Designation: C670 – 13

StandardPractice for Preparing Precision and Bias Statements for Test Methods for Construction Materials[1]

This standard is issued under the fixed designation C670; 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.

This standard has been approved for use by agencies of the Department of Defense.

1. Scope*

1.1 The *Form and Style for ASTM Standards* requires that all test methods contain statements on precision and bias. Further, the precision statement is required to contain a statement on single-operator precision (repeatability) and a statement on multilaboratory precision (reproducibility). This practice provides guidance for preparing precision and bias statements that comply with these requirements. Discussion of the purpose and significance of precision and bias statements for users of test methods is also provided. Examples of precision statements that conform to this practice are included in Appendix X1. This practice supplements Practice E177 and has been developed to meet the needs of ASTM Committees dealing with construction materials.

NOTE 1—Although this practice is under the jurisdiction of Committee C09, the current version was developed jointly by Committees C01 and C09 and has subsequently been adopted for use by other committees dealing with construction materials.

1.2 This practice assumes that an interlaboratory study (ILS) has been completed in accordance with Practice C802 or Practice E691. The interlaboratory study provides the necessary statistical values to write the precision and bias statements.

1.3 The system of units for this practice is not specified. Dimensional quantities in the practice are presented only in examples of precision and bias statements.

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

2. Referenced Documents

2.1 *ASTM Standards:*[2]

C802 Practice for Conducting an Interlaboratory Test Program to Determine the Precision of Test Methods for Construction Materials

C1067 Practice for Conducting a Ruggedness Evaluation or Screening Program for Test Methods for Construction Materials

D6607 Practice for Inclusion of Precision Statement Variation in Specification Limits

E177 Practice for Use of the Terms Precision and Bias in ASTM Test Methods

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.

3.2 *Definitions of Terms Specific to This Standard:*[3]

3.2.1 *test determination, n*—the value of a characteristic of a single test specimen obtained by a specified test method.

3.2.1.1 *Discussion*—The term "replicate" is often used for a test determination.

3.2.2 *test result, n*—the value of a characteristic of a material obtained by carrying out a specified test method.

3.2.2.1 *Discussion*—A test result may be a single test determination or the average of a specified number of test determinations, or replicates (see 4.1 for additional discussion).

3.2.3 *identical test specimens, n*—test specimens selected at random and made from a single quantity or batch of material that is as homogeneous as possible.

[1] This practice is under the jurisdiction of ASTM Committee C09 on Concrete and Concrete Aggregates and is the direct responsibility of Subcommittee C09.94 on Evaluation of Data (Joint C09 and C01).

Current edition approved July 1, 2013. Published July 2013. Originally approved in 1971. Last previous edition approved in 2010 as C670-10. DOI: 10.1520/C0670-13.

[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.

[3] Terms are listed in order of hierarchy beginning with the basic concept.

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

3.2.3.1 *Discussion*—In interlaboratory studies of test methods for fresh cementitious mixtures, a practicable approach for obtaining identical tests specimens is to assemble technicians from different laboratories at one location and test specimens are made from the same batch of the fresh mixture. For interlaboratory studies of nondestructive test methods, the same test specimens can be circulated among participating laboratories, provided the characteristic of interest does not change during the time to complete the study.

3.2.4 *single-operator standard deviation, s_r, (or coefficient of variation, CV_r), n*—the standard deviation (or coefficient of variation) of test determinations obtained on identical test specimens by a single operator using the same apparatus in the same laboratory over a relatively short period of time.

