Designation: E 1488 – 02ε1

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 E 1488; 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—The introduction was deleted editorially in March 2003.

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:*

E 29 Practice for Using Significant Digits in Test Data to Determine Conformance with Specifications[2]

E 105 Practice for Probability Sampling of Materials[2]

E 122 Practice for Choice of Sample Size to Estimate a Measure of Quality for a Lot or Process[2]

E 141 Practice for Acceptance of Evidence Based on the Results of Probability Sampling[2]

E 177 Practice for Use of the Terms Precision and Bias in ASTM Test Methods[2]

E 178 Practice for Dealing with Outlying Observations[2]

E 456 Terminology Relating to Quality and Statistics[2]

E 691 Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method[2]

E 1169 Guide for Conducting Ruggedness Tests[2]

E 1301 Guide for Proficiency Testing by Interlaboratory Comparisons[2]

E 1325 Terminology Relating to Design of Experiments[2]

E 1402 Terminology Relating to Sampling[2]

2.2 *ISO Standards:*

ISO 5725 Accuracy (Trueness and Precision) of Measurement Methods and Results[3]

ISO 17025 General Requirements for the Competence of Testing and Calibration Laboratories[3]

ISO Guide to the Expression of Uncertainty in Measurement[3]

3. Terminology

3.1 *Definitions:*

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

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.

3.2 For all other formal definitions of statistical terms, see Terminology E 456.

4. Significance and Use

4.1 All ASTM test methods are required to include statements on precision and bias.[4]

4.2 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 very beneficial to follow. This document is designed to provide a brief overview of these procedures and to suggest an appropriate sequence of carrying out these procedures.

4.3 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.

[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 Oct. 10, 2002. Published December 2002. Originally published as E 1488 – 92. Last previous edition E 1488 – 96.

[2] *Annual Book of ASTM Standards*, Vol 14.02.

[3] Available from American National Standards Institute, 11 W. 42nd St., 13th Floor, New York, NY 10036.

[4] 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.

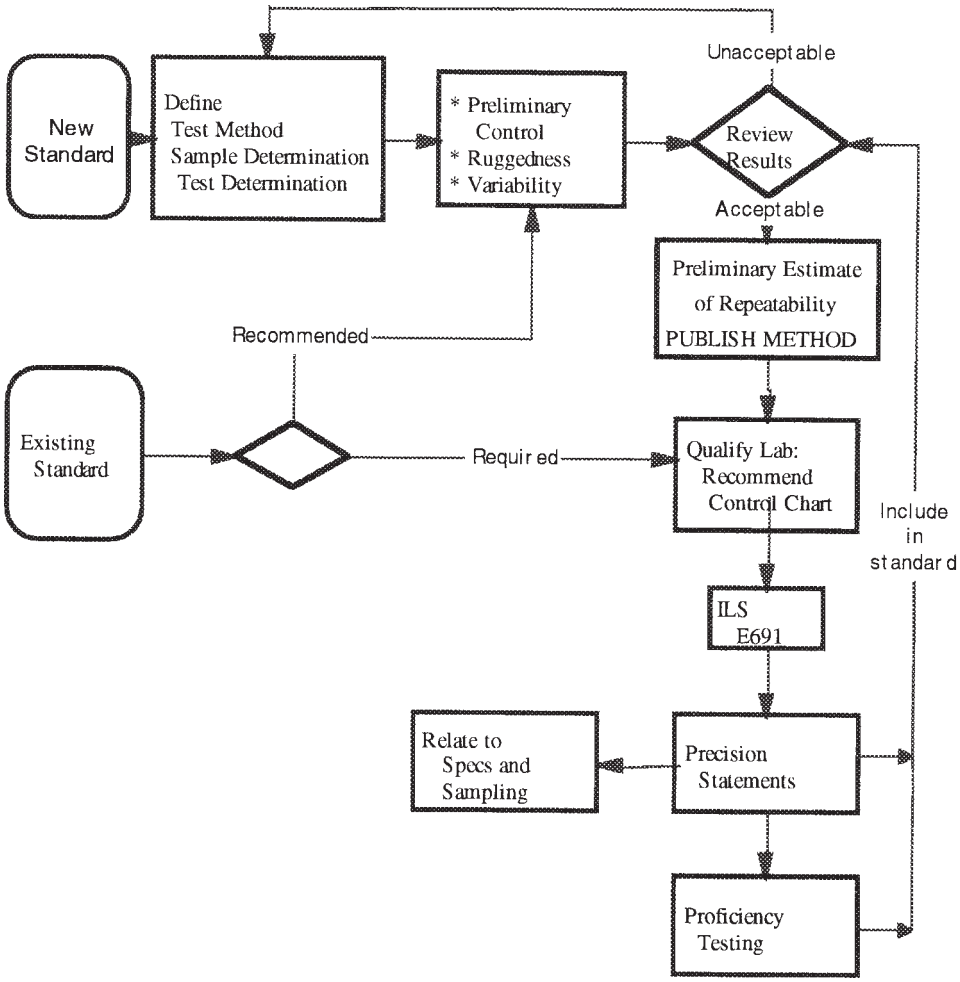

FIG. 1 Sequence of Steps

5. 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.

