~~Designation:C802–09~~ Designation: C802 – 09a

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 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 in Practice C670 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 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.

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
C670 Practice for Preparing Precision and Bias Statements for Test Methods for Construction Materials
C1067 Practice for Conducting A Ruggedness 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

3. Significance and Use

3.1 Certain criteria need to be met before undertaking an interlaboratory study to determine the precision of 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 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.

---

[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 Nov. 1, 2009. Published December 2009. Originally approved in 1974. Last previous edition approved in 2008 as C802–96(2008) ε1. DOI: 10.1520/C0802-09.~~

Current edition approved Dec. 15, 2009. Published February 2010. Originally approved in 1974. Last previous edition approved in 2009 as C802 – 09. DOI: 10.1520/C0802-09a.

[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.**

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.

3.2 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 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.

4. Laboratories

4.1 The problem of obtaining competent participating laboratories for an interlaboratory study is often one of the most difficult ones 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.

4.2 For the purposes of programs using this recommended practice, it is recommended that at least ten participating 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.

4.3 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 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 **(3).** If a laboratory has 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.

5. Materials

5.1 The number and type of materials to be included in an interlaboratory study will depend on the following:

5.1.1 The range of the values of the property being measured on a given material and how the precision varies over that range,

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

5.1.2 The number of different materials to which the test method is to be applied,

5.1.3 The difficulty and expense involved in obtaining, processing, and distributing samples,

5.1.4 The difficulty of, length of time required for, and expense of performing the tests, and

5.1.5 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 **(4).**

5.2 In general, a minimum of three materials should be considered acceptable.

6. Estimates of Precision

6.1 In accordance with Recommended Practice C670, the procedure described in this practice is designed to provide two basic estimates of the precision of a test method: (*a*) single-operator precision, and (*b*) multilaboratory precision. If other estimates of precision are desired, other references should be consulted (see Practice E177) **(5).**

6.2 *Single-operator precision* provides an estimate of the difference that may be expected between duplicate 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.

6.3 *Multilaboratory precision* provides an estimate of the difference that may be expected between measurements made on the same material in two different laboratories.

7. Collection of Data

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

7.2 *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 be cautioned to conduct tests and report results exactly as specified in the test method, with the one exception as noted in 7.2.2.

7.2.1 *Averaging Test Results*—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 and tested at the same period shall be averaged and reported. In this case, the directions for the interlaboratory test should specify the number of determinations to be obtained and reported. Whenever 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 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.

7.2.2 *Rounding of Data*:

7.2.2.1 Generally, laboratories should be required 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 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 **(6).** For example, Method C136 calls for reporting of percentages 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 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.

7.2.2.2 In cases where a test result 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.

7.3 *The Data Sheet*—This practice is based on the following assumptions: p laboratories each have made n replicate measurements on each of q materials (see Ref **7**). Table 1 and Table 2 are sample data sheets for an individual laboratory and for a summary of data for the entire program for a program with: p = ten laboratories, n = four replicates, (test results on each material in each laboratory), and, q = five materials. These data sheets suggest the forms to be used when an individual measurement constitutes the basic test result. In cases where individual measurements are averaged or otherwise subjected to calculation to produce a test result, the form of the individual laboratory sheet may be altered or a secondary sheet provided to permit recording of the fundamental measurements and the test results.

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

Laboratory:

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	_____	_____	_____	_____	_____

7.4 *Number of Replicates*:

7.4.1 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 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 precision but has no effect on between-laboratory precision **(8).** It is recommended that, for 10 to 15 participating laboratories, at least three replicates should be required. In cases where it is not possible to obtain 10 participating laboratories, the number of replicates, n , should be equal

to or greater than (30/*p*) + 1. For more than 15 laboratories, the number of replicates may be reduced to two. (If 30 is not a multiple of *p*, 30/*p* is rounded to the next higher integer.) This will give an adequate estimate of within-laboratory precision, but information about between-laboratory precision is not as good as desired with fewer than 10 laboratories.

7.4.2 The variation among replicate measurements 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, 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, with the same operator and apparatus in order to provide meaningful replicate measurements.

