Designation: E 177 – 90a

AMERICAN SOCIETY FOR TESTING AND MATERIALS
1916 Race St. Philadelphia, Pa 19103

Standard Practice for Use of the Terms Precision and Bias in ASTM Test Methods[1]

This standard is issued under the fixed designation E 177; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 The purpose of this practice is to present concepts necessary to the understanding of the terms "precision" and "bias" as used in quantitative test methods. This practice also describes methods of expressing precision and bias and, in a final section, gives examples of how statements on precision and bias may be written for ASTM test methods.

NOTE 1—The term "accuracy", used in earlier editions of Practice E 177, embraces both precision and bias (see Section 20 and Note 4).

1.2 Informal descriptions of the concepts are introduced in the text as the concepts are developed, and appear in the following sections:

1.3 *This standard does not purport to address the safety problems 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:*
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]
2.2 *ANSI/ASQC Standard:*
A1-1978 Definitions, Symbols, Formulas and Tables for Control Charts[3]
2.3 *Other Documents:*
TAPPI Collaborative Reference Program, Reports 25 through 51, Aug. 1973 through Jan. 1978[4]
ASQC Glossary and Tables for Statistical Quality Control[3]

3. Terminology

3.1 The terminology defined in Terminology E 456 applies in all areas affected by this practice, except where modified by this practice.

3.2 This practice is specifically concerned with the development of statements on precision and bias for inclusion as descriptors of the performance of a test method. This application requires refinement of the Terminology E 456 definitions, as discussed herein.

3.3 The informal descriptions of concepts developed in this practice have been collected in Appendix X1, and have been arranged alphabetically for easy reference.

4. Significance and Use

4.1 Part A of the "Blue Book," *Form and Style for ASTM Standards*, requires that all test methods include statements of precision and bias. This practice discusses these two concepts and provides guidance for their use in statements about test methods.

4.2 *Precision*—A statement of precision allows potential users of a test method to assess in general terms the test method's usefulness with respect to variability in proposed applications. A statement on precision is not intended to contain values that can be exactly duplicated in every user's

[1] This practice is under the jurisdiction of ASTM Committee E-11 on Quality and Statistics and is the direct responsibility of Subcommittee E11.04 on Development and Evaluation of Test Methods.
Current edition approved June 29, 1990. Published August 1990. Originally published as E 177 – 61. Last previous edition E 177 – 90.

[2] *Annual Book of ASTM Standards*, Vol 14.02.
[3] Available from American Society for Quality Control, 230 West Wells St., Milwaukee, WI 53203.
[4] Available from the Technical Association of the Pulp and Paper Industry, Technology Park/Atlanta, P.O. Box 105113, Atlanta, GA 30348.

laboratory. Instead, the statement provides guidelines as to the kind of variability that can be expected between test results when the method is used in one or more reasonably competent laboratories. For a discussion of precision, see Section 18.

4.3 *Bias*—A statement on bias furnishes guidelines on the relationship between a set of typical test results produced by the test method under specific test conditions and a related set of accepted reference values (see Section 19).

GENERAL CONCEPTS

5. Test Method

5.1 Section 2 of the ASTM Regulations describes a *test method* as "a definitive procedure for the identification, measurement, and evaluation of one or more qualities, characteristics, or properties of a material, product, system or service that produces a test result."

5.2 In this practice only quantitative test methods that produce numerical results are considered. Also, the word "material" is used to mean material, product, system or service; the word "property" is used herein to mean that a quantitative test result can be obtained that describes a characteristic or a quality, or some other aspect of the material; and "test method" refers to both the document and the procedure described therein for obtaining a quantitative test result for one property. For a discussion of test result, see Section 9.

5.3 During its development, a test method should be subjected to a screening procedure and ruggedness test in order to establish the proper degree of control over factors that may affect the test results (see Guide E 1169).

NOTE 2—A screening procedure or ruggedness test is a procedure for investigation of the effects of variations in environmental and other pertinent factors on the test results obtained from a test in order to determine how control of such factors should be specified in the written description of the method. For example, temperature of the laboratory or of a heating device used in the test may have a significant effect in some cases and less in others. In a screening procedure, deliberate variations in temperature would be introduced to establish the limits of significant effect, (1, 2, 3).[5]

5.4 A well-written test method specifies control over such factors as the test equipment, the test environment, the qualifications of the operator (explicitly or implicitly), the preparation of test specimens, and the operating procedure for using the equipment in the test environment to measure some property of the test specimens. The test method will also specify the number of test specimens required and how measurements on them are to be combined to provide a test result (Section 9), and might also reference a sampling procedure appropriate for the intended use of the method.

5.5 It is necessary that the writers of the test method provide instructions or requirements for every known outside influence.

6. Measurement Terminology

6.1 The following terms have been used to describe both the measurement process and the partial or complete result of the process: measurement, observation, observed value, test, test determination, test result, and others. These terms have often been used loosely and interchangeably.

6.2 For clarity, it is necessary to select certain of these terms for specific use. However, the word "measurement" will be used in a generic sense to cover observation (or observed value), test determination and test result. The use of the word "test" by itself is discouraged.

6.3 A quantitative test method may have three distinct stages: (*1*) the direct measurement or observation of dimensions or properties; (*2*) the arithmetical combination of the observed values to obtain a single determination; and (*3*) the arithmetical combination of a number of determinations to obtain the test result of the test method. These three stages are explained and illustrated in Sections 7 to 9.

7. Observation

7.1 For the purposes of this practice, *observation* or *observed value* should be interpreted as the most elemental single reading or corrected reading obtained in the process of making a measurement. This statement is a narrower interpretation than is given in Terminology E 456 in that the latter applies to nonquantitative as well as quantitative test methods.

7.2 An observation may involve a direct reading (for example, a zero-adjusted micrometer reading of the thickness of a test strip at one position along the strip) or it may require the interpolation of the reading from a calibration curve.