3.2.4.1 *Discussion*—The single-operator standard deviation, or coefficient of variation, is the fundamental statistic underlying the single-operator indexes of precision. The single-operator standard deviation, or coefficient of variation, is an indication of the variability of a large group of test determinations by the same operator on the same material. This value is obtained from an interlaboratory study and is equal to the pooled standard deviation of test determinations obtained by the operators. The coefficient of variation (ratio of standard deviation to the average expressed as a percentage) is used if the standard deviation is proportional to the level of the characteristic being measured. The single-operator standard deviation, usually considered a property of the test method, will generally be lower than the multilaboratory standard deviation. In Practice E177, the single-operator standard deviation is referred to as the *repeatability standard deviation*, and the subscript *r* is used. In previous versions of Practice C670, the terms *one-sigma limit (1s)* or *one sigma limit in percent (1s%)* were used for the single-operator standard deviation or single-operator coefficient of variation, respectively. In some publications, the term *within-test standard deviation* (or *coefficient of variation*) has been used. The term *within-laboratory standard deviation* (or coefficient of variation) should not be used for this statistic (see 4.2.3).

3.2.5 *multilaboratory standard deviation, s_R (or coefficient of variation, CV_R), n*—the standard deviation or coefficient of variation of test results obtained with the same test method on identical test specimens in different laboratories with different operators using different equipment.

3.2.5.1 *Discussion*—The multilaboratory standard deviation, or coefficient of variation, is the fundamental statistic underlying the indexes of precision under multilaboratory conditions. The multilaboratory standard deviation is an indication of the variability of a group of test results obtained by different laboratories for identical test specimens. The multilaboratory standard deviation (or coefficient of variation) is usually greater than the single-operator standard deviation (or coefficient of variation), because different operators and different apparatus have been used in different laboratories for which the environments may have differed. In Practice E177, the multilaboratory standard deviation is referred to as the *reproducibility standard deviation* and the subscript *R* is used.

3.2.6 *difference limit (d2s or d2s%), n*—the difference between two test results that is expected to be exceeded with a probability of about 5 % in the normal and correct operation of the test method; used as an index of precision of the test method.

3.2.6.1 *Discussion*—The difference limit has been selected as the appropriate index of precision in most precision statements. A difference limit (d2s) indicates the maximum acceptable difference between two results obtained on identical test specimens (see 3.2.3.1) under the applicable system of causes (single-operator or multilaboratory conditions). The (d2s%) limit is the maximum acceptable difference between two test results expressed as a percentage of their average. These difference limits are calculated by multiplying the appropriate standard deviation (s_r or s_R) or coefficient of variation (CV_r or CV_R) by the factor 1.96 $\sqrt{2}$, which for the purpose of this Practice is taken to be equal to 2.8. In Practice E177, the terms *repeatability limit* and *reproducibility limit* are used for these difference limits under single-operator and multilaboratory conditions, respectively.

3.2.7 *acceptable range, n*—the difference between the largest and smallest of three or more test determinations or test results that is expected to be exceeded with a probability of about 5 % in the normal and correct operation of the test method; used as an index of precision of the test method, if applicable.

3.2.7.1 *Discussion*—This index is usually reported in precision statements of test methods that define a test result as the average of three or more determinations. Otherwise, the difference limit (d2s or d2s%) is used. See 4.3 for additional discussion on how to determine this index.

4. General Concepts

4.1 *Test Result*—The result of a test method may be a single test determination or the average of two or more test determinations (or replicates). The precision statement of a test method applies to a test result as defined in the test method and should state clearly this fact.

4.1.1 *Number of Test Determinations*—The number of test determinations required to obtain a test result by a test method must be taken into account when evaluating testing variations. The statistic used in evaluating single-operator precision is based usually on the standard deviation (or coefficient of variation) of single test determinations. The single-operator standard deviation (or coefficient of variation) may be used in evaluating the acceptable range of test determinations.

4.1.2 *Test Result Based on Averages of Determinations*—For test methods that define a test result as the average of two or more test determinations (or replicates), the fundamental statistic is still the standard deviation (or coefficient of variation) of single test determinations. The report of the analysis of the interlaboratory study (see 5.2) must include this statistic. The single-operator standard deviation of test determinations can be used to calculate the standard deviation of a test result that is the average of multiple determinations and thereby define the maximum acceptable difference between two test

results obtained by the same operator on identical test specimens. The precision statement may also include the maximum acceptable range of individual determinations that comprise the test result (see 4.3).

