				
	A	B	C	D	E
a	_____	_____	_____	_____	_____
b	_____	_____	_____	_____	_____
c	_____	_____	_____	_____	_____
d	_____	_____	_____	_____	_____

TABLE 2 Summary Data Sheet for an Interlaboratory Test Program for an ASTM Test Method

Laboratory	Replicate	Material				
		A	B	C	D	E
1	a					
	b					
	c					
	d					
2	a					
	b					
	c					
	d					
3	a					
	b					
	c					
	d					
4	a					
	b					
	c					
	d					
5	a					
	b					
	c					
	d					
6	a					
	b					
	c					
	d					
7	a					
	b					
	c					
	d					
8	a					
	b					
	c					
	d					
9	a					
	b					
	c					
	d					
10	a					
	b					
	c					
	d					

four ~~replicates~~ required replicate determinations on a given material from ~~a~~laboratory, ~~*given*~~*i*, ~~laboratory~~ out of 10 laboratories is missing, the three ~~remaining results~~reported determinations should be added and ~~then~~ divided by ~~3~~three to ~~get~~obtain the average, $\bar{X}_{i,i}$. The ~~within-laboratory~~single-operator variance, s_{iri}^2, should ~~also~~ be calculated using ~~3~~three for the number of ~~results.~~determinations. From then on, both ~~results~~values should be used as though they were based on four ~~measurements.~~determinations. If the number of missing results exceeds ~~1 %~~ 3 % of the total, some of the tests should be repeated in order to obtain proper measurements for the missing values. Missing values handled in this way must be individual values distributed throughout the mass of data, and ~~should~~are not to be concentrated as a group in one laboratory-material ~~cell.~~ cell (see Note 2). If the latter occurs, the laboratory should provide another group of measurements on the material in question. ~~Analysis-of-variance procedures exist for the analysis of such unbalanced sets of data. The advice of a statistical consultant should be obtained when such practices are used.~~

NOTE 2—A cell refers to the group of replicate test determinations for a particular combination of laboratory and material. Appendix X3 describes an analysis-of-variance procedure that can be used to analyze unbalanced sets of data. The advice of a statistical consultant should be obtained if such procedures are used.

10. Analysis of Data

10.1 An ILS is a type of experiment design known as a nested design or a hierarchal design (**6**). The general purpose of a nested design is to identify and quantify the sources of variation in a process. In the case of the ILS, the objective is to quantify the

single-operator and between-laboratory components of variance. Fig. 1 is a diagram of a single-stage nested design that is representative of the basic ILS described in this practice. The laboratories participating in the ILS are the first stage representing the factor "laboratory." For each laboratory, there are *n* replicate determinations. The replicate determinations are considered to be "nested" within the factor "laboratory."

10.2 The procedure described herein is simplified, and statistical terms are avoided to the greatest extent possible in order to make the ~~recommended~~ practice ~~easily~~ usable by persons with little statistical background. This exposes the ~~recommended~~ practice to the danger that, although the technique recommended is widely applicable to many situations using many kinds of data, it may be used mechanically in situations ~~in~~for which it is not ~~applicable by persons who are not familiar with the statistical background of the recommended procedures.~~ applicable. For this reason, it is recommended to seek the advice of a person who is familiar with the statistical procedures before undertaking analysis of ~~an interlaboratory study~~ ILS data by this or any other published procedure. An example of the procedure is given in Appendix X1. For further description of the method, see Ref (**51**). The procedure does not require sophisticated software and can be implemented using an electronic spreadsheet.[4]

10.3 *Single-Operator and Between-Laboratory Components of Variance for Each Material*—Before starting the analysis, plot the data. This can be done by making a scatter plot of the test determinations for each laboratory. A separate plot can be made for each material, or all data can be shown in one plot. These plots will reveal any potential data inconsistencies that will be investigated further in accordance with 10.4. The first step in the analysis is to obtain estimates of single-operator and between-laboratory components of variance for each material. This may be done by setting up the data as shown in Table 3 and using the equations presented in this section. Table 3 is set up as an example using Material A for tests in ten laboratories (p = 10) with four replicate determinations per laboratory (n = 4) to correspond with the example summary data sheet in Table 2. Each row of data represents a particular laboratory-material combination and often called a *cell*. Tables similar to Table 3 would be used for each material in the study. In the equations that follow, the subscript g is used to designate a single test determination for a particular material in one laboratory and goes from 1 to n, where n is the number of replicates for each laboratory. The lower case letters (a, b, c, d) in Table 3 represent the replicate test determinations. The subscript i is used to designate a particular laboratory in the analysis and goes from 1 to p, where p is the total number of laboratories. The subscript j is used to designate the different materials, and goes from 1 to q, where q is the total number of materials. As shown in Table 1 and Table 2, the different materials are identified with capital letters A, B, C, and so forth.