8. Test Determination

8.1 For a quantitative test method, a *test determination* may be described as (*1*) the process of calculating from one or more observations a property of a single test specimen, or as (*2*) the value obtained from the process. Thus, the test determination may summarize or combine one or more observations.

8.2 *Examples:*

8.2.1 The measurement of the density of a test specimen may involve the separate observation of the mass and the volume of the specimen and the calculation of the ratio mass/volume. The density calculated from the ratio of one pair of mass and volume observations made on one specimen is a test determination.

8.2.2 The determination of the thickness of a test specimen strip may involve averaging micrometer caliper observations taken at several points along the strip.

9. Test Result

9.1 A *test result* is the value obtained by carrying out the complete protocol of the test method once, being either a single test determination or a specified combination of a number of test determinations.

9.2 In general, a test method describes not only the manner in which each test determination is to be made, but also the number of test determinations to be made and how these are to be combined to provide the test result.

9.3 *Examples:*

9.3.1 The test method on density might require that the mass and volume observations of a specimen be combined to give a test determination of density (8.2.1) and the test

[5] The boldface numbers in parentheses refer to a list of references at the end of this standard.

determination of each of five specimens be averaged to give a test result.

9.3.2 The test method for paper thickness may require that the determination of strip thickness in 8.2.2 be made on ten strips and that the ten test determinations be averaged to give the test result.

9.3.3 The test method for a tensile strength test of paper may specify that a tensile strength determination be performed on each of ten specimens and that the ten tensile test determinations be averaged to get the test result.

9.3.4 In chemical analyses a variety of situations may occur. Thus, in some cases, the method may call for the preparation of a single solution from a test unit, and measurement on three aliquots (specimens) of the solution made up to a specified volume. The average of the three analytical determinations would then be called the test result. In other cases of chemical analysis, the method may call for two individual test determinations, each one made on a different specimen with recalibration of the measuring instrument for each of the two determinations. The average of the two determinations would then be the test result.

9.3.5 In rubber testing, the method may describe not only the shape of the test specimen to be taken from a sheet of rubber, but also the preparation of the sheet, including compounding and curing. For example, one rubber test method specifies that four sheets be individually compounded and cured and three specimens tested from each sheet. The test result is then defined as the average of the four medians, each median being the middle determination, in the order of magnitude, of the three values obtained from a sheet.

9.3.6 Some test methods, such as those for analytical chemistry, involve calibration with known standard substances. The originally collected test determinations may be subjected to complex computational and statistical treatment prior to being converted into a test result. Such treatment might include separation of the analytical response for the substance of interest from the chromatographic absorption data, elimination or other treatment of outliers (see Practice E 178) in the data for the known standard substances, and preparation of a calibration curve to determine the test result.

9.4 Precision statements for ASTM test methods are applicable to comparisons between test results, not test determinations nor observations, unless specifically and clearly indicated otherwise (see Section 18).

SOURCES OF VARIABILITY

10. Experimental Realization of a Test Method

10.1 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.

10.2 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 Sections 11 through 15.

11. Operator

11.1 *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.

11.2 *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.

11.2.1 If requirements are not explicitly stated in the test method (see 5.5), they must be included in the instructions for the interlaboratory study (see Practice E 691).

11.3 *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.

12. Apparatus

12.1 *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.

12.2 *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.

13. Environment

13.1 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

though it may be stated quantitatively as the reciprocal of the standard deviation, precision is usually expressed as the standard deviation or some multiple of the standard deviation (see Section 27).

18.2 A measurement process may be described as precise when its test results are in a state of statistical control and their dispersion is small enough to meet the requirements of the testing situations in which the measurement process will be applied. The test results of two different processes expressed in the same units may be statistically compared as to precision, so that one process may be described as more (or less) precise than the other.

18.3 The precision of the measurement process will depend on what sources (Sections 10 through 15) of variability are purposely included and may also depend on the test level (see Section 21). An estimate of precision can be made and interpreted only if the experimental situation (prescribed like conditions) under which the test results are obtained is carefully described. There is no such thing as *the* precision of a test method; a separate precision statement will apply to each combination of sources of variability. The precision of a particular individual test result depends on the prescribed conditions for which it is considered a random selection. For example, will it be compared with other results obtained within the laboratory or with results obtained in other laboratories? No valid inferences on the precision of a test method or a test result can be drawn from an individual test result.

18.4 In order to minimize the effect of material variability in evaluating the precision of a test method, it is desirable to select a relatively uniform material for each of several test levels (magnitudes) chosen for the property being tested (see Practice E 691 for further information).

19. Bias

19.1 The *bias* of a measurement process is a generic concept related to a consistent or systematic difference between a set of test results from the process and an accepted reference value of the property being measured. The measuring process must be in a state of statistical control; otherwise the bias of the process has no meaning. In determining the bias, the effect of the imprecision is averaged out by taking the average of a very large set of test results. This average minus the accepted reference value is an estimate of the bias of the process (test method). Therefore, when an accepted reference value is not available, the bias cannot be established.

19.2 The magnitude of the bias may depend on what sources of variability are included, and may also vary with the test level and the nature of the material (see Section 21).

19.3 When evaluating the bias of a test method, it is usually advisable to minimize the effect of the random component of the measurement error by using at each test level the average of many (30 or more) test results, measured independently, for each of several relatively uniform materials, the reference values for which have been established by one of the alternatives in 16.1 (see 23.3 and 25.3).

19.4 If the bias of a test method is known, an adjustment for the bias may be incorporated in the test method in the section on calculation or in a calibration curve and then the method would be without bias.

19.5 The concept of bias may also be used to describe the systematic difference between two operators, two test sites (see 23.3), two seasons of the year, two test methods, and so forth. Such bias is not a direct property of the test method, unless one of the test sites or test methods provides the accepted reference value. The effect of such bias may be reflected in the measured reproducibility of the test method.

