Designation: E1488 – 12ε1

An American National Standard

Standard Guide for Statistical Procedures to Use in Developing and Applying Test Methods[1]

This standard is issued under the fixed designation E1488; 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.

ε[1] NOTE—Editorial corrections were made to 3.1.25 in April 2013.

1. Scope

1.1 This guide identifies statistical procedures for use in developing new test methods or revising or evaluating existing test methods, or both.

1.2 This guide also cites statistical procedures especially useful in the application of test methods.

2. Referenced Documents

2.1 *ASTM Standards:*[2]

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
E1169 Practice for Conducting Ruggedness Tests
E1402 Guide for Sampling Design
E2282 Guide for Defining the Test Result of a Test Method
E2489 Practice for Statistical Analysis of One-Sample and Two-Sample Interlaboratory Proficiency Testing Programs
E2554 Practice for Estimating and Monitoring the Uncertainty of Test Results of a Test Method Using Control Chart Techniques
E2586 Practice for Calculating and Using Basic Statistics
E2587 Practice for Use of Control Charts in Statistical Process Control
E2655 Guide for Reporting Uncertainty of Test Results and Use of the Term Measurement Uncertainty in ASTM Test Methods

3. Terminology

3.1 *Definitions*—For a more extensive list of terms in E11 standards, see Terminology E456.

3.1.1 *bias, n*—the difference between the expectation of the test results and an accepted reference value. **E177**

3.1.1.1 *Discussion*—Statistical procedures include the sampling considerations or the experiment design for the collection of data, or both, and the numerical and graphical approaches to summarize and analyze the collected data.

3.1.2 *coefficient of variation, CV, n*—for a nonnegative characteristic, the ratio of the standard deviation to the mean for a population or sample. **E2586**

3.1.3 *component of variance, n*—a part of a total variance identified with a specified source of variability.

3.1.4 *control chart, n*—chart on which are plotted a statistical measure of a subgroup versus time of sampling along with limits based on the statistical distribution of that measure so as to indicate how much common, or chance, cause variation is inherent in the process or product. **E2587**

3.1.5 *observation, n*—the process of obtaining information regarding the presence or absence of an attribute of a test specimen, or of making a reading on a characteristic or dimension of a test specimen. **E2282**

3.1.6 *observed value, n*—the value obtained by making an observation. **E2282**

3.1.7 *precision, n*—the closeness of agreement between independent test results obtained under stipulated conditions. **E177**

3.1.8 *proficiency testing, n*—determination of laboratory testing performance by means of interlaboratory comparisons. **E2489**

3.1.9 *repeatability, n*—precision under repeatability conditions. **E177**

3.1.10 *repeatability conditions, n*—conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time. **E177**

[1] This guide is under the jurisdiction of ASTM Committee E11 on Quality and Statistics and is the direct responsibility of Subcommittee E11.20 on Test Method Evaluation and Quality Control.

Current edition approved Aug. 1, 2012. Published September 2012. Originally approved in 1992. Last previous edition approved in 2009 as E1488 – 09. DOI: 10.1520/E1488-12E01.

[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.

3.1.11 *repeatability limit r, n*—the value below which the absolute difference between two individual test results obtained under repeatability conditions may be expected to occur with a probability of approximately 0.95 (95 %). **E177**

3.1.12 *repeatability standard deviation, s_r, n*— the standard deviation of test results obtained under repeatability conditions. **E177**

3.1.13 *reproducibility, n*—precision under reproducibility conditions. **E177**

3.1.14 *reproducibility conditions, n*—conditions where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment. **E177**

3.1.15 *reproducibility limit, R, n*—the value below which the absolute difference between two test results obtained under reproducibility conditions may be expected to occur with a probability of approximately 0.95 (95 %). **E177**

3.1.16 *reproducibility standard deviation, s_R, n*—the standard deviation of test results obtained under reproducibility conditions. **E177**

3.1.17 *ruggedness, n*—insensitivity of a test method to departures from specified test or environmental conditions. **E1169**

3.1.18 *ruggedness test, n*—a planned experiment in which environmental factors or test conditions are deliberately varied in order to evaluate the effects of such variation. **E1169**

3.1.19 *standard deviation, n—of a population*, σ, the square root of the average or expected value of the squared deviation of a variable from its mean – *of a sample* $\bar{x}$, the square root of the sum of the squared deviations of the observed values in the sample divided by the sample size minus 1. **E2586**

3.1.20 *state of statistical control, n*—process condition when only common causes are operating on the process. **E2587**

3.1.21 *statistical procedures, n*—the organized techniques and methods used to collect, analyze, and interpret data.