10.3.1 *Single-operator analysis*—The averages, $\bar{X}_{iA}$, and variances, S^2_{riA}, in the last two columns of Table 3 are the single-operator averages and variances for the given material (in this example, it is Material A). These quantities are calculated from the n replicate test determinations within each of the p laboratories as follows:

x_{gij} = single test determination g by laboratory i for material j

$$\bar{X}_{ij} = \frac{\Sigma x_{gij}}{n} \quad (1)$$

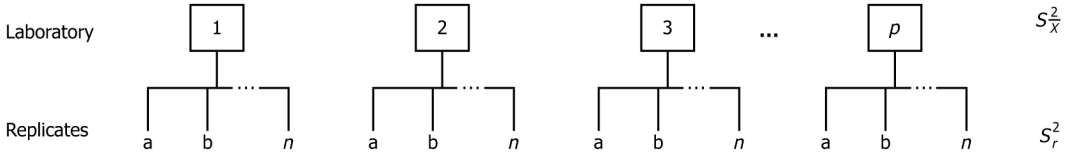
= average of n replicate test determinations for laboratory i on material j

$$s^2_{rij} = \frac{\Sigma\left(x_{gij} - \bar{X}_{ij}\right)^2}{n-1} \quad (2)$$

= single-operator variance of replicate determinations for laboratory i on material j

10.3.2 *Between-laboratory analysis*—From the single-operator averages, Eq 1, and variances, Eq 2, for each laboratory, the following quantities are calculated for the given material: (1) the pooled single-operator variance, (2) the overall average, (3) the

[4] A statistical software package called Dataplot® is available from the National Institute of Standards and Technology that will perform the plotting and calculations described in this practice. The program can be downloaded from homepage: http://www.nist.gov/itl/sed/dataplot.cfm. Instructions on how to analyze ILS data can be found at this site: http://www.itl.nist.gov/div898/software/dataplot/refman1/auxillar/e691.htm.



Between-laboratory component of variance = $S^2_L = S^2_{\bar{X}} - \frac{S^2_r}{n}$

FIG. 1 Diagram of a Single-Stage Nested Experiment Design

TABLE 3 ~~Between and Within~~ Analysis for Material A[A]

~~colwidth="68pts"> colname="col2" colwidth="84pts"> colname="col3" colwidth="84pts"> colname="col4" colwidth="84pts">~~

Laboratory	Data				~~Average~~	~~Within-Laboratory Variance~~
	a	b	c	d	~~τ~~	~~s_1^2~~
~~1~~					~~$\bar{x}_1$~~	~~s_1^2~~
~~2~~					~~$\bar{x}_2$~~	~~s_2^2~~
~~3~~					~~$\bar{x}_3$~~	~~s_3^2~~
~~4~~					~~$\bar{x}_4$~~	~~s_4^2~~
~~5~~					~~$\bar{x}_5$~~	~~s_5^2~~
~~6~~					~~$\bar{x}_6$~~	~~s_6^2~~
~~7~~					~~$\bar{x}_7$~~	~~s_7^2~~
~~8~~					~~$\bar{x}_8$~~	~~s_8^2~~
~~9~~					~~$\bar{x}_9$~~	~~s_9^2~~
~~10~~					~~$\bar{x}_{10}$~~	~~s_{10}^2~~

TABLE 3 Single-Operator and Between-Laboratory Analysis for Material A[A]

Laboratory	Data (Replicates)				Average,	Single-Operator Variance,
	a	b	c	d	$\bar{X}_{iA}$	s_{riA}^2
1					$\bar{X}_{1A}$	s_{r1A}^2
2					$\bar{X}_{2A}$	s_{r2A}^2
3					$\bar{X}_{3A}$	s_{r3A}^2
4					$\bar{X}_{4A}$	s_{r4A}^2
5					$\bar{X}_{5A}$	s_{r5A}^2
6					$\bar{X}_{6A}$	s_{r6A}^2
7					$\bar{X}_{7A}$	s_{r7A}^2
8					$\bar{X}_{8A}$	s_{r8A}^2
9					$\bar{X}_{9A}$	s_{r9A}^2
10					$\bar{X}_{10A}$	s_{r10A}^2