20. Accuracy

20.1 Accuracy is a generic concept of exactness related to the closeness of agreement between the average of one or more test results and an accepted reference value. Unless otherwise qualified, the use of the word "accuracy" by itself is to be interpreted as the accuracy of a test result. The *accuracy of a test result* is the closeness of agreement between the test result and the accepted reference value. It depends on both the imprecision and the bias of the test method.

20.2 There are two schools of thought on defining the *accuracy of a measuring process* **(5, 7).** In either case, the measurement process must be in a state of statistical control, otherwise the accuracy of the process has no meaning:

20.2.1 The closeness of agreement between the accepted reference value and the average of a large set of test results obtained by repeated applications of the test method, preferably in many laboratories.

20.2.2 The closeness of agreement between the accepted reference value and the individual test results **(8, 9).**

20.3 In 20.2.1 the imprecision is largely eliminated by the use of a large number of measurements and the accuracy of the measuring process depends only on bias. In 20.2.2 the imprecision is not eliminated and the accuracy depends on both bias and imprecision. In order to avoid confusion resulting from use of the word "accuracy", only the terms precision and bias should be used as descriptors of ASTM test methods.

21. Variation of Precision and Bias with Material

21.1 A test method is intended to cover a class of materials. Any one material within the class differs from any other in the following two basic ways: the level of the property that is being measured; and the matrix of the material. The matrix is the totality of all properties, other than the level of the property to be measured, that can have an effect on the measured value. Thus the precision and the bias of the test method may be functions of the property level and of the material matrix.

21.2 In some cases, a test method may be intended to be applied to more than one class of materials. If so, it may be advisable to provide separate statements of precision for each class (see 31.3).

22. Variation of Precision and Bias with Sources of Variability

22.1 The precision and bias of test results obtained by repeated applications of a test method depend upon what combinations of the sources of variability (Sections 10 through 15) affect the variability of the test results. For example, test results obtained by all possible operators within one laboratory using one set of test apparatus would have a bias based in part on that laboratory's apparatus and environment and a precision that would depend in part on the quality of training and supervision of operators in that laboratory. Many combinations of sources of variability are possible.

Some of the combinations used by ASTM committees are described in Sections 23 through 25.

COMBINATIONS OF SOURCES OF VARIABILITY (TYPES OF PRECISION AND BIAS)

23. Repeatability and Laboratory Bias

23.1 *Within-Laboratory Precision*—Information about a frequently used within-laboratory precision, sometimes called single-operator-day-apparatus precision, can be obtained from at least the three experimental situations described in 23.1.1 through 23.1.3, the last situation being most reliable; that is, the estimate of this precision is improved progressively by pooling additional information.

NOTE 3—If the test method requires a series of steps, the "single-operator-equipment" requirement means that for a particular step the same combination of operator and equipment is used for every test result and on every material. Thus one operator may prepare the test specimens, a second measure the dimensions and a third measure the breaking force. The "single-day" requirement means that the test results, at least for a particular material are obtained in the shortest practical period of time, whether this be a fraction of a day or several days.

23.1.1 *Precision From an Experiment Involving One Operator, Day and Apparatus*—A single, well-trained operator using one set of equipment obtains two or more test results in a short period of time during which neither the equipment nor the environment is likely to change appreciably. The variability is due primarily to small changes in equipment, calibration, reagents, environment, and operator's procedure, and possibly to some heterogeneity in the material tested. The last is kept small by use of test specimens from a reasonably uniform lot of material. The precision estimate for this operator, day, and equipment is determined from the variability of the test results. In this situation and the other experiments listed below, all potential sources of variability must be carefully controlled within the tolerances specified in the test method.

23.1.2 *Precision from Repeated Experiments Within a Laboratory*—In order to get an expression of precision that applies to any operator and day with a specific set of equipment at a given laboratory, the experiment of 23.1.1 must be repeated on different days by the same and different operators. Then the precision estimates, obtained as in 23.1.1, for each operator-day combination must be suitably combined or pooled to obtain an estimate of single-operator-day precision that applies to this laboratory and equipment. If the laboratory has several sets of equipment for this test method, the experiment may be enlarged to include tests on each set of equipment and the test results pooled in order to obtain an overall single-operator-day-equipment precision for that laboratory.

23.1.3 *Precision from Within-Laboratory Experiments in Several Laboratories*—In order to obtain an estimate of within-laboratory precision that is characteristic of the test method and may reasonably be applied to any laboratory, the whole within-laboratory experiment of 23.1.2 could be repeated in a number of laboratories. Alternatively, this desired broadly-applicable estimate may be obtained by pooling within-laboratory information from only one operator-day-equipment combination carried out in each of a number of laboratories. Although only one operator, one day, and one set of equipment are combined in each laboratory, the use of many laboratories, as in an interlaboratory study such as described in Practice E 691, provides an evaluation based on many operators, many days and many units of equipment. This abbreviated approach, only one operator-day-equipment combination in each laboratory, is based on the assumption that this estimate of within-laboratory precision does not change, or should not be expected to change, significantly from laboratory to laboratory. Consequently, this measure of precision can be treated as a characteristic of the test method. This pooled within-laboratory precision is called the **repeatability** of the test method.

23.2 *Repeatability Conditions*—While other conditions (Section 24) have sometimes been used for obtaining repeated test results in the determination of repeatability, the preferred conditions (illustrated above in 23.1 through 23.1.3) are those under which test results are obtained with the same test method in the same laboratory, by the same operator with the same equipment, in the shortest practical period of time, using test units or test specimens (see Practice E 691, 10.3) taken at random, from a single quantity of material that is as nearly homogeneous as possible. For meaning of "same operator, same equipment" and "shortest practical period of time," see Note 3 above.