7.5 *Outliers*—In general, the practice of discarding individual test results that appear to differ by suspiciously large amounts from the others, should 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 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 not be repeated, however, just because the results don't look good. Further discussion of the problems of outliers is given in Appendix X2, Practice E178, and in Refs (**9** and **10**).

7.6 *Missing Data*—Sometimes individual items of data 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 % 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 four replicates on a given material from a given laboratory is missing, the three remaining results should be added and then divided by 3 to get the average, $\bar{x}_i$. The within-laboratory variance, s_i^2, should also be calculated using 3 for the number of results. From then on, both results should be used as though they were based on four measurements. If the number of missing results exceeds 1 % 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 not be concentrated as a group in one laboratory-material cell. 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.

8. Analysis of Data

8.1 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 which it is not applicable by persons who are not familiar with the statistical background of the recommended procedures. 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 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 **(5)**.

8.2 *Between-Laboratory and Within-Laboratory Analysis for Each Material*—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 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

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

Laboratory	Data				Average $\bar{x}_1$	Within-Laboratory Variance s_1^2
	a	b	c	d		
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

[A] p = 10 laboratories.
n = 4 replicate test results on each material in each laboratory.
Overall average $\bar{x}_A$ =
Pooled within-laboratory variance s_A^2 (pooled) =
Variance of laboratory averages $s_{\bar{x}_A}^2$ =
Between-laboratory component of variance $s_{L_A}^2$ =

with the sample summary data sheet in Table 2. Similar tables should be set up for each material in the study. The subscript i is used to designate a particular laboratory in the analysis and goes from 1 to p, the total number of laboratories. Capital letter subscripts, A, B, etc., are used to designate quantities calculated for the different materials. The averages, $\bar{x}_i$, 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 n replicate test results for each of the p laboratories as follows:

$\bar{x}_i$ = $\Sigma\, x_i/n$ = sum of n replicate test results for laboratory i divided by n.
s_i^2 = $(\Sigma\, x_i^2 - n\,\bar{x}_i^2)/(n-1)$ = 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.

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:

$\bar{x}_A$ = $\Sigma\, \bar{x}_i/p$ = sum of p averages for the laboratories divided by p
s_A^2 (pooled) = $\Sigma s_i^2/p$ = average of within-laboratory variances (see Note 54).
$s_{\bar{x}_A}^2$ = $[\Sigma\, \bar{x}_i^2 - p\,(\bar{x}_A)^2]/(p-1)$ = 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.

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.

8.2.1 Before proceeding with the analysis, it is necessary to investigate agreement of the data with the following two assumptions: (*a*) the variances are the same in different laboratories (homogeneity of variance), and (*b*) the results show the same pattern of change from one material to another in different laboratories (lack of interactions). These two aspects of the analysis are discussed in 8.2.2 and 8.2.3.

8.2.2 *Investigation of Agreement of Variances*—This method is based on the assumption that the within-laboratory variances in different laboratories (of which the s_i^2 in Table 3 and its variations, are estimates) are the same. This does not mean that the s_i^2 have to be very close together, since an individual variance can be about four times the average variance (for $p = 10$ and $n = 4$) when all the calculated variances are really estimates of 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.

8.2.2.1 Table 4 gives approximate values (upper 5 % level) for the ratio of the largest variance to the sum of the variances that should not be exceeded **(11)**.

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

TABLE 4 Approximate Values (Upper 5 % Level) for the Ratio of the Largest Variance to the Sum of the Variances

No. of Laboratories	No. of Replicates				
	2	3	4	5	6
5	0.8412	0.6838	0.5981	0.5441	0.5065
6	0.7808	0.6161	0.5321	0.4803	0.4447
7	0.7271	0.5612	0.4800	0.4307	0.3974
8	0.6798	0.5157	0.4377	0.3910	0.3595
9	0.6385	0.4775	0.4027	0.3584	0.3286
10	0.6020	0.4450	0.3733	0.3311	0.3029
11	0.5700[A]	0.4140[A]	0.3480[A]	0.3070[A]	0.2810[A]
12	0.5410	0.3924	0.3264	0.2880	0.2624
13	0.5140[A]	0.3630[A]	0.3080[A]	0.2690[A]	0.2470[A]
14	0.4920[A]	0.3450[A]	0.2910[A]	0.2530[A]	0.2320[A]
15	0.4709	0.3346	0.2758	0.2419	0.2195
20	0.3894	0.2705	0.2205	0.1921	0.1735
30	0.2929	0.1980	0.1593	0.1377	0.1237

[A] Values obtained by graphic interpolation.