4.1.3 *Standard Deviation of an Average*—The standard deviation of the average of *n* test determinations obtained from identical specimens taken from the same population is equal to the standard deviation of the individual determinations divided by the square root of *n*. This relationship is valid, however, only if the determinations are obtained using identical specimens. It is not applicable to averages obtained on specimens made from different batches of cementitious mixtures as discussed in 4.2.3.

4.2 *Types of Precision*—A precision statement meeting the requirements of this practice normally contains two main elements: (*1*) single-operator precision, and (*2*) multilaboratory precision. For test methods that require test results on specimens made from more than one batch, the single-operator, multi-batch precision is also included.

4.2.1 *Single-Operator Precision*—The pooled, single-operator standard deviation (or coefficient of variation) of test determinations obtained from the interlaboratory study is the underlying statistic of the test method. This is used to calculate the greatest difference between two or more determinations that would be considered acceptable when properly conducted repetitive determinations are made on the same material by a competent operator. As discussed in 4.1.2, the single-operator standard deviation (or coefficient of variation) of test determinations is also used to calculate the greatest acceptable difference between test results defined as the average of two or more determinations. The single-operator precision provides a quantitative guide to acceptable performance by an operator. If two determinations or test results by the same operator differ by more than the difference limit, (d2s) or (d2s%), or if the range of more than two determinations or test results exceeds the values defined in 4.3, there is a high probability that an error has occurred and retests should be made.

NOTE 2—It is beyond the scope of this practice to describe in detail what action should be taken in all cases if two test results differ by more than the (d2s) or (d2s%) limits or the range of more than two determinations exceeds the maximum expected range. Such an occurrence is a warning that there may have been some error in the test procedure, or some departure from the prescribed conditions of the test on which the limits appearing in the test method are based; for example, faulty or misadjusted apparatus or improper conditions in the laboratory. In judging whether or not results are in error, information other than the difference between two test results is needed. Often a review of the circumstances under which the test results in question were obtained will reveal some reason for a departure. In this case, the data should be discarded and new test results obtained and evaluated separately. If no physical reason for a departure is found, retests should still be made, but the original tests should not be ignored. If the second set of results also differs by more than the applicable limit, the evidence is very strong that something is wrong or that a real difference exists between the specimens tested. If the second set produces a result within the limit, it may be taken as a valid test, but the operator or laboratory may then be suspected of producing erratic results, and a closer examination of the procedures would be in order. If knowledge about the test method in question indicates that certain actions may be appropriate in cases where deviant results occur, then such information should be included in the test method, but details of how this should be done will depend upon the particular test method.

4.2.2 *Multilaboratory Precision*—The multilaboratory standard deviation (or coefficient of variation) obtained from the interlaboratory study provides a measure of the greatest difference between two test determinations that would be considered acceptable when properly conducted tests are made by two different operators in different laboratories on portions of a material that are intended to be identical, or as nearly identical as possible. If results differ by more than the difference limit (d2s) or (d2s%), there is a high probability that one or both laboratories are in error or that a difference exists in the characteristics of the test specimens used for the tests. In such cases, retests should be made. If possible, newly drawn test specimens should be used for such retests.

4.2.2.1 If the test method calls for reporting the average of more than one test determination, multilaboratory precision is expressed as a maximum allowable difference between averages of such groups obtained by two laboratories (Note 3). In this case, the multilaboratory standard deviation derived from the interlaboratory study is based on the number of replications required to obtain a test result as defined by the test method.

NOTE 3—Example 5 in Appendix X1 shows an example of this situation. If a test result is based on tests of specimens made from different batches of the cementitious mixture, the consideration in 4.2.3 apply, and Example 6 provides an example of this situation.