23.3 *Repeatability*—The closeness of agreement between test results obtained under repeatability conditions.

23.4 *Bias of a Particular Laboratory*, relative to the other laboratories may be calculated by averaging test values obtained as described in 23.1.2 for that laboratory and comparing the result with the average of all test values obtained as described in 23.1.3. The *bias of the test method* may be calculated by comparing the latter average with the accepted reference value (Section 16), or it may be determined as described in 25.3. Once the bias is known, the method should be modified to correct for it (see 19.5).

24. Other Within-A-Single Laboratory Precisions

24.1 *Single-Operator-Apparatus, Multi-Day Precision*—A single operator using one set of equipment obtains replicate test results as in Section 23, but one on each of two or more days. Since the time interval is greater than in Section 23, there is a greater chance that the equipment (including its calibration) and the environment may change, and that the change will depend on the degree of control or supervision maintained by the laboratory over these factors. Therefore, the precision calculated in this between-day within-laboratory situation, may vary appreciably from laboratory to laboratory and often cannot be regarded as a universal parameter of the test method. While this multi-day precision has been called "repeatability" by some ASTM committees, it is better to reserve the term for the precision estimate described in 23.1.3, which is more likely to be an estimate of a universal characteristic of the test method. If information on multi-day precision is needed by a laboratory, it should be studied in that laboratory, since the estimate may vary widely from laboratory to laboratory.

24.2 *Multi-Operator, Single-Day-Apparatus Precision*—Each of several operators in one laboratory using the same set of equipment obtains a test result. Since the operator effect may depend on the degree of training and supervision exercised in the laboratory, the precision among test results

(between operators within laboratory) may vary widely from laboratory to laboratory, and therefore may not be regarded as a universal parameter of the test method (see Note in example in 31.7). If information on multi-operator precision is needed by a laboratory, it should be studied by that laboratory.

25. Reproducibility and Bias of the Test Method

25.1 *Between-Laboratory Precision*—Each of several laboratories, each with its own operator, apparatus, and environmental conditions, obtains a test result on randomly-selected specimens from the same reasonably-uniform sample of material. The variability of the test results may be used to calculate the between-laboratory precision, which, when based on a single test result from each laboratory, is also called the **reproducibility** of the test method. The laboratories being compared in order to obtain the between-laboratory reproducibility of the test method should be independent of each other. Independent means that the laboratories should not be under the same supervisory control, nor should they have worked together to resolve differences. The value found for the between-laboratory precision will depend on the choice of laboratories and the selection of operators and apparatus within each laboratory.

25.1.1 The precision within a single laboratory or facility having multiple test stations will depend largely on the degree of supervision provided. If information on this precision is required, the laboratory should run its own internal study, possibly using Practice E 691, with each station treated as a laboratory. The precision determined (that is, "between station reproducibility"), can be expected to be somewhat better than the reproducibility of the test method, depending on the degree of common supervision of the test stations.

25.2 *Reproducibility*, as used in 25.1 and 25.1.1, is a general term for a measure of precision applicable to the variability between single test results obtained in different laboratories using test specimens taken at random from a single sample of material. This use of the word "reproducibility" is narrower then that defined in Terminology E 456 because it assumes the simpler interlaboratory study of 23.1.3 and Practice E 691 where only one operator-day-apparatus combination is involved in each laboratory.

25.3 *Bias of Test Method*—The bias of the test method, for a specific material, may be calculated by comparing the average of all the test results obtained in 25.1 for that material with the accepted reference value (see Section 16) for that material. If no accepted reference value is available, bias cannot be calculated (however, see 29.2). For a valid determination of bias, the results of the test method must indicate a state of statistical control (see Section 17).

26. Range of Materials

26.1 The estimates of precision and bias described in Sections 23 through 25 are based on test results from a material at one level of the property of interest. The experiments should be extended to other related materials yielding test results at other test levels. Related materials are materials that may have similar matrixes of other properties (see Section 21) and are likely to be compared by means of the test method.

26.2 Precision and bias may be constants or simple functions of the test level or they may depend so appreciably on the matrix of other properties of the materials that the test method will have to be modified to take into account these other, possibly-interfering, properties before reasonable and consistent values for precision and bias can be obtained.

METHODS OF EXPRESSING PRECISION AND BIAS

27. Indexes of Precision

27.1 *General*—Precision may be stated in terms of an index consisting of some positive value, a. The index is expressed in the same units as those of the test result, or as a percent of the test result. The numerical value of a will be smaller when the individual test results from repeated applications of the test method are more closely grouped. The larger the index, the less precise the measurement process. A test method has a separate index of precision for each type of precision (see Sections 22 through 25) and this index may vary in a systematic way with the test level or it may vary from material to material even at the same test level.

27.2 *Basis*—The usual source of the index of precision is the sample estimate of the standard deviation, (denoted by the symbol s), of a random set of test results for that type of precision (for example, from an interlaboratory study such as Practice E 691), where standard deviation has its usual meaning (for example, see Terminology E 456). The number of test results in the set should be sufficiently large (at least 30) so that the sample standard deviation(s) computed from the randomly-selected set be a good approximation to the standard deviation of the population of all test results (denoted by the symbol σ) that could be obtained for that type of precision. See Practice E 691 for an example of the design of an interlaboratory study to determine within-laboratory and between-laboratory standard deviations, also called repeatability and reproducibility standard deviations.

27.3 *Possible Indexes of Precision:*

27.3.1 *Standard Deviations(s)*—See 27.2.

27.3.2 *"Two"-Standard Deviation Limits (2s)*—Approximately 95 % of individual test results from laboratories similar to those in an interlaboratory study can be expected to differ in absolute value from their average value by less than 1.960 s (about 2.0 s).