3.1.21.1 *Discussion*—Statistical procedures include the sampling considerations or the experiment design for the collection of data, or both, and the numerical and graphical approaches to summarize and analyze the collected data.

3.1.22 *test determination, n*—the value of a characteristic or dimension of a single test specimen derived from one or more observed values. **E2282**

3.1.23 *test method, n*—a definitive procedure that produces a test result. **E2282**

3.1.24 *test observation, n*—see **observation**. **E2282**

3.1.25 *test result, n*—the value of a characteristic obtained by carrying out a specified test method. **E2282**

4. Significance and Use

4.1 The creation of a standardized test method generally follows a series of steps from inception to approval and ongoing use. In all such stages there are questions of how well the test method performs.

4.1.1 Assessments of a new or existing test method generally involve statistical planning and analysis. This standard recommends what approaches may be taken and indicates which standards may be used to perform such assessments.

4.2 This standard introduces a series of phases which are recommended to be considered during the life cycle of a test method as depicted in Fig. 1. These begin with a *design phase* where the standard is initially prepared. A *development phase* involves a variety of experiments that allow further refinement and understanding of how the test method performs within a laboratory. In an *evaluation phase* the test method is then examined by way of interlaboratory studies resulting in precision and bias statistics which are published in the standard. Finally, the test method is subject to a *monitoring phase*.

4.3 All ASTM test methods are required to include statements on precision and bias.[3]

4.4 Since ASTM began to require all test methods to have precision and bias statements that are based on interlaboratory test methods, there has been increased concern regarding what statistical experiments and procedures to use during the development of the test methods. Although there exists a wide range of statistical procedures, there is a small group of generally accepted techniques that are beneficial to follow. This guide is designed to provide a brief overview of these procedures and to suggest an appropriate sequence of carrying out these procedures.

4.5 Statistical procedures often result in interpretations that are not absolutes. Sometimes the information obtained may be inadequate or incomplete, which may lead to additional questions and the need for further experimentation. Information outside the data is also important in establishing standards and in the interpretation of numerical results.

5. Summary of Guide

5.1 Outlined below is a suggested sequence of four phases useful in the development of a test method. A flowchart is provided in Fig. 1. Such a sequence of analyses may need to be modified in specific situations. The assistance of a qualified statistician is recommended at each review phase.

5.2 *Design Phase:*

5.2.1 This phase includes the formalization of the scope and the significance and use sections. It may include determining the purpose and describing a general approach to the test method but usually does not involve statistical studies.

5.2.2 Studies may be conducted to evaluate the basic performance of the method. The draft test method is prepared and sampling requirements and the test result (see Guide E2282) are clearly defined.

5.2.3 A flow chart is extremely valuable to identify the sequence of operations involved in a test method, for example, the sampling steps required to obtain the test specimens, definition of the test determination, how a test result is to be computed, and running the tests on the specimens.

5.3 *Development Phase:*

[3] See the Form and Style Manual for ASTM Standards that specifies, when possible, precision statements shall be estimated based on the results of an interlaboratory test program.