[A] ~~p = 10 laboratories.~~ = 10 laboratories

~~n = 4~~ = 4 replicate test ~~results~~determinations on each material in each ~~laboratory~~laboratory

~~Overall~~Overall average $\bar{X}_A = \frac{\Sigma \bar{X}_{iA}}{p}$ ~~$\bar{x}_A$~~

~~Pooled within-laboratory~~Pooled single-operator variance ~~$s_{rA}^2 = \frac{\Sigma s_{riA}^2}{p}$~~ ~~(pooled)~~

~~Variance~~Variance of laboratory averages ~~$s_{\bar{x}_A}^2$~~ $s_{\bar{X}_A}^2 = \frac{\Sigma(\bar{X}_{iA} - \bar{X}_A)^2}{p-1}$

~~Between-laboratory~~Between-laboratory component of variance ~~$s_{LA}^2 = s_{\bar{X}_A}^2 - \frac{s_{rA}^2}{n}$~~

variance of laboratory averages, and (4) the between-laboratory component of variance. These values are entered at the bottom of Table 3 and are calculated as follows (Note 3):

$$s_{rj}^2 = \frac{\Sigma s_{rij}^2}{p} \quad (3)$$

= pooled single-operator variance for material j (Note 4)

$$\bar{X}_j = \frac{\Sigma \bar{X}_{ij}}{p} \quad (4)$$

= overall average for all laboratories for material j

$$s_{\bar{X}_j}^2 = \frac{\Sigma(\bar{X}_{ij} - \bar{X}_j)^2}{p-1} \quad (5)$$

= variance of laboratory averages for material j

$$s_{Lj}^2 = s_{\bar{X}_j}^2 - \frac{s_{rj}^2}{n} \quad (6)$$

= between-laboratory component of variance for material j. If the calculated values is negative, the between-laboratory component of variance is taken as zero.

NOTE 3—Appendix X1 includes an example showing how these calculations are made for each material.

NOTE 4—The method of pooling variances used here applies only if the individual variances are based on the same number of replicate tests. In general, a pooled estimate of a variance is not obtained by averaging individual variances if the number of replicate determinations is not the same for all laboratories. Refer to a textbook on basic principles of statistics for the method to pool variances if the number of replicates is not constant.

10.4 ~~*Between-Laboratory and Within-Laboratory Analysis for Each Material—*~~*Data Consistency—*~~The first step in the analysis is to obtain estimates of between-laboratory and within-laboratory variances for each material. This may be done by using the form~~

shown in Before continuing with the remaining analysis of the ILS data to determine the precision of the test method, it is necessary to check each laboratory's data for consistency in terms of the average and the dispersion of the results. If data from one laboratory are not consistent with data from the other laboratories, it may be necessary Table 3. Table 3 is set up as an example, using material A in ten laboratories with four replicate test results per laboratory to correspond with the sample summary data sheet into eliminate that laboratory's data before completing Table 2. Similar tables should be set up for each material in the study. The subscriptthe analysis. Inconsistent data imay is used to designate a particular laboratory in the analysis and goes from 1inflate the calculated precision values and thereby encourage laboratories to p,tolerate the total number of laboratories.less careful testing. The approach for checking data consistency used in this practice is the same as in Practice E691Capital letter subscripts,. Two statistics A,B, etc., are used to designate quantities calculated for the different materials. The averages, evaluate data consistency: (1) $\bar{x}_i$, and variances, s_i^2, in the last two columns are the within-laboratory averages and variances for the given material, and are calculated from the *nh* replicate test results for each of -value and (2) the *pk* laboratories as follows:-value.

$\bar{x}_i$ = $\sum x_i/n$ = sum of *n* replicate test results for laboratory *i* divided by *n*.
s_i^2 = $(\sum x_i^2 - n\bar{x}_i^2)/(n-1)$ = sum of squares of *n* replicate test results for laboratory *i* less *n* times the square of the average for laboratory *i*, divided by one less than the number of replicate test results.