27.3.3 *Difference "Two"-Standard-Deviation Limit (d2s)*—Approximately 95 % of all pairs of test results from laboratories similar to those in the study can be expected to differ in absolute value by less than $1.960\sqrt{2}$ s (about $2.0\sqrt{2}$ s) = 2.77 s (or about 2.8 s). This index is also known as the 95 % limit on the difference between two test results. For the two cases described in Sections 23 and 25, these limits are known as the repeatability and reproducibility limits.

27.3.4 *Multiplier for 95 % Limit:*

27.3.4.1 The multiplier 1.960 or 2.0 used in 27.3.2 and 27.3.3 assumes an underlying normal distribution for the test results being compared. For methods in which the average of several test determinations is reported as a single test result, the assumption of normality is usually reasonable, even for skewed or bimodal distributions. When normality cannot be assumed, it is usually satisfactory to continue to use the multiplier 2.0 but recognize that the actual probability limit will differ somewhat from the nominal 95 % limit.

27.3.4.2 It may be thought that the use of the multiplier

1.960 (or approximately 2.0) in 27.3.2 and 27.3.3 requires that the sample standard deviation (s) be assumed to be equal to the population (or "true") standard deviation (σ). No within or between-laboratory study will yield a standard deviation (s) exactly equal to the "true" standard deviation (σ), and few will come close unless at least 30 laboratories are included in the study. No multiplier for s will ensure an actual limit of exactly 95 %. The use of the multiplier t, (Student's t), instead of the multiplier, 1.960 does not remedy the situation. In order to resolve this problem, a range of probabilities around 95 % must be accepted as defining the "95 % limit". For appropriate choices of the defining range, the multiplier 1.960 (or 2.0) may still be used. It has been shown that 1.960 is the best choice for achieving the desired (but approximate) 95 % coverage **(10)**. The multiplier is independent of the number of test results in the within-laboratory study or the number of laboratories in the study for between-laboratory precision. However, a within- or between-laboratory study must be of reasonably large size in order to provide reliable information on which to base a precision statement.

27.3.5 *Indexes in Percent*—In some instances (see 28.5) there may be some advantage in expressing the precision index as a percentage of the average test result; that is, percent coefficient of variation (CV %). The notation may then be (CV %), (2CV %), (d2CV %), etc.

27.3.6 *Other Indexes*—For some applications, limits based on 95 % probability are not adequate. Basic multipliers other than 1.960 (or about 2.0) may be used, yielding probabilities other than "approximately 0.95". As discussed below, however, the (d2s) = (2.8 s) and (d2CV %) = (2.8 CV %) indexes are recommended, unless there is a special need.

28. Preferred Indexes of Precision for ASTM Test Methods

28.1 *Preferred Types of Precision and Preferred Indexes*—The types of precision described in 23.1.3 and 25.1, namely, repeatability and reproducibility, are the preferred types of precision statements for ASTM test methods. The preferred index for each of these types is the 95 % limit on the difference between two test results (see 27.3.3), namely, 2.8 s or 2.8 CV %. Also the corresponding standard deviation (s) or percent coefficient of variation (CV %) shall be indicated.

28.2 *Recommended Terminology for Preferred Indexes:*

r = 95 % repeatability limit
R = 95 % reproducibility limit

or, to help prevent confusion between r and R, use:

r = 95 % repeatability limit (within a laboratory)
R = 95 % reproducibility limit (between laboratories)

Similarly, the recommended terminology for the corresponding standard deviations is:

s_r = repeatability standard deviation (within a laboratory)
s_R = reproducibility standard deviation (between laboratories)

and for the coefficients of variation:

CV $\%_r$ = repeatability coefficient of variation in percent (within a laboratory)
CV $\%_R$ = reproducibility coefficient of variation in percent (between laboratories)

where:

$r = 1.960\ \sqrt{2}\ s_r = 2.8\ s_r$ or $r = 1.960\ \sqrt{2}$ CV $\%_r = 2.8$ CV $\%_r$

$R = 1.960\ \sqrt{2}\ s_R = 2.8\ s_R$ or $R = 1.960\ \sqrt{2}$ CV $\%_R = 2.8$ CV $\%_R$

depending on how the indexes vary with the test level (see 28.5). For other than the preferred types, the more general terminology "95 % limit" may be used with a description of the sources of variability; for example:

95 % limit (operator-to-operator, within-laboratory) and similarly for the corresponding standard deviation:

operator-to-operator within-laboratory standard deviation.

28.3 *Whenever the general terms "repeatability" and "reproducibility" or the more specific terminology "repeatability limit" and "reproducibility limit" are stated with numerical values, users will have to assume that the 95 % limits are intended,* unless otherwise specified.

28.4 Quantitative estimates of repeatability and reproducibility may be obtained from an interlaboratory study conducted as directed in Practice E 691.

28.5 *Variation of Index With Test Level*—The choice between 2.8 s and 2.8 CV % and the form for the statement of the precision indexes depends upon how the indexes vary with the test level.

28.5.1 If a 2.8 s index is approximately constant throughout the test range, then the 2.8 s index is recommended. Express the index in the units of the measured property.

28.5.2 If a 2.8 s index is approximately proportional to the test level, then use the 2.8 CV % index. Express the index in percentage of the test level.

28.5.3 In either case, express the index as a single average (or pooled) number followed parenthetically by the actual range of the index values (highest and lowest) encountered in the interlaboratory study.

28.5.4 If a 2.8 s index is neither approximately constant nor approximately proportional to the test level, plot the index versus the test level to determine how they are related. If the index varies systematically with the test level, express the index by a combination of 2.8 s and 2.8 CV % (see example 31.3), by a simple formula, or by a plot. If the index varies in no systematic way with the level, but jumps from material to material (perhaps because some materials are inherently more variable than others), express the index by a table (see 31.6) or by a single compromise value selected by judgment. Carefully describe each material in the table. The jumping may be due to interfering properties in the material matrixes (Section 21) and the description may eventually allow identification of the cause.