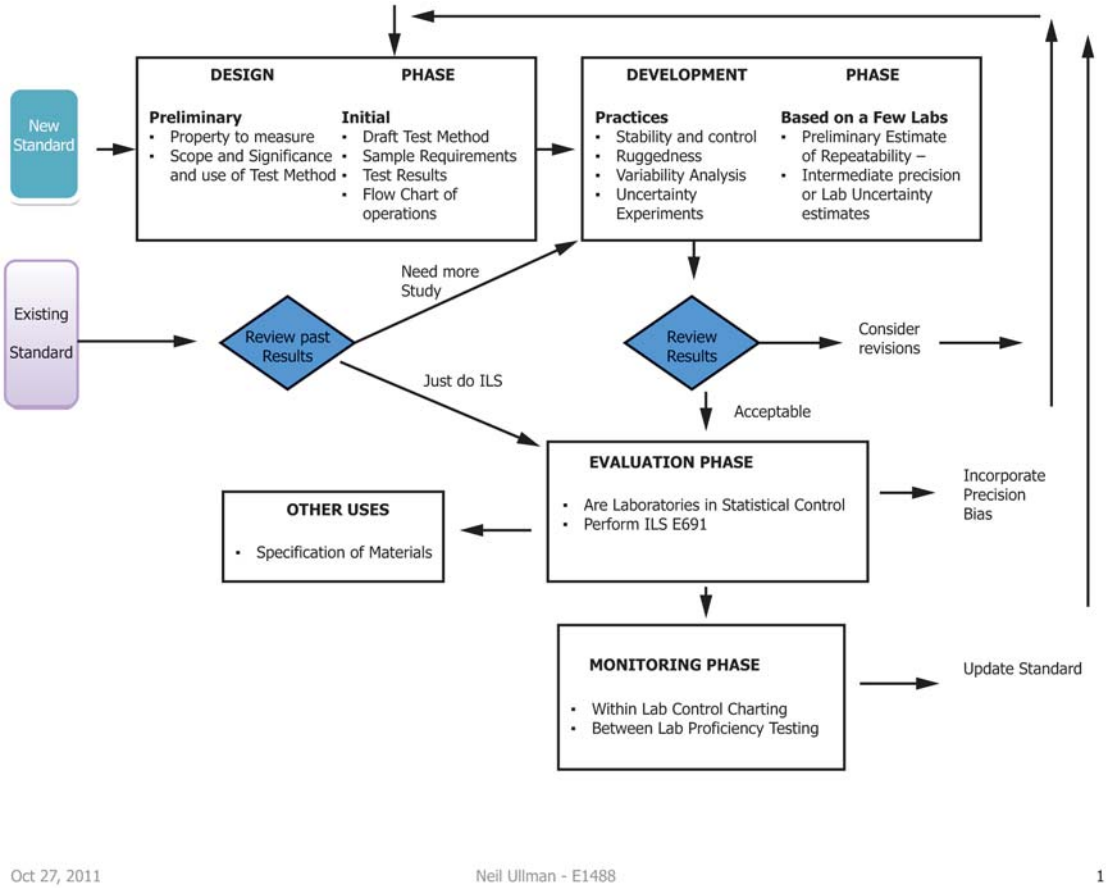

FIG. 1 Sequence of Steps

5.3.1 The test method is examined for such concerns as its stability, ruggedness, statistical control and the contributions to variability. The completion of this phase should result in preliminary estimates of precision and the identification and suggested ways to estimate potential contributors to uncertainty.

5.3.2 *Evaluation of Short Term Control of Test Method*—A test method must exhibit an ability to provide consistent results at least over short time periods. Preliminary studies or a pilot test should be conducted to evaluate the short term stability of the test method. A small series of repeated tests should be conducted.

5.3.3 *Analysis of Variability*—Statistically designed experiments conducted in one or two laboratories can be used to assess the relative magnitudes of different sources or potential contributors to variability of the test results. Such studies can provide estimates of intermediate measures of precision.

5.3.4 *Ruggedness Test*—A ruggedness test (see Practice E1169) is a statistically designed experiment that helps identify problems in running the test method, clarifies errors, and points out possible environmental conditions, which may adversely affect the test method or point out need for tightening requirements. The ruggedness test can assist in locating ways of reducing variability in the test method.

5.3.5 *Preliminary Estimates of Precision*—From the various studies conducted in accordance with 5.3.2 – 5.3.4, preliminary estimates of repeatability standard deviations should be developed and published in this test method. Until an interlaboratory study is performed, these estimates generally are considered to be provisional. Information on how a lab should develop uncertainty estimates should also be provided.

5.3.6 *Statistical Control*—A test method must show capability of performing in a consistent way over time. The use of control charts (see Guide E2655)[3] to monitor a proposed, or existing, test method over time is one recommended way to examine the controllability or stability of a test method. This statistical control should be demonstrated in one or two laboratories using homogeneous material (test specimen).

5.4 *Evaluation Phase:*

5.4.1 The test method is subjected to interlaboratory studies to provide estimates of within-laboratory repeatability and between-laboratory reproducibility. Additional information is supplied from proficiency studies when conducted.