4.2.3 *Single-Operator, Multi-Batch Precision*—Some test methods require reporting the averages of two or more determinations obtained on specimens from two or more batches made using the same materials. The single-operator, multi-batch standard deviation is a measure of the variation of the averages among the batches. This standard deviation will usually be greater than the value obtained by dividing the single-operator standard deviation by the square root of the number of determinations used to obtain the average test result for each batch. This is because the single-operator, multi-batch standard deviation includes the batch-to-batch variability. The precision statement for this type of test method will include three indexes of precision: (*1*) the single-operator precision, (*2*) the single-operator, multi-batch precision, and (*3*) the multilaboratory precision. In some test methods, the term *within-laboratory precision* has been used. The preferred term, however, is *single-operator, multi-batch precision* because this is more descriptive of the conditions under which results are obtained. The single-operator, multi-batch precision statement would indicate the acceptable range (or difference limit, if only two batches are involved) among batch averages. The advice of a statistical consultant should be sought in planning the interlaboratory study for this type of test method so that the necessary statistics can be determined.

4.2.4 *Other Measures of Precision*—The elements described in 4.2.1, 4.2.2, and 4.2.3 involve the main systems of causes that are of interest to users of test methods involving construction materials. In cases where other systems of causes apply (for example, *single-operator-apparatus*, *multi-day precision*; or *multi-operator, single-day-apparatus precision*), the appropriate statistics for those systems of causes need to be developed and the appropriate combination of modifiers given in Practice E177 should be used to describe those statistics. These should not, however, be taken as the fundamental

precision parameters for the test method. The advice of a statistical consultant should be sought in planning the interlaboratory study so that the correct statistics can be determined.

4.3 *Acceptable Range Among Results*—If the test method requires more than two test results, as so defined in the method, the difference between highest and lowest test results in the group must be compared to the maximum acceptable range for the applicable system of causes. The range among different numbers of test results in the group, including two, that would be expected to be exceeded with no more than about 5 % probability is obtained by multiplying the appropriate standard deviation or coefficient of variation by the corresponding factor from the second column of Table 1. If more than two test results are obtained, the index of precision for the difference between two results cannot be used as a criterion for judging acceptability of the differences between pairs of results selected from the group.

4.4 *Uses of Indexes of Precision*

4.4.1 *In Setting Specification Limits*—The indexes of precision described in this practice are applicable to test results obtained on identical test specimens and provide information on the inherent variability of the test method. In routine quality control or acceptance testing for a project, the variation of the test results will be affected by the inherent variability of the test method, the variability of the materials, and the variability associated with the sampling method. Specifiers need to consider these sources of variability in setting specification limits so as to control the producer's risk of rejection of a lot of acceptable material and the purchaser's risk of accepting a lot of deficient material.[4] Practice D6607 provides a methodology for setting specification limits that accounts for the inherent variability of the test method along with the material variability. The variability associated with the test method may be reduced by requiring a test result to be the average of two or more test determinations. A balance, however, needs to be achieved between the incremental cost of additional testing and the corresponding incremental reduction in uncertainty. Also, increasing the number of determinations to obtain a test result may have a minor affect on multilaboratory variability if the between-laboratory component of variance is greater than the single-operator variance. Because specification limits should be established with consideration of testing variability, it is not appropriate to consider the indexes of precision of the test method as tolerances to be added to statistically-derived specification limits for the purpose of judging acceptance or rejection of materials.

4.4.2 *For Qualifying an Operator*—As discussed in 4.2.1, indexes of single-operator precision are sometimes used as a basis for qualifying an operator. The assumption is that results that do not differ by more than the stated index are indicative of proper performance of the test. This assumption, however, is not necessarily correct. Uniform misunderstanding of instructions, incorrect specimen preparation, or maladjustments of equipment may produce consistent but erroneous test results. Whenever possible, tests conducted for the purpose of qualifying an operator should be made on materials for which the measured characteristic is known, so that bias as well as precision can be evaluated. Participation in proficiency sample programs is an effective way to evaluate operator performance among peers.