29. Preferred Statements of Bias for ASTM Test Methods

29.1 Some information may be available concerning the bias or part of the bias of a test method as determined from an interlaboratory study (25.3 and 23.3) or from known effects of environmental or other deviations as determined in ruggedness tests (see 5.3). An adjustment for what is known about the bias can be incorporated in the calculations or calibration curves. The statement on bias should then state how this correction is provided for in the test method.

29.2 If the bias of a test method, or the uncorrected balance of the bias, is not known because there is no accepted reference value (see 25.3), but upper and lower bounds can be estimated by a theoretical analysis of potential systematic errors, credible bounds for this uncorrectable balance of the

bias should be given in the bias statement (see example Ex 9 of 31.9). (9)

NOTE 4—No formula for combining the precision and the bias of a test method into a single numerical value of accuracy is likely to be useful. Instead separate statements of precision and bias should be presented. The value may then be used jointly in any specific application of the test method.

30. Elements of a Statement of Precision and Bias

30.1 The precision and bias section of a test method should include, as a minimum, the elements specified in 30.2 through 30.5 and in 30.7:

30.2 A brief description of the interlaboratory test program on which the statement is based, including (*1*) what materials were tested, (*2*) number of laboratories, (*3*) number of test results per laboratory per material, and the (*4*) interlaboratory practice (usually Practice E 691) followed in the design of the study and analysis of the data. This section should give the ASTM Research Report number for the interlaboratory data and analysis.

30.3 A description of any deviation from complete adherence to the test method for each test result, such as preparation in one laboratory of the cured test sheets and distribution thereof to the participating laboratories, when curing is a specified part of the test method.

30.4 The number of test determinations and their combination to form a test result, if not clearly defined in the body of the test method.

30.5 A statement of the precision between test results expressed in terms of the 95 % repeatability limit and the 95 % reproducibility limit (see 28.2), including any variation of these statistics with test level or material (see 28.5 and 28.6). Report the repeatability and reproducibility standard deviations (or percent coefficients of variation) among test results as indicated in 28.2. Finally, state that repeatability and reproducibility are used as directed in ASTM Practice E 177.

30.6 If precision under additional conditions (for example, operator-to-operator or day-to-day) has been determined, report the number of operators or days per laboratory. Include a careful description of the additional conditions, and the precision values obtained, using such terminology as 95 % limit (operator-to-operator within laboratory).

30.7 A statement concerning what is known about bias, including how the method has been modified to adjust for what is known about bias and that it is now without known bias. If the value of the property being measured can be defined only in terms of the test method, state this and whether the method is generally accepted as a reference method. If an estimate of the maximum bias of the method can be made on theoretical grounds (for example, by examining the maximum probable contributions of various steps in the procedure to the total bias), then describe these grounds in this section. Give the ASTM Research Report number on the theoretical or experimental study of bias.

STATEMENTS OF PRECISION AND BIAS

31. Statements of Precision and Bias

31.1 *Example Statements of Precison and Bias*—In the simplest case, the statement will appear essentially as shown in illustrative example Ex.1. Ex.1 is a simplified example. Normally, at least six laboratories and at least three materials should be included in the study in accordance with Practice E 691. (No general conclusions about the test method can be considered valid from so few materials and laboratories.)

Ex.1 Precision and Bias

Ex.1.1 *Interlaboratory Test Program*—An interlaboratory study of the permanent deformation of elastomeric yarns was run in 1969. Each of two laboratories tested five randomly drawn test specimens from each of three materials. The design of the experiment, similar to that of Practice E 691, and a within-between analysis of the data are given in ASTM Research Report No. XXXX.

Ex.1.2 *Test Result*—The precision information given below for average permanent deformation in percentage points at 100-min relaxation time is for the comparison of two test results, each of which is the average of five test determinations.

Ex.1.3 *Precision:*

95 % repeatability limit (within laboratory)	0.8 %
95 % reproducibility limit (between laboratories)	2.9 %

The above terms (repeatability limit and reproducibility limit) are used as specified in Practice E 177. The respective standard deviations among test results, related to the above numbers by the factor 2.8, are:

repeatability standard deviation	= 0.3 %
reproducibility standard deviation	= 1.0 %.

Ex.1.4 *Bias*—This method has no bias because permanent deformation of elastomeric yarns is defined in terms of this method.

31.2 The illustrative example Ex.2 is another simplified example in which only two materials have been used but with the required minimum number (six) of participating laboratories:

Ex.2 Precision and Bias

Ex.2.1 *Interlaboratory Test Program*—An interlaboratory study was run in which randomly drawn test specimens of two materials (kraft envelope paper and wove envelope paper) were tested for tearing strength in each of six laboratories, with each laboratory testing two sets of five specimens of each material. Except for the use of only two materials, Practice E 691 was followed for the design and analysis of the data, the details are given in ASTM Research Report No. XXXY.

Ex.2.2 *Test Result*—The precision information given below in the units of measurement (grams) is for the comparison of two test results, each of which is the average of five test determinations:

Ex.2.3 *Precision:*

	Material A	Material B
Average Test Value	45 gf	100 gf
95 % repeatability limit (within laboratory)	3 gf	7 gf
95 % reproducibility limit (between laboratories)	6 gf	12 gf

The above terms (repeatability limit and reproducibility limit) are used as specified in Practice E 177. The respective standard deviations among test results may be obtained by dividing the above limit values by 2.8.

Ex.2.4 *Bias*—The original draft of this abbreviated method was experimentally compared in one laboratory with the appropriate reference method (ASTM DXXXX) and was found to give results approximately 10 % high, as theoretical considerations would suggest (See ASTM Research Report No. XXXW). An adjustment for this bias is now made in Section XX on calculations, so that the final result is now without known bias.