5.4.2 *Interlaboratory Study (ILS)*—In accordance with ASTM Form and Style Manual, whenever feasible, an interlaboratory study must be conducted. This procedure will provide specific estimates of variation anticipated when using the test method.

5.4.3 Protocol for the ILS, Practice E691, provides a guide for developing the ILS for the test method. A first step is the writing of an ILS Protocol, which will set out what needs to be done before the test specimens (or test materials) are distributed to the participating laboratories.

5.4.4 *Precision Statements*—Using the estimates of variation obtained in the interlaboratory test, one may prepare precision statements using Practices E691 and E177 or equivalent procedures.

5.5 *Monitoring Phase:*

5.5.1 After a test method is approved and in use it is important to ensure that the published precision and bias statistics for the test method remain achievable and consistent over time or amongst different groups conducting the tests.

5.5.2 *Monitoring Within a Single Location*—It is important for any laboratory or organization that will use a particular test method over time that a means of monitoring to ensure the method results using quality control samples are stable and in control. Regular evaluation of the uncertainty (Practice E2554) or use of a control charting method (Practice E2587) are two ways to monitor the test method.

5.5.3 *Between Laboratory Comparisons*—Proficiency testing programs measure the typical variation amongst ordinary laboratories. The specific laboratories involved also obtain information about how well they perform compared to other laboratories.

6. Development of Test Method

6.1 Proposed standards that are under development should be treated in a formal manner following as many of the suggested procedures as possible. Standards that are already in existence as approved test methods or in general practice require periodic review that would include selected procedures.

6.2 *Under Development*—The development stage involves test methods that are in the preliminary stages during which equipment may not have been fully tested, practices are not agreed upon, and operators have yet to be adequately trained. Often this stage also applies to standards that have not yet been approved.

6.2.1 It is essential that tests for statistical control, ruggedness, and variability analyses be conducted prior to any interlaboratory test programs.

6.2.2 After all major environmental contributors have been identified, controlled, and incorporated into the test method, and after adequate standardized equipment is available, an interlaboratory test can be conducted. The interlaboratory test program must be completed prior to the first five-year review. The committee should strive to have interlaboratory results as soon as possible.

6.2.3 After evaluating data from ruggedness tests, variability analysis, or an interlaboratory test program, changes to the test method may be suggested.

6.2.4 If major changes are made to the test method, a repeat of the various steps is usually necessary. Precision and bias statements should reflect the most current version of the test method.

6.3 *Existing Standards*—These standards comprise test methods that are in common use for which standard equipment may exist and for which experienced operators have been trained and are available.

6.3.1 Control charting, ruggedness tests, and variability analyses will be useful, especially if they have not previously been conducted. Such tests may provide better information about variation and necessary tolerances than has previously been available.

6.3.2 If precision estimates have not been established through an actual interlaboratory test program, then such a program should be initiated.

7. Data and Sampling

7.1 *Sample Determination:*

7.1.1 The sampling section of a standard should indicate clearly what constitutes the primary sampling unit, how that sampling unit is further subdivided, and how multiple test values are designated. (See Guide E1402.)

7.1.2 In considering the implication of test results as they relate to the material, the test method should be clear as to whether the sampling method or the test is destructive or nondestructive.

7.1.3 The user of the test method should be aware of whether the standard calls for a random sample. In some standards, as for example in sampling from coils or rolls of material, samples may be taken only from certain portions of the material.

7.2 *Test Result Determination*—The procedure for determining a test result must be clear and unambiguous.

7.2.1 An observation leads to an observed value.

7.2.2 Several observed values may lead to a test determination. The observed values need not be the same type of measurements (for example, they may consist of three readings such as length, width, and mass).

7.2.3 Several Test determinations may lead to a test result, as by averaging three test determinations.

7.2.4 A test result is the consequence of a single execution of the entire test method.

7.3 *Type of Data*—The kind of data that results from the application of the test method determines the types of statistical analyses to be performed.

7.3.1 *Numerical versus Categorical/Attribute Data*—Most of the statistical procedures referred to in this guide deal with numerical data. Control charts are available for all types of data, but all interlaboratory test procedures currently in use depend on numerical data.

7.3.2 *"Normally" Distributed Data*—Most of the statistical procedures referred to in this guide consider that the unknown distribution of the test results can be modeled by a normal distribution.