5.3 *Development Phase—*

5.3.1 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 are clearly defined.

5.3.2 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.

5.4 Validation Phase

5.4.1 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.

5.4.2 *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.

5.4.3 *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.

5.4.4 *Ruggedness Test*—A ruggedness test (see Guide E 1169) 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.

5.4.5 *Preliminary Estimates of Precision*—From the various studies conducted in accordance with 5.4.2–5.4.4, 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.

5.4.6 *Statistical Control*—A test method must show capability of performing in a consistent way over time. The use of control charts (see Manual 7)[4] 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).

5.5 *Evaluation Phase*:

5.5.1 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.

5.5.2 *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.

5.5.3 Protocol for the ILS, Practice E 691 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.

5.5.4 *Precision Statements*—Using the estimates of variation obtained in the interlaboratory test, one may prepare precision statements using Practices E 691 and E 177 or equivalent procedures.

6. Development of Test Method — Sampling and Test Result

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 5-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.

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 standard 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. Preliminary Evaluation of Short Term Control

8.1 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.

8.2 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) 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 11).

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.

NOTE 3—New standards are being developed to provide further guidance.

9. Analysis of Variation

9.1 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.

9.2 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.

9.3 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).

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

10. Ruggedness Testing

10.1 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.

10.1.1 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.

10.2 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.

10.3 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.

10.4 Guide E 1169 is suggested to provide guidance in determining how to proceed.

11. State of Statistical Control

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

11.2 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.

11.3 One strongly recommended method for determining if a process, or test method, is in statistical control is the use of control charts. Guidance for preparing and using control charts is given in Manual 7 (the revised STP 15D)[5]

11.4 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. Precision and Bias

12.1 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.

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

12.2.1 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 4—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.

12.2.2 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.

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

12.2.4 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.

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

12.3.1 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.

13. Preliminary Estimates of Precision

13.1 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

[5] Manual on Presentation of Data and Control Chart Analyses, ASTM Manual 7, 6th Edition, 1991, available from ASTM Headquarters. Request PCN:28-007089-34.

charts all can provide preliminary estimates of precision that can be obtained in individual typical laboratories.

13.2 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.

14. Uncertainty

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

NOTE 5—In many cases these uncertainty values are to be developed based on the ISO Guide to the Expression of Uncertainty (GUM) in Measurement. This is particularly necessary for laboratories undergoing accreditation following ISO 17025.

14.2 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.

14.3 The methods described here are only of the Type A uncertainty as described in the GUM. Combining these with Type B evaluations is discussed in the GUM.

14.4 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, and inclusion of budget templates (without numerical entries) are appropriate.

14.5 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.

15. Interlaboratory Tests

15.1 *Purpose of the Study*:

15.1.1 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.

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

15.1.3 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.

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

15.2 *Standard to Use*:

15.2.1 In those cases where only the within-laboratory repeatability (12.1.1) and between-laboratory reproducibility (12.1.2) are of interest, the use of Practice E 691 is preferred.

15.2.2 When statistical designs more complex than prescribed in Practice E 691 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 E 691 study.

15.3 *Range of Materials*:

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

15.3.2 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.

15.4 *Sample Size*:

15.4.1 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.

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

15.4.3 Two tests per laboratory must be conducted at all times. A minimum of three tests are recommended by many statisticians. A large number of repeated tests in each laboratory is unnecessary. 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.

16. Using the Estimates of Standard Deviation

16.1 *Precision Statements*:

16.1.1 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.

16.1.2 Guidance for preparing formal precision statements is found in Practice E 177.

16.2 *Comparison of Repeatability and Reproducibility*:

16.2.1 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.

16.2.2 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.

16.2.3 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.

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

16.3.1 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.

16.3.2 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.

16.4 *Additional Considerations*:

16.4.1 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.

16.4.2 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.

16.4.3 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.

16.4.4 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.

17. Proficiency Testing

17.1 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 Guide E 1301).

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

17.3 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.

18. Reporting Statistical Results

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

18.2 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.

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

18.3.1

Uncertainty. General information about how to develop laboratory uncertainty values.

18.3.2 Precision. Results of preliminary precision studies and interlaboratory programs as they are conducted.

18.3.3 Bias. This will depend on the availability of reference materials or values.

19. Keywords

19.1 statistical procedure