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**).

8.2.2.3 Often the data from one laboratory may indicate a high or low variance compared to the others, and elimination of that laboratory from the analysis results in a set of data with similar variances for the remaining laboratories (see Appendix X1). If all the variances 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 there is doubt about eliminating a high or low variance.

8.2.3 *Interactions*—A common problem with test results obtained from an interlaboratory study is the presence of interactions between laboratories and materials. This means that the pattern of change of the results obtained on a given group of materials in one laboratory differs from the pattern obtained in another laboratory. In extreme cases, different laboratories may even fail to rate materials in the same order. The accepted statistical technique for finding significant interactions is an analysis of variance. A reasonably reliable method for checking to see if troublesome interactions may exist, however, is to make a plot of the averages obtained on the materials by each laboratory (see X1.3.5). These plots should show similar patterns of change from material to material for all laboratories. One laboratory may show a noticeably different pattern from the others and may be eliminated. However, if the patterns vary for more than one or two of the laboratories, the test method needs to be reinvestigated, and the causes of the interactions discovered and eliminated. The advice of a statistical consultant should be obtained.

8.3 *Within-Laboratory and Between-Laboratory Variances*—After the analysis described in 8.2 has been completed for all materials and investigations for homogeneity of variance and for interactions have been completed, the quantities indicated are assembled in Table 6. The averages in Column 2 are the overall averages for each material, which are arranged in increasing order of the magnitude of the average. The components of variance in Columns 3 and 4 are the pooled within-laboratory variance and the component of between-laboratory variance, respectively, for each material. The variances in Columns 5 and 6 are the pooled within-laboratory variances (same as Column 3) and the sum of the two components of variance (Column 3 plus Column 4), respectively (see Note 5).

NOTE 5—These within-laboratory and between-laboratory variances apply to single determinations in a laboratory, even though the data from which they are derived involve replicate measurements in a laboratory. Thus, precision statements based on these variances will apply to comparisons between two single measurements within a laboratory or between laboratories, respectively.

8.4 *Estimates of Precision*—The reason for listing the materials in increasing order of magnitude in Table 6 is to permit examination of the precision and how it varies with the level of the property measured, and thus to make a decision about the proper form of the precision statement. For this purpose, the quantities listed in Table 7 are calculated and entered as shown, still in increasing order of magnitude of the average. Column 2 in Table 7 is the same as Column 2 in Table 6. Columns 3 and 4 contain the square roots of the numbers in Columns 5 and 6 of Table 6. Columns 5 and 6 of Table 7 contain the corresponding coefficients of variation, expressed as percents, that is, the within-laboratory or between-laboratory standard deviation, respectively, divided by the corresponding average and multiplied by 100.

8.4.1 *Determination of Form of Precision Statement* —The appropriate form of a precision statement depends on the relationship between the average level of the property measured for the different materials and the within-laboratory and between-laboratory standard deviations. There are three main forms of the relationship that cover most of the cases which are pertinent to ASTM test methods: (*a*) cases in which the standard deviation is relatively constant over the range of materials; (*b*) cases in which the standard deviation has an approximately linear relationship with the average level and the coefficient of variation is relatively constant; and (*c*) cases where the materials fall into two or more distinct groups within which condition (*a*) or (*b*) holds approximately, and for each of which a characteristic precision can be determined. In most cases, the determination of which of these alternatives applies, or whether some more complicated situation exists can be determined for practical purposes by plotting the standard deviations and coefficients of variation against the average level. Two separate graphs, one for the two standard

TABLE 5 Approximate Values (Upper 5 % Level) for the Ratio of Highest to Lowest Variance