31.3 If a sufficient number of different materials to cover the test range are included in the interlaboratory study (6 or more in accordance with Practice E 691), then the approximate variation in precision with test level may be determined.

Since two distinctly separate classes of material are tested by the method shown in illustrative example Ex.3, two separate interlaboratory studies were made. In the first study, the repeatability was found to be essentially proportional to the test value (with minor variation from material to material as shown), whereas the reproducibility had a more complex linear relationship (that is, a constant as well as a proportional term). In the second study, the repeatability and the reproducibility were each found to be proportional to the test value.

Ex.3 Precision

Coarse-fiber materials	
Test range	30 to 150 g
95 % repeatability limit (within laboratory)	7 % (6 to 8.5 %) of the test result
95 % reproducibility limit (between laboratories)	2 g + 10 % (8 to 12 %) of the test result
Well beaten (fine-fiber) materials	
Test range	20 to 75 g
95 % repeatability limit (within laboratory)	4 % (3.5 to 5 %) of the test result
95 % reproducibility limit (between laboratories)	7 % (5 to 8 %) of the test result

Ex.3.1 The values shown above for the limits are the average (and range) in each case as found in separate interlaboratory studies for the coarse and fine-fiber materials. The terms repeatability limit and reproducibility limit are used as specified in Practice E 177. The respective standard deviations among test results may be obtained by dividing the above limit values by 2.8.

31.4 Precision information can often be obtained from studies made for other purposes. Example below illustrates this approach and also illustrates another way of showing variation from material to material.

Ex.4 Precision

Ex.4.1 *Interlaboratory Test Program*—The information given below is based on data obtained in the TAPPI Collaborative Reference Program for self-evaluation of laboratories, Reports 25 through 51 (Aug. 1973 through Jan. 1978). Each report covers two materials with each of approximately 16 laboratories testing 5 specimens of each material.

Ex.4.2 *Test Result*—The precision information given below has been calculated for the comparison of two test results, each of which is the average of 10 test determinations.

Ex.4.3 *95 % Repeatability Limit (within laboratory)*—The repeatability is 5.4 % of the test result. For the different materials the repeatability ranged from 3.7 to 9.6 %. The range of the central 90 percent of the repeatability values was 3.9 to 8.7 %.

Ex.4.4 *95 % Reproducibility Limit (between laboratories)*—The reproducibility is 19.2 % of the test result. For the different materials the range of all of the calculations of reproducibility was 6.4 to 45.4 %. The range of the central 90 percent of the calculations was 12.2 to 25.5 %.

Ex.4.5 *Definitions and Standard Deviations*—The above terms repeatability limit and reproducibility limit are used as specified in Practice E 177. The respective percent coefficients of variation among test results may be obtained by dividing the above numbers by 2.8.

31.5 Precision is often constant for low test values and proportional for higher test values, as shown in the following example:

Ex.5 Precision

Test range	0.010 to 1200 mm
95 % repeatability limit (within laboratory)	0.002 mm or 2.5 % of the average, whichever is larger
95 % reproducibility limit (between laboratories)	0.005 mm or 4.2 % of the average, whichever is larger

The above terms repeatability limit and reproducibility limit are used as specified in Practice E 177. The respective standard deviations and percent coefficients of variation among test results may be obtained by dividing the above limit values by 2.8.

31.6 A table may be used especially if the precision indexes vary irregularly from material to material. Note in the following example that the materials have been arranged in increasing order of test value:

Ex.6A Precision[A]

Material	Glucose in Serum, Average	Repeatability Standard Deviation	Reproducibility Standard Deviation	Repeatability Limit	Reproducibility Limit
A	41.518	1.063	1.063	2.98	2.98
B	79.680	1.495	1.580	4.19	4.42
C	134.726	1.543	2.148	4.33	6.02
D	194.717	2.625	3.366	7.35	9.42
E	294.492	3.935	4.192	11.02	11.74

[A] The table was calculated using the relationship: limit = 2.8 × standard deviation. The quantity 1.960 $\sqrt{2}$ rounds to 2.77 or 2.8.

Ex.6A.1 *Interlaboratory Test Program*—An interlaboratory study of glucose in serum was conducted in accordance with Practice E 691 in eight laboratories with five materials, with each laboratory obtaining three test results for each material. See ASTM Research Report No. XXXX.

Ex.6A.2 The terms repeatability limit and reproducibility limit in Ex.6A are used as specified in Practice E 177.

Ex.6B Precision

Material	Pentosans in Pulp, Average	Repeatability Standard Deviation	Reproducibility Standard Deviation	Repeatability Limit	Reproducibility Limit
A	0.405	0.015	0.114	0.04	0.32
B	0.884	0.032	0.052	0.09	0.14
C	1.128	0.143	0.196	0.40	0.55
D	1.269	0.038	0.074	0.11	0.21
E	1.981	0.040	0.063	0.11	0.18
F	4.181	0.032	0.209	0.09	0.58
G	5.184	0.133	0.243	0.37	0.68
H	10.401	0.194	0.585	0.54	1.64
I	16.361	0.216	1.104	0.60	3.09

Ex.6B.1 *Interlaboratory Test Program*—An interlaboratory study of pentosans in pulp was conducted in accordance with Practice E 691 with seven participating laboratories each obtaining three test results of each of nine materials. See ASTM Research Report No. YYYY.

Ex.6B.2 The terms repeatability limit and reproducibility limit in Ex.6B are used as specified in Practice E 177.

31.7. If multi-operator precision (23.1) as well as repeatability and reproducibility has been evaluated, its variation among laboratories may be shown as in illustrative example Ex.7.