5. Basis for Precision Statement

5.1 In order to be valid, the indexes of precision to be included in the precision statement must be based on estimates of the precision of the test method obtained from a statistically designed interlaboratory study. Before proceeding with the interlaboratory study, the ruggedness of the test method should be investigated in accordance with Practice C1067. A ruggedness evaluation requires the involvement of a few laboratories and the use of several materials that encompass the range of the level of the characteristics to be measured by the test method. This evaluation will provide a preliminary estimate of single-operator precision and indicate whether tighter tolerances may be needed for key aspects of the test method. The interlaboratory study, on the other hand, must involve a sufficient number of laboratories, materials, and replicate measurements so that the results obtained provide reliable estimates of the precision of the test method (Note 4). The procedures described in this practice are based on the assumption that proper estimates of precision have been obtained. Practice C802 is a companion document that describes how to organize and conduct a suitable interlaboratory study and how to analyze the data to obtain the relevant estimates of precision.

NOTE 4—The requirement of "reliable estimates of the precision" presupposes an estimate obtained from a properly designed and executed interlaboratory series of tests involving at least 30 degrees of freedom for single-operator standard deviation and at least 10 laboratories. See Practice C802.

5.2 The *Form and Style for ASTM Standards* requires that data and details of the interlaboratory study used to determine precision and bias be filed as a research report at ASTM International Headquarters.

5.3 The ASTM International Interlaboratory Study Program (ILS) can support subcommittees in the development of precision statements by assisting in the design of an interlaboratory study, distribution of materials or test specimens, data analysis, and preparation of a draft research report.

[4] Philleo, R. E., "Establishing Specification Limits for Materials," *Cement, Concrete, and Aggregates*, CCAGDP, Vol. 1, No. 2, 1979, pp. 83-87.

TABLE 1 Maximum Acceptable Range of Test Results[A]

Number of Test Results	Multiplier of Standard Deviation or Coefficient of Varitation[B]
2	2.8
3	3.3
4	3.6
5	3.9
6	4.0
7	4.2
8	4.3
9	4.4
10	4.5

[A] A test result can be a single determination or the average of two or more determinations as defined in the test method.

[B] Values were obtained from Table A7 of "Order Statistics and Their Use in Testing and Estimation," Vol 1, by Leon Harter, Aerospace Research Laboratories, United States Air Force.

5.4 A subcommittee may wish to postpone the organization of the interlaboratory study until a new test method has been approved. In such cases, the precision statement of the new test method must include as a minimum the single-operator standard-deviation obtained in at least one laboratory. Preferably, the standard deviation should be obtained by using materials with different levels of the characteristic being measured. A ruggedness evaluation in accordance with Practice C1067 can be a source of data to develop a temporary precision statement. This temporary precision statement is permitted for five years at which time it needs to be replaced with a complete statement based on an interlaboratory study.

5.5 When an approved test method is being revised, the responsible subcommittee should determine whether the proposed change(s) to the test method will affect the validity of the precision statement in the existing standard. If the subcommittee believes the precision of the method may be affected by the revision, a new interlaboratory study should be conducted to provide data for updating the precision statement.

5.6 For some tests under the jurisdiction of Committees C01, C09, D04, and D18 there may be an extensive database of interlaboratory test data obtained from various proficiency sample programs. If such data are available for a particular test method, a precision statement can be prepared by carrying out the data analysis described in Practice C802 based upon a much larger population of data than can normally be assembled in an interlaboratory study. Care is needed, however, in evaluating the data because the requirement for identical test specimens may not be met by data from some proficiency sample programs. For example, participating laboratories may be shipped the dry ingredients to prepare specimens of cementitious mixtures for testing. The resulting specimens among the laboratories are not identical test specimens and the resulting multilaboratory precision includes an additional source of variation associated with making the test specimens. This needs to be mentioned in the precision statement.