No. of Laboratories	No. of Replicates				
	2	3	4	5	6
5	[A]	202	51	25	16
6	[A]	266	62	30	19
7	[A]	333	73	34	21
8	[A]	403	84	38	23
9	[A]	475	94	41	25
10	[A]	550	104	45	26
11	[A]	626	114	48	28
12	[A]	704	124	51	30
13	[A]	790[B]	135[B]	54[B]	31[B]
14	[A]	885[B]	145[B]	57[B]	32[B]
15	[A]	995[B]	155[B]	59[B]	33[B]

[A] Although it may be possible to calculate this value, it is suggested that all values be included in the analysis when only 2 replicates are used.
[B] Value obtained by graphic extrapolation.

TABLE 6 Averages, Components of Variance, and Variances for All Materials

Material	Average[A]	Components of Variance		Variance	
		Within-Laboratory	Between-Laboratory	Within-Laboratory	Between-Laboratory
A					
B					
C					
D					
E					

[A] Listed in increasing order of magnitude

TABLE 7 Averages, Standard Deviations, and Coefficients of Variation for All Materials

Material	Average[A]	Standard Deviations		Coefficient of Variation	
		Within-Laboratory	Between-Laboratory	Within-Laboratory	Between-Laboratory
A					
B					
C					
D					
E					

[A] Listed in increasing order of magnitude

deviations and one for the two coefficients of variation, are usually adequate (see Note 6). If more sophisticated techniques are desired, they may be found in other references **(5, 13).** The appropriate measures of precision described in 8.4.2-8.4.5 become the indexes of precision as described in Practice C670.

NOTE 6—Usually, the same case should be applicable to both between-laboratory and within-laboratory precision. Sometimes, however, one of the two types of measures of precision is dependent on the level and the other is not. In situations like this, it may be possible to select a suitable compromise in order to have the two precision statements in the same form. The advice of a statistical consultant should be obtained.

8.4.2 *Constant Standard Deviation*—In this case the pooled within-laboratory standard deviation over all materials becomes the single-operator standard deviation or one-sigma limit (1s) and the pooled between-laboratory standard deviation becomes the multilaboratory standard deviation or one-sigma limit (1s) as described in Sections 3 and 4 of Practice C670. The pooled standard deviations are derived by adding Columns 5 and 6 of Table 6 for within-laboratory and between-laboratory estimates, respectively, dividing each of the two totals by *q*, and taking the square roots.

8.4.3 *Constant Coefficient of Variation*—In this case the average within-laboratory coefficient of variation becomes the single-operator, one-sigma limit in percent (1s %) and the average between-laboratory coefficient of variation becomes the multilaboratory one-sigma limit in percent (1s %) as described in 3.1.3 of Practice C670. Since it is not possible to pool coefficients of variation in the same manner as variances and standard deviations, the simple arithmetic averages of Columns 5 and 6 in Table 7 are used.

8.4.4 *Separate Groups with Constant Standard Deviation or Coefficient of Variation (see* Note 7*)*—In this case the single-operator and multilaboratory one-sigma limits or one-sigma limits in percent are calculated separately for each group in the same manner as described in 8.4.2 or 8.4.3 above. For each group, the range of average values over which the index of precision applies is supplied with the estimates. Refer to 6.2.2.2 and 6.2.2.3 of Practice C670.

NOTE 7—Situations of the type described in 8.4.4 and 8.4.5 are often indications that something is wrong with the experimental situation or the test method. If the standard deviation and coefficient of variation are so erratic that it is difficult to write an applicable precision statement without giving separate indexes of precision for each material tested in the interlaboratory program, this is very possibly an indication that the test method itself may be subject to erratic variations and may need to be restudied and revised. Also interactions or non-normal distributions may exist in the data. See X1.3.4 and X1.3.5 and Ref. **(8).** In cases of erratic precision, a precision statement in the test method may really be more misleading than helpful to persons trying to use or interpret the results of the test method. It may actually provide invalid information about what should be expected when the test method is used.