10.4.1 *Check Laboratory Averages*—The *h*-value is used to check whether the average value for a laboratory is consistent with the overall average of the other laboratories for a given material. The *h*-value is calculated for each laboratory and material as follows:

$$h_{ij} = \frac{\bar{X}_{ij} - \bar{X}_j}{s_{\bar{X}j}} \tag{7}$$

where:

$\bar{X}_{ij}$ = average of results for laboratory *i* and material *j* (Eq 1),
$\bar{X}_j$ = overall average of results for material *j* (Eq 4), and
$s_{\bar{X}j}$ = standard deviation of laboratory averages for material *j*, which is the square root of Eq 5.

NOTE 3—The results of the calculations described here may be subject to a rounding error if the numbers involved are large. See Appendix X1 and Note X1.1 for an example of this and a discussion of how to deal with this problem.

From the *p* individual within-laboratory averages and variances, four quantities for the given material; namely, the overall average, pooled within-laboratory variance, variance of laboratory averages, and between-laboratory component of variance, are calculated and entered on Table 3 as follows:

10.4.2 *Check Laboratory Dispersion*—The *k*-value for each laboratory is used to check the consistency of the single-operator variability for a given material. The *k*-value is calculated for each laboratory and material as follows:

$$k_{ij} = \frac{s_{rij}}{s_{rj}} \tag{8}$$

where:

s_{rij} = single-operator standard deviation of replicate determinations for laboratory *i* and material *j*, which is the square root of Eq 2, and
s_{rj} = pooled single-operator standard deviation for material *j*, which is the square root of Eq 3.

-
-

$\bar{x}_A$ = $\sum \bar{x}_i/p$ = sum of *p* averages for the laboratories divided by *p*
s_A^2 (pooled) = $\sum s_i^2/p$ = average of within-laboratory variances (see Note 4).
$s_{\bar{x}_A}^2$ = $[\sum \bar{x}_i^2 - p(\bar{x}_A)^2]/(p-1)$ = sum of squares of *p* within-laboratory averages less *p* times the overall average squared, divided by $p-1$.
$s_{L_A}^2$ = $s_{\bar{x}_A}^2 - [s_A^2 \text{ (pooled)}/n]$ = the variance of laboratory averages less $1/n$ times the pooled variance.

10.4.3 *Critical h- and k-values*—The calculated *h*- and *k*-values are compared with the critical values shown in Table 4, which is extracted from a larger table in Practice E691. The second column gives the critical *h*-value, which depends only on the number of laboratories. The subsequent columns give the critical *k*-values, which depend on the number of laboratories and the number of replicate test determinations. The *h*-values can be positive or negative, while the *k*-values are always positive. The critical values in Table 4 are the 0.5 % significance level. According to Practice E691, this significance level was chosen on the basis of judgment and experience so that not too many nor too few laboratories are flagged for further investigation. Appendix X1 of Practice E691 provides the basis for the critical values of *h* and *k*. Refer to Practice E691 for applicable values of *h* and *k* if more than 20 laboratories or more than 6 replicate determinations are involved in the ILS.

A sample work sheet showing exactly how these calculations are made appears in Appendix X1.

NOTE 4—The method of pooling variances used here applied only when all the individual variances being pooled are based on the same number of

measurements. In general, a pooled estimate of a variance is not obtained by averaging individual variances.

10.4.4 *Summary of h- and k-values*—BeforeThe *proceedingh* with- and *k*the analysis, it is necessary to investigate agreement of the data with the following two assumptions: (-values calculated for each laboratory and each material are assembled in a table. Values that exceed the critical values and values that approach the critical values should be highlighted. The *ah*) the variances are the same in different laboratories (homogeneity of variance), and (- and *bk*) the results show the same pattern -values should also be plotted as bar graphs grouped in two ways: (1) by laboratories and (2) by materials. The critical values should be drawn on the plots. The plots of *changeh* fromand *onek* material to another in different laboratories (lack of interactions). These two aspects of the analysis are discussed inand the marked tables give a picture of the overall character of the 8.2.2 andvariability of 8.2.3.the test method as well as singling out particular laboratories that should be investigated.