Ex.7 Precision

Average test value	100 g
95 % repeatability limit (within a laboratory)	7 % (6 to 8 %) of the test result
95 % reproducibility limit (between laboratories)	15 % (13 to 16 %) of the test result
95 % limit (operator-to-operator, within laboratory)	6 % to 15 % of the test result

Ex.7.1 The values shown above for the limits are, in each case, the average (and range) found in the interlaboratory study. The terms, repeatability, reproducibility and operator-to-operator limit,

are used as specified in this practice. The respective standard deviations may be obtained by dividing the above limit values by 2.8.

NOTE—Since the lower value for the operator-to-operator effect was obtained in a laboratory that has a continuing training program for its operators, it appears that the operator-to-operator effect may be reduced by training. Furthermore, since the upper value for the operator-to-operator effect in some laboratories is as high as the reproducibility between laboratories, it is possible that reproducibility also may be improved by better operator training.

31.8. An example of a bias statement when bias has been removed through comparison with a reference method is given in 31.2 and Ex.2.4. A similar statement would apply for any accepted reference value, for example, from an accepted reference material. If bias depends on other properties of the material, a statement such as the following might be used:

Ex.8 Bias

Ex.8.1. *Bias*—A ruggedness study (ASTM Research Report No. XXXZ) showed that test results are temperature dependent, with the dependence varying with the type of material. Therefore, if the test temperature cannot be maintained within the specified limits, determine the temperature dependence for the specific material being tested and correct test results accordingly.

31.9 A maximum value for the bias of a test method may be estimated by an analysis of the effect of apparatus and procedural tolerances on the test results, as illustrated below:

Ex.9 Bias

Ex.9.1. *Bias*—Error analysis shows that the absolute value of the maximum systematic error that could result from instrument and other tolerances specified in the test method is 3.2 % of the test result.

31.10 Even when a quantitative statement on bias is not possible, it is helpful to the user of the method to know that the developers of the method have considered the possibility of bias. In such cases, a statement on bias based on one of the following examples may be used:

Ex.10.1 *Bias*—This method has no bias because (insert the name of the property) is defined only in terms of this test method.

Ex.10.2 *Bias*—Since there is no accepted reference material, method, or laboratory suitable for determining the bias for the procedure in this test method for measuring (insert the name of the property), no statement on bias is being made.

Ex.10.3 *Bias*—No justifiable statement can be made on the bias of the procedure in this test method for measuring (insert the name of the property) because (insert the reason).

APPENDIX

(Nonmandatory Information)

X1. DESCRIPTIONS OF TERMS

X1.1 The following brief descriptions have been extracted from the text. For fuller discussions of the concepts, see the referenced sections.

X1.1.1 *accepted reference value*—a value that serves as an agreed-upon reference for comparison. (Section 16)

X1.1.2 *accuracy*—a generic concept of exactness related to the closeness of agreement between the average of one or more test results and an accepted reference value. (Section 20)

X1.1.3 *bias*—a generic concept related to a consistent or systematic difference between a set of test results from the process and an accepted reference value of the property being measured. (Section 19)

X1.1.4 *observation or observed value*—the most elemental single reading or corrected reading obtained in the process of making a measurement. (Section 7)

X1.1.5 *precision*—a generic concept related to the closeness of agreement between test results obtained under prescribed like conditions from the measurement process being evaluated. (Section 18).

X1.1.6 *repeatability*—the closeness of agreement between test results obtained under repeatability conditions.

X1.1.7 *repeatability conditions*—conditions under which test results are obtained with the same test method in the same laboratory, by the same operator with the same equipment, in the shortest practical period of time, using test units or test specimens taken at random, from a single quantity of material that is as nearly homogeneous as possible.

X1.1.8 *reproducibility*—a general term for a measure of precision applicable to the variability between single test results obtained in different laboratories using test specimens taken at random from a single sample of material. (Section 25)

X1.1.9 *statistical control*—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. (Section 17)

X1.1.10 *test determination*—(*1*) the process of calculating from one or more observations a property of a single test specimen, or (*2*) the value obtained from the process. (Section 8)

X1.1.11 *test method*—a definitive procedure for the identification, measurement, and evaluation of one or more qualities, characteristics, or properties of a material, product, system, or service that produces a test result. (Section 5)

X1.1.12 *test result*—the value obtained by carrying out the complete protocol of the test method once, being either a single test determination or a specified combination of a number of test determinations. (Section 9)

REFERENCES

(1) Youden, W. J., "Experimental Design and ASTM Committee", *Materials Research and Standards*, ASTM, November 1961, pp. 862–867.

(2) Wernimont, Grant, "Ruggedness Evaluation of Test Procedures", *Standardization News*, Vol. 5, No. 3, March 1977, pp. 13–16.

(3) Youden, W. J., *Statistical Techniques for Collaborative Tests*, Association of Official Analytical Chemists, Washington, DC, 1967, pp. 29–32.

(4) Shewhart, Walter A., *Statistical Method from the Viewpoint of Quality Control*, The Graduate School of the Department of Agricultural, Washington, DC, 1939.

(5) Mandel, John, *The Statistical Analysis of Experimental Data*, Interscience-Wiley Publishers, New York, NY, 1964 (out of print); corrected and reprinted by Dover Publishers, New York, NY, 1984, p. 105.

(6) Manual on Presentation of Data and Control Chart Analysis, *ASTM STP 15D*, ASTM 1976.

(7) Murphy, R. B., "On the Meaning of Precision and Accuracy", *Materials Research and Standards*, ASTM, April 1961, pp. 264–267.

(8) Eisenhart, Churchill, "The Reliability of Measured Values—Part I: Fundamental Concepts", *Photogrammetric Engineering*, June 1952, pp. 542–561.

(9) Eisenhart, Churchill, "Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems", *Journal of Research of the National Bureau of Standards, 67C*, 1963, pp. 161–187.

(10) Mandel, John and Lashof, T. W., "The Nature of Repeatability and Reproducibility," *Journal of Quality Technology*, Vol 19, No. 1, January 1987, pp. 29–36.