8.4.5 *Irregular or Nonlinear Relationship Between Standard Deviation, Coefficient of Variation, and Average Level (see* Note 7*)*—One way of dealing with situations that do not apply to 8.4.2-8.4.4 is to use the largest estimate of the standard deviation or coefficient of variation (whichever comes closest to being constant) and to use the abbreviation" max" after the indexes of precision (see section 6.2.2.1 of Practice C670). This practice should be discouraged because the resulting indexes of precision are certain to be more lenient than they should be. The maximum limit applies strictly to the level at which the maximum standard deviation or coefficient of variation occurred. Tests done at other levels, for which lower precision limits apply, will be judged on the basis of a wider tolerance than they should be. Also, as pointed out in 8.2.2, individual estimates of variance can vary widely from each other yet still be estimates of the same underlying variance, and it may often be that the pooled or averaged estimates are still the most appropriate ones to use, even if upon superficial examination, the individual variances appear to scatter rather wildly. It is again emphasized that the advice of a statistical consultant is needed here.

8.4.5.1 Cases where the standard deviation or coefficient of variation is a nonlinear function of the average level are dealt with in Ref. **(5).** Very often the amount of data, especially the number of laboratories and materials, is insufficient to establish the form

of such a relationship beyond question, and estimates of precision based on one of the cases already described will serve. In addition, the difficulty of writing a precision statement based on a nonlinear relation, that can be easily understood and applied by the user of a test method indicates that such statements should be avoided if possible.

APPENDIXES

(Nonmandatory Information)

X1. EXAMPLE OF ANALYSIS OF DATA FROM AN INTERLABORATORY STUDY

X1.1 *Introduction*—The following example is based on data from an interlaboratory study of tests of expansive cements conducted by ASTM Subcommittee C01.12 on Blended Cements.

X1.2 *Characteristics of the Study and Data*—The major test program was preceded by preliminary tests in three of the participating laboratories, each using a different kind of cement and the test procedures to be used in the major test program. The purpose of the preliminary tests was to determine whether there were any glaring deficiencies connected with the test procedures, and to alter the procedures, if necessary, before undertaking the major test program. This is the screening procedure referred to in 3.1.1 (see Note X1.1).

X1.2.1 In the main test program, eleven laboratories tested five cements for several properties, including restrained and unrestrained expansion of mortar, compressive strength at several ages of unrestrained and restrained modified 2-in. (50.8-mm) cubes, time of setting, false set, and restrained expansion and compressive strength of concrete.

X1.2.2 The data used in this analysis are the 3-day unrestrained compressive strengths of the 2-in. (50.8-mm) mortar cubes. In each laboratory, three rounds of mortar specimens were made with four of the cements, with three cube specimens per round. Four rounds were made with the fifth cement, but in this analysis only the data from the first three rounds reported were used. The average of the three cubes from a single round was used as an individual test result. Thus, the data used constitute a study with $p = 11$ laboratories; q = five materials; and n = three replicates (each an average of three specimens). Certain minor adjustments were made in the data in order to present a better illustration of the technique of analyzing the results of an interlaboratory study, but the data are essentially as reported by the laboratories.

X1.3 *Analysis of Data:*

NOTE X1.1—Since the discussion and tables following constitute directions for manipulating the numbers only, and what is of interest is the process of analysis itself, rather than the analyzed data, it is not deemed appropriate to include metric equivalents, and in the tables, no units at all are given. Units in both systems, however, are given in the finished precision statements, which are examples of what would be published as a result of the analysis. For those who are interested in converting the numbers, however, the following information is given:

(*1*) All the numbers given are in psi or $(psi)^2$.

(*2*) For individual values, averages, and sums of individual values,

psi × 6.9 = kPa and

(*3*) For sums of squares, variances, etc.