10.4.5 *Investigation of Agreement of Variances—Plots by Laboratories*—This method is based on the assumption that the within-laboratory variances in different laboratories (of which the Examples of plots of sh_r^2 in Table 3and its variations, are estimates) are the same. This does not mean that the sk_r^2 have to be very close together, since an individual variance can be about four times the average variance (for by laboratories p = 10 and n = 4) when all the calculated variances are really estimates ofare shown in Appendix X1 the same variance. In order to check for agreement among variances, it is helpful to plot the individual variances against the laboratories, draw a horizontal line across the plot at the level of the average variance, and examine the lowest and highest individual variances. A variance that is very low compared to the others may indicate that the laboratory is not permitting the normal causes for variation between results to show up, while a high variance indicates the lack of proper control of the testing process.based on the illustrative data. For each laboratory, the materials are grouped in increasing order of the overall average property value. The following guidelines can be used to evaluate differences between laboratories.

10.4.5.1 *h-Plot*—Table 4The *givesh* approximate values (upper 5 % level) for the ratio -plot indicates how the laboratory average property values for each material compare with the overall average for that material. There are several general patterns in these plots. In one, all laboratories have both positive and negative *h*-values among the materials. In the second, individual laboratories tend to have either positive or negative *h*of the largest variance to the sum of the variances -values for all materials, and the number of laboratories with negative values is approximately the same as the number of laboratories with positive values. Neither of these patterns is unusual or requires investigation, although they may tell something about the nature of the test method variability. In the third pattern, one laboratory has all positive (or negative) *h*-values compared with the other laboratories, which have substantially all negative (or positive) *h*that should not be exceeded -values. Such a pattern calls for an investigation of that laboratory. Another pattern to look for occurs within one laboratory, in which the (*hH*).-values for materials with low property levels are of one sign and for materials with high property levels the *h*-values are of the opposite sign. If the *h*-values are extreme, investigation is warranted. As described in Practice E691, the investigation should consider examination of potential clerical errors in recording or transcribing data and it should consider examination of the laboratory reports for deviations from the test method or ILS protocol.

8.2.2.2 The case of a small variance is not usually as troublesome as that of a variance that is too large. However, when one laboratory performs its tests in such a way that the normal causes of variation do not affect the results, an unrealistically low variance may occur. If no significantly high variance is present, as judged by the criterion given above, the following method may be used to examine a suspiciously low variance. The statistic used is the ratio of highest to lowest variance in the group. Table 5 gives the approximate values (upper 5 % level) for this ratio that should not be exceeded **(12)**.

10.4.5.2 *k-Plot*—OftenThe *k*the data from one -plot compares the single-operator variability among the laboratories. As stated, *k*laboratory may indicate a high-values are always positive. The primary pattern to look for is whether a laboratory has large (or small) *k*or low variance compared to the others, and elimination -values for all or most of the materials. Elimination of that laboratory from the analysis results may result in a set of data with similar *variancesk*-values for the remaining laboratories (seelaboratories. High Appendix X1).*k*-values represent high single-operator variability. A check for outliers may be used to examine the data for the particular laboratory-material combination. The case of a small *k*-value is not usually as troublesome as that of a large *k*If -value. If one laboratory, however, performs its tests in such a way that the normal causes of variation are not permitted to occur, there may be an unrealistically low single-operator standard deviation. Small *k*-values may indicate an insensitive measurement scale or other measurement problems. If all the *variancesk*-values are erratic, however, the test method is in trouble. Efforts to develop precision statements from the data should be suspended and further study of the test method should be undertaken to determine the causes for such erratic behavior. The advice of a statistical consultant should be obtained whenever if there is doubt about eliminating a laboratory with a high or low *variance.k*-value.

10.4.6 *Plots by Material*—If a plot by laboratory shows several *h*- or *k*-values near the critical values, look at the corresponding plot by material to see how that laboratory differs from the rest for a given material. Often an *h*-value that seems strong in the plot by laboratory, because of its relation to the values for the other materials, will turn out to be reasonably consistent with the other laboratories for the same material. On the other hand, the *h*- or *k*-value for the one laboratory may be revealed as strongly different from the values for the other laboratories in the plot by material. If so, this behavior should be investigated.

10.4.7 *Interactions*—A common problem with test results obtained from an interlaboratory study is the presence of interactions betweenamong laboratories and materials. This means that the pattern of change of the results obtained on a given group of materials in the material by one laboratory differs from the pattern obtained in another laboratory. by the other laboratories. In extreme cases, different laboratories may even fail to rate materials in the same order. order based on the measured average