8. Sources of Variability

8.1 *Experimental Realization of a Test Method:*

8.1.1 A realization of a test method refers to an actual application of the test method to produce a test result as specified by the test method. The realization involves an interpretation of the written document by a specific test operator , who uses a specific unit and version of the specified test apparatus, in the particular environment of his testing laboratory, to evaluate a specified number of test specimens of the material to be tested. Another realization of the test method may involve a change in one or more of the above emphasized experimental factors. The test result obtained by another

realization of the test method will usually differ from the test result obtained from the first realization. Even when none of the experimental factors is intentionally changed, small changes usually occur. The outcome of these changes may be seen as variability among the test results.

8.1.2 Each of the above experimental factors and all others, known and unknown, that can change the realization of a test method, are potential sources of variability in test results. Some of the more common factors are discussed in 8.2 – 8.6.

8.2 *Operator:*

8.2.1 *Clarity of Test Method*—Every effort must be made in preparing an ASTM standard test method to eliminate the possibility of serious differences in interpretation. One way to check clarity is to observe, without comment, a competent laboratory technician, not previously familiar with the method, apply the draft test method. If the technician has any difficulty, the draft most likely needs revision.

8.2.2 *Completeness of Test Method*—It is necessary that technicians, who are generally familiar with the test method or similar methods, not read anything into the instructions that is not explicitly stated therein. Therefore, to ensure minimum variability due to interpretation, procedural requirements must be complete.

8.2.3 *Differences in Operator Technique*—Even when operators have been trained by the same teacher or supervisor to give practically identical interpretations to the various steps of the test method, different operators (or even the same operator at different times) may still differ in such things as dexterity, reaction time, color sensitivity, interpolation in scale reading, and so forth. Unavoidable operator differences are thus one source of variability between test results. The test method should be designed and described to minimize the effects of these operator sources of variability.

8.3 *Apparatus:*

8.3.1 *Tolerances*—In order to avoid prohibitive costs, only necessary and reasonable manufacturing and maintenance tolerances can be specified. The variations allowed by these reasonable specification tolerances can be one source of variability between test results from different sets of test equipment.

8.3.2 *Calibration*—One of the variables associated with the equipment is its state of calibration, including traceability to national standards. The test method must provide guidance on the frequency of verification and of partial or complete recalibration; that is, for each test determination, each test result, once a day, week, etc., or as required in specified situations. In some test methods the calibration may also depend on the levels. Linearity and constancy of variation may depend on the range of levels.

8.4 *Environment:*

8.4.1 The properties of many materials are sensitive to temperature, humidity, atmospheric pressure, atmospheric contaminants, and other environmental factors. The test method usually specifies the standard environmental conditions for testing. However, since these factors cannot be controlled perfectly within and between laboratories, a test method must be able to cope with a reasonable amount of variability that inevitably occurs even though measurement and adjustment for the environmental variation have been used to obtain control. Thus, the method must be both robust to the differences between laboratories and require a sufficient number of test determinations to minimize the effect of within-laboratory variability.

8.5 *Sample (Test Specimens):*

8.5.1 A lot (or shipment) of material must be sampled. Since it is unlikely that the material is perfectly uniform, sampling variability is another source of variability among test results. In some applications, useful interpretation of test results may require the measurement of the sampling error.

8.5.2 In interlaboratory evaluation of test methods to determine testing variability, special attention is required in the selection of the material sample) in order to obtain test specimens that are as similar as possible. A small residual amount of material variability is almost always an inseparable component of any estimate of testing variability.

8.6 *Time:*

8.6.1 Each of the above sources of variability (operator performance, equipment, environment, test specimens) may change with time; for example, during a period when two or more test results are obtained. The longer the period, the less likely changes in these sources will remain random (that is, the more likely systematic effects will enter), thereby increasing the net change and the observed differences in test results. These differences will also depend on the degree of control exercised within the laboratory over the sources of variability. The material properties may also change over time. This is especially problematic when materials are stored or shipped. In conducting an interlaboratory evaluation of a test method, the time span over which the measurements are made should be kept as short as reasonably possible.

9. Preliminary Evaluation of Short Term Control

9.1 A test method must be capable of providing consistent results over short time periods. The first efforts at evaluating a test method should include repeating the method on the same or as close to the same materials under constant conditions over a short time period. This will provide some initial information about how close measurements can be repeated. This type of experiment should be repeated several times to determine how well the test method can perform at different time periods.