X1.3.1 *Step 1: Assembling the Data*—Table X1.1 shows the data sheet for one laboratory for all the tests. The individual cube values and averages for each round are given. In this laboratory, the individual results are given to the nearest 25 psi, which is the

TABLE X1.1 Data Sheet for Interlaboratory Test for Compressive Strength of 2 by 2-in. (50.8 by 50.8-mm) Mortar Cubes at 3 Days

Laboratory: ABC Testing Agency, Washington, DC						
Round	Specimen	Material				
		A	B	C	D	E
1	1	2950	3750	2750	1775	2025
	2	2800	3575	2825	1875	2075
	3	2825	3825	2800	1900	2100
	Avg.	2858	3717	2792	1850	2067
2	1	2875	3800	2775	2025	2225
	2	2875	4050	2900	2025	2275
	3	2850	3925	2725	2025	2225
	Avg.	2867	3925	2800	2020	2242
3	1	2875	3925	2625	1975	2075
	2	2875	3925	2775	1975	2100
	3	2975	4000	2650	1975	2125
	Avg.	2908	3950	2683	1980	2100

case with most of the compressive strength data from the study. Some of the laboratories, however, gave results of individual cube strengths to the nearest 1 psi. The testing machines used in testing 2-in. cubes generally are graduated to 50-lb intervals and can be estimated to half a graduation. The total load is divided by four to get pounds per square inch. Some laboratories give individual results to the nearest 25 psi and some to the nearest 50 psi. Others round the result of the division to the nearest 1 psi, which gives different results in the final digit depending on whether or not half divisions on the dial are estimated and how the results of the division are rounded. In addition, different methods were used in averaging the three individual cube measurements to get a test result.

X1.3.1.1 In setting up the instructions for an interlaboratory study, directions for handling details such as this should be spelled out, so that the laboratories use the same procedure. In the analysis presented here, the individual cube results were used as reported, and the averages of the three were recalculated and rounded to the nearest 1 psi in cases where the averages given by the laboratories had been rounded to the nearest 10 psi.

X1.3.1.2 Table X1.2 is a Summary Data Sheet for all laboratories and all materials. The individual entries are the averages of the three cubes per round. The arithmetic labor of an analysis such as this can often be reduced by coding the data, that is, by subtracting a constant from all the numbers. This treatment has no effect on the variances and standard deviations that form the essential parts of the estimates of precision. Averages are reduced by the amount of the constant subtracted and can be restored by merely adding it again. The practice is helpful in avoiding difficulties that sometimes arise from squaring and summing the squares of large numbers when a calculator that carries a limited number of digits is used. The data were not coded for the analysis given in this example, but they could have been. For Material A, with Laboratory 2 excluded (see X1.3.3.1), the subtraction of 2550 (or any number in the range 2526 to 2567) would have reduced all the data to three-digit numbers. The most appropriate number would have been different for the different materials.

TABLE X1.2 Summary Data Sheet for Interlaboratory Test Program for Compressive Strength of 2 by 2-in. (50.8 by 50.8-mm) Mortar Cubes at 3 Days

Laboratory		Material				
		A	B	C	D	E
1	a	2858	3717	2792	1850	2067
	b	2867	3925	2800	2025	2242
	c	2908	3950	2683	1975	2100
2	a	1813	2382	1737	1784	1314
	b	2388	3588	2475	1992	2113
	c	2625	3396	2321	1825	1650
3	a	3083	3750	2867	1757	2100
	b	3033	3600	2908	1758	2058
	c	3158	3667	2825	1803	1937
4	a	2783	3783	2608	1925	2175
	b	2692	3867	2800	1925	2125
	c	2817	3800	2600	1808	2042
5	a	3098	3742	2575	1675	2000
	b	3133	3617	2958	1983	2125
	c	3058	3392	2817	1875	2092
6	a	3263	4117	2850	2204	2529
	b	3259	4325	2917	2171	2596
	c	3300	4004	3025	2163	2525
7	a	3225	3983	2717	1867	2250
	b	3125	3817	2533	1875	2242
	c	2850	3550	2775	1925	2300
8	a	3267	4075	2950	2166	2358
	b	3525	4425	3317	2383	2442
	c	3158	4025	3075	2217	2200
9	a	3177	3920	2747	1978	2203
	b	3230	3933	2800	1983	2200
	c	3303	4000	2797	1973	2215
10	a	2840	3907	2425	1767	1942
	b	2858	3858	2158	1817	2000
	c	2667	3967	2600	1883	2067
11	a	2967	3775	2658	1850	1900
	b	2900	3842	2600	1865	1992
	c	3033	3758	2675	1867	2025