6. Form of Precision Statement

6.1 *Background Information*

6.1.1 *Description of the Interlaboratory Study*—The *Form and Style for ASTM Standards* requires that the precision statement include a summary of the interlaboratory study that will permit the user of the test method to judge the reliability of the precision statement. This summary should include the number of laboratories, number of materials, range of material characteristics measured, and number of test determinations (replicate tests) for each material. The research report (see 5.2) should be referenced for the details of the interlaboratory study and the data analysis leading to the precision statement. This summary information should be provided in a note.

NOTE 5—Example 1 in Appendix X1 illustrates the wording that may be used in a note to summarize the interlaboratory study. The subcommittee should exercise its discretion in choosing the exact wording for the note as this will depend on the nature of the test method and the actual interlaboratory study.

6.1.2 *Information on Units*—Many precision indexes for test methods of construction materials are based on data obtained using the inch-pound version of a combined standard and these indexes have been converted to SI units. The following examples provide recommended wording for a note to the precision statement, if applicable, and how the conversion should be performed.

6.1.2.1 *Case 1*—Precision is stated in terms of a coefficient of variation. The precision indexes are non-dimensional and there would be no need for dual presentations. In this case, it is only necessary to state that the data were obtained in the inch-pound system.

Example 1:

The data used to develop the precision statement were obtained using an earlier inch-pound version of this test method.

6.1.2.2 *Case 2*—For a combined standard in which both systems of units are to be used separately:

Example 2:

The data used to develop the precision statement were obtained using the inch-pound version of this test method. The precision indexes shown in SI units are exact conversions of the values in inch-pound units.

6.1.2.3 *Case 3*—For a inch-pound standard that has been converted to an SI standard and the inch-pound units have been dropped:

Example 3:

The data used to develop the precision statement were obtained using the previous inch-pound version of this test method. The indicated precision indexes in SI units are exact conversions of the previous values obtained originally in inch-pound units.

6.2 *Manner of Expression*

6.2.1 If the interlaboratory study data, which are the basis for the precision statement, indicate that the standard deviation is essentially the same for all levels of the characteristic in question, the precision statement shall be expressed in the units of the measured characteristic.

6.2.2 If the standard deviation is essentially proportional to the average for different levels of the characteristic in question, that is, the coefficient of variation is essentially constant, the coefficient of variation and difference limit in percent (d2s%) shall be given. The coefficient of variation is determined as the ratio of the standard deviation to the average value of the results multiplied by100 %.

6.2.3 If neither of these conditions is met, the applicable precision limits for specific ranges of the measured characteristic shall be stated. See Example 3 in Appendix X1.

6.2.4 The phrase “are not expected to differ by more than” is used to introduce the applicable difference limits (d2s or d2s%). The intent of this wording is to recognize that the difference limits are expected to be exceeded in the long run with a probability of about 5 %. For introducing the acceptable range of more than two results, the corresponding phrase is “the range is not expected to exceed.”

6.2.5 The abbreviations (d2s) and (d2s%) are given in a footnote to the precision statement and reference is made to Practice C670 as shown in the examples in Appendix X1.

6.2.6 If the standard deviation varies erratically or the coefficient of variation is not constant over the range of the characteristic tested, the maximum value of the index of

precision shall be used. The word "maximum" shall be used in the first sentence of the precision statement and the abbreviation "max" shall be added as a subscript to the abbreviation for the difference limit in the footnote, that is, use $(d2s)_{max}$, or $(d2s\%)_{max}$. This form should be used rarely, and then only as a last resort. For additional discussion of this situation, see the section on Irregular or Nonlinear Relationship Between Standard Deviation, Coefficient of Variation and Average Level in Practice C802. Example 4 in Appendix X1 gives an example of this type of precision statement.

7. Bias Statement

7.1 *Introduction*—Bias is a systematic error inherent in the test method that contributes to the difference between the mean of the test results and an accepted reference or true value. In any test method, tolerances are placed on the accuracy of measuring equipment. Tests made with a given set equipment that has an error within the permitted tolerance will produce results with a small consistent bias, but that bias is not inherent to the test method and is not included in the bias statement for the test method.