9.2 Since the tests may involve only a few sets of sample measurements, an experimental design model is the appropriate mode of evaluation of the results.

NOTE 1—We recommend that the Analysis of Means (ANOM)[4] procedure be utilized to determine how well the mean level remains at the same target level. This also permits an easy graphical and conceptual transition to a future control chart (as recommended in Section 12).

NOTE 2—Each sample will consist of small number of repeats. To determine if the variability remains consistent from sample to sample an Analysis of Ranges (ANOR) can be similarly conducted.[5]

[4] Ott, Schilling, and Neubauer, *Process Quality Control*, ASQ Quality Press.

[5] Ullman, "The Analysis of Means (ANOM) for Signal and Noise," *JQT*, Vol 21, No. 2, April 1989.

10. Analysis of Variation

10.1 Important contributions to variability must be ascertained. These sources may involve applying the test method at different laboratories, with different operators, over different days, with different apparatus, using different samples, and so on.

10.2 A statistically designed experiment for estimating "Components of Variance" is usually conducted to identify the relative contribution to the variation due to each of the factors under consideration.

10.3 A study of variability may be conducted in one or only a few laboratories because of the difficulty of managing the experiment (in contrast to an ILS).

10.4 A qualified statistician should be involved in organizing and working with the task group throughout the project.

11. Ruggedness Testing

11.1 The committee should attempt to identify all variables that are believed to have possible major influence on the precision or bias of the test method.

11.1.1 The ruggedness test usually is conducted in one or two laboratories with each "treatment" set at two levels. These levels are based on the conditions specified in the test method, and the low and high levels for each treatment are derived from the reasonable extremes that might be encountered in use. This test often should be one of the first procedures carried out and may need to be repeated when significant changes in the test method are made.

11.2 The test should include each such variable at levels as reasonably extreme as possible and likely to be encountered in practice. The ruggedness test then consists of an experiment conducted at one or two laboratories.

11.3 The statistical design is usually one in which a small set of the possible combinations of variables are tested at the selected two levels of each variable.

11.4 Practice E1169 is suggested to provide guidance in determining how to proceed.

12. State of Statistical Control

12.1 A test method, in order to be useful, must demonstrate long-term stability within a laboratory. The variation over long-term periods ideally should be no greater than the short-term variability.

12.2 A process is in a state of statistical control if the variations between the observed test results from it can be attributed to a constant system of chance causes. By "chance causes" is meant unknown factors, generally numerous and individually of small magnitude, that contribute to variation, but that are not readily detectable or identifiable.

12.3 The measurement process is in a state of statistical control when the test results obtained vary in a predictable manner, showing no unassignable trends, cycles, abrupt changes, excess scatter, or other unpredictable variations as determined by application of appropriate statistical methods. The ensurance of a state of statistical control is not a simple matter, but may be helped by the use of control charts (see Practices E2587 and E2554).

12.4 One measure of repeatability can be determined from the control chart for variability (range or standard deviation control chart). It is good laboratory practice to maintain a control chart for each test method in regular use.

12.5 The presence of outliers (Practice E178) may be evidence of a lack of statistical control in the production process or in the measurement process. It is quite proper to discard outliers for which a physical explanation is known. Discarding outliers in the measurement process on the basis of statistical evidence alone may yield biased results since one can truly measure the value of the property of interest only if the measurement process is in control. The presence of one or more outliers may indicate a weakness in the test method or its documentation.

12.6 Before a laboratory is to participate in any major comparative programs, it should demonstrate that the method exhibits such a state of statistical control within that laboratory.

12.6.1 If the set of test results to be considered in terms of statistical control is obtained in different laboratories, it may be possible to view the laboratories as a "sample" of all qualified laboratories that are likely to use the given test method, or as a set comprising a special category of such laboratories, and that the differences between the laboratories represent random variability. "Qualified" may mean, for example, laboratories that have used this test method for a year or more.

13. Precision and Bias

13.1 Although statistical procedures aid in the understanding of a test method, the primary purpose of including results of various studies, including an interlaboratory study, are to provide estimates of precision and bias.

13.2 *Precision*—A measure of the variability among test results conducted on the same material (or type of material).

13.2.1 The smallest variation occurs with replicated values obtained under the most reasonably similar conditions, usually within a single laboratory. This measure (when pooled over a set of participating laboratories of an ILS) is often referred to as repeatability of the test method.

NOTE 3—Some test methods may involve the taking of duplicate results. Variation among such duplicate observations usually will be smaller than between replicates. The estimate of precision that is of interest is between replicated test results.

13.2.2 The largest variation occurs with values obtained, for example, in different laboratories, which will involve different units of the specified equipment, and different operators. This measure is often referred to as reproducibility.

13.2.3 Variation amongst readings conducted within a single laboratory, with the same material but under different conditions and times, should be expected to fall between the repeatability and reproducibility values. This is often referred to as intermediate precision and depends on the conditions included in the estimates. This has also been described as "site precision."

13.2.4 Variation in test results also may be due to sampling of the material.

13.2.5 It is necessary for writers of the test methods to clarify for the user what types of variation may be encountered and how each source of variation should be controlled.

13.3 *Bias*—Bias refers to the difference between a population mean of the measurements or test results and an accepted reference or true value.

13.3.1 If no standard reference material exists and no such material can be prepared, then no estimate of bias can be determined. In such cases, all that is required is a statement saying that no bias estimate can be obtained.

14. Preliminary Estimates of Precision

14.1 Prior to the committee completing an interlaboratory study, results of experiments conducted in individual laboratories should be included in the standard. Studies such as pilot experiments, ruggedness tests, variability analyses, and control charts all can provide preliminary estimates of precision that can be obtained in individual typical laboratories.

14.2 Specific information on the type of materials, test conditions, and the number of laboratories and sets of repeated measurements should accompany the resulting estimates of precision.

15. Uncertainty

15.1 Users of many test methods are being required to prepare estimates of uncertainty. This is especially true for laboratories undergoing accreditation.

15.2 The standards developers are not expected to provide numerical values that will satisfy uncertainty estimation for any particular laboratory. The studies described in the previous sections may give guidance on the possible results and types of studies, but every laboratory must undertake its own studies.

15.3 Practice E2554 provides a recommended practice for determining the uncertainty of a particular test method within a single laboratory. Practice E2554 may then be used for ongoing monitoring of the uncertainty within the laboratory.

15.4 The test method developers should carefully evaluate the method and describe the procedures that a laboratory or other user should undertake to estimate the uncertainty of the measurements in their laboratory. Presentation of lists of factors to consider, identification of sources of variability, and possible level of effect are appropriate.

15.5 It is neither appropriate for, nor the responsibility of the test method to provide values of uncertainty that a user should use as their estimate of uncertainty.

NOTE 4—The methods described here are only of the Type A uncertainty. Guide E2655 discusses this as well as the process of combining Type A with Type B evaluations.

16. Interlaboratory Tests

16.1 *Purpose of the Study:*

16.1.1 The first objective is to obtain measures of how well the standard operates in a typical laboratory. The standard deviations obtained in each of the laboratories are averaged to give a measure called repeatability standard deviation that provides a guide to the user on how well different instruments or laboratory setups function on various materials (how repeatable the test results are in single laboratories). Separate estimates may be needed for different materials.

16.1.2 The second objective is to obtain measures of how well the standard operates among different laboratories (reproducibility of the test method).

16.1.3 In some cases a committee may be interested in investigating other specific types of variation. For example, the committee may consider it is useful to know how much variation is associated with day to day effects, with operator to operator effects, or for different calibration times. These sources of variation are often better investigated in one or a few laboratories.

16.1.4 The estimates of variability that are obtained are strictly for guidance purposes in assessing the general performance of the test method.

16.2 *Standard to Use:*

16.2.1 In those cases where only the within-laboratory repeatability (13.2.1) and between-laboratory reproducibility (13.2.2) are of interest, the use of Practice E691 is preferred.

16.2.2 When statistical designs more complex than prescribed in Practice E691 are used, the study should only be conducted with the assistance of a trained statistician. A statistician also may need to be consulted to help interpret results from a Practice E691 study.

16.3 *Range of Materials:*

16.3.1 The wider the range of material types, sizes, or compositions utilized in the interlaboratory study, the more useful will be the overall results.