7.2 *Estimating Bias*—There are two conditions that permit the bias of a test method to be estimated: (*1*) a standard reference sample with a known value of the characteristic in question has been tested by the test method, and (*2*) the test method has been applied to a sample that has been compounded in such a manner that the true value of the characteristic being measured is known, such as may be the case, for example, in a test for cement content of concrete. Judgment is required to determine whether a potential reference sample is suitable for the purpose. For example, a metal bar of accurately known physical properties might not be suitable for establishing the bias of a test for the corresponding concrete properties because the level of the values may differ by an order of magnitude. If it is possible to examine bias, it is necessary to determine whether there are enough data to determine statistically that the mean of the test results is significantly different from the true value. If there is a difference, an absolute measure of bias cannot be made, but confidence limits can be placed on the bias.

7.3 *If Bias Cannot be Estimated*—For most test methods there is no reference value available or the characteristic can be measured only by using that test method. In those cases, a bias statement based on one of the following may be used:

Example 1:

Bias—This test method has no bias because the values determined can be defined only in terms of the test method.

Example 2:

Bias—There is no accepted reference material suitable for determining the bias in this test method, therefore, no statement on bias is made.

Example 3:

Bias—No justifiable statement can be made on the bias of this test method because (insert here the reason).

7.4 *Procedure to Estimate Bias*—If it is possible to obtain data to determine if bias exists, a two-tailed t-test can be used in accordance with 7.4.1-7.4.4.

7.4.1 Obtain at least 30 results from separate specimens of the reference material or the material compounded to a known value of the characteristic in question. Calculate the quotient using Eq 1.

$$t = \frac{\bar{X} - X_r}{\frac{s}{\sqrt{N}}} \quad (1)$$

where:

X_r = reference value,
$\bar{X}$ = mean of the measured values,
s = standard deviation of the measured values, and
N = number of measured values on separate specimens.

7.4.2 The quotient obtained using Eq 1 has a *t*-distribution with *N*-1 degrees of freedom. Usually, the level of significance for the *t*-test, α, is taken to be 0.05; and because a two-tailed *t*-test is used, α/2 = 0.025 is used to determine the critical *t*-values. The null hypothesis that no bias exists is rejected if the value of t calculated by Eq 1 is less than $-t_{\alpha/2}$ or greater than $t_{\alpha/2}$. For α = 0.05 and *N*-1 = 29 degrees of freedom, the critical *t*-values for a two-tailed test are ±2.045, and the inequalities for rejecting the null hypothesis are: $t < -2.045$ or $t > 2.045$. Thus if the calculated value of t for 30 measurements falls between -2.045 and 2.045, there is no strong evidence to reject the null hypothesis and it may be concluded that there is no bias.

7.4.3 If the calculated value of t falls in the rejection region, it may be concluded that there is a bias in the test method and the 95 % confidence limits for the bias are:

$$\left(\bar{X} - X_r\right) \pm t_{\alpha/2}\frac{S}{\sqrt{N}} \quad (2)$$

7.4.4 In some cases, the bias may be a function of level of the characteristic being measured. If the differences $\bar{X} - X_r$ for different levels of X_r are statistically different from each other, the above procedure may be applied to each such level. A different bias may be applicable for different levels.

7.5 *Form of Bias Statement*

7.5.1 If a study for bias has been made, a bias statement based on one of the following examples may be used:

Example 1—No bias:

Bias—If measured results are compared with accepted reference values (or known values from accurately compounded specimens), the test method is found to have no bias.

Example 2—Bias exists:

Bias—If measured results are compared with accepted reference values (or known values from accurately compounded specimens), the bias of the test method is found with 95 % confidence to lie between 0.0062 and 0.0071.

Example 3—Bias depends on level:

Bias—If measured results are compared with accepted reference values (or known values from accurately compounded specimens), the bias of the test method is found with 95 % confidence to lie between -0.0004 and -0.0001 for results in the range of 6 to 10 and between -0.0006 and -0.0002 for results in the range of 10 to 15.