16.3.2 Sometimes, regular additional interlaboratory studies are conducted to extend the range of materials. For example, a general test method measuring tensile strength might be further evaluated by an interlaboratory test procedure conducted by a special materials committee. Other committees may conduct periodic interlaboratory studies but add new materials or different test levels of the material.

16.4 *Sample Size:*

16.4.1 The first consideration should be toward having as many laboratories as possible. It is often difficult to obtain a large number of cooperating laboratories. When there are many laboratories, however, the number of tests per laboratory may often be reduced.

16.4.2 The number of types of materials or the range of levels, sizes, compositions, and so on should be the second consideration.

16.4.3 At least two tests per laboratory must be conducted within each laboratory. A minimum of three tests are recommended by many statisticians, although Practice E691 recommends two to four tests. The inclusion of more materials would be the preferred use of any additional test per laboratory. The ultimate goal of finding estimates of repeatability is accomplished by averaging the variability of sets of tests in many different laboratories. A possible exception to the rule of few tests per laboratory may occur when the execution of the test method is quick and simple, and the sample units are easy to obtain and are inexpensive.

17. Using the Estimates of Standard Deviation

17.1 *Precision Statements:*

17.1.1 Precision statements are to inform the committee and the ultimate user of the test method how close or far apart different test results may occur, or may be considered as not unusual.

17.1.2 Guidance for preparing precision statements is found in Practice E177.

17.2 *Comparison of Repeatability and Reproducibility:*

17.2.1 These two measures of variability obtained for a test method should be compared. The within-laboratory repeatability estimate is usually smaller than the between laboratory reproducibility estimate.

17.2.2 If the repeatability and reproducibility are similar in magnitude, it may be concluded that the test method has good stability between laboratories and that test results can be readily compared from one laboratory to another.

17.2.3 If the repeatability and reproducibility are quite different, the committee should consider reexamining the test method to determine the cause of a wide variation among laboratories. One implication is that the test method performs well at one place (at one time with a given set of equipment), but that different machines, different operators, and different laboratory conditions and equipment, at different times, may lead to quite different results.

17.3 *Coefficient of Variation (CV) versus Standard Deviation:*

17.3.1 If the standard deviations are similar over the range of levels of the measurement (from low to high), then only standard deviations should be reported.

17.3.2 The coefficient of variation may be useful when the standard deviation is a linear function of the average levels of the materials used. Note, however, that the CV covers up information—the magnitudes of both the average and the standard deviation are lost when reporting this ratio.

17.4 *Additional Considerations:*

17.4.1 The number of test determinations required for a test result may be established based on the estimates of within-laboratory repeatability and the precision desired for individual test results.

17.4.2 If multiple determinations are used and good precision is obtained, it may be possible to reduce the number of determinations. The guidance of an experienced statistician is desirable here.

17.4.3 If the repeatability standard deviation is found to be too large for intended purposes, then one consideration may be to increase the number of test determinations that are included in a test result.

17.4.4 Material specifications may call out the number of test results to be obtained for a particular material. Precision estimates obtained through ILSs and associated studies can assist in determining the appropriate number of such test results to be conducted.

18. Proficiency Testing

18.1 Proficiency testing is the use of interlaboratory test comparisons to determine the performance of individual laboratories for specific tests and to monitor the consistency and comparability of a laboratory's test data (see Practic E2489).

18.2 Repeatability and reproducibility precision data from previous interlaboratory studies should be used to establish initial guidelines for acceptable performance.

18.3 The results of the proficiency test should be provided to the subcommittee responsible for maintaining the standard test method. Summaries of the repeatability and reproducibility obtained during the proficiency program should be included in future revisions. Updates should be added to assist the observing trends in improvement to the test precision.

19. Reporting Statistical Results

19.1 Summaries of the results of all statistical studies should be included in an annex (mandatory).

19.2 For ASTM Standards, research reports corresponding to the studies should be prepared in accordance with the Form and Style of ASTM Standards and filed at ASTM headquarters.

19.3 Sections should be included in the standard to address at least the following:

19.3.1 *Uncertainty*—General information about how to develop laboratory uncertainty values.

19.3.2 *Precision*—Results of preliminary precision studies and interlaboratory programs as they are conducted.

19.3.3 *Bias*—This will depend on the availability of reference materials or values.

20. Keywords

20.1 interlaboratory study; precision; standards development; statistical control; statistical procedure; test method; test method development; uncertainty; variability