



Standard Practice for Use of the Terms Precision and Bias in ASTM Test Methods¹

This standard is issued under the fixed designation E177; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ϵ) indicates an editorial change since the last revision or reapproval.

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

1. Scope

1.1 The 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.

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

2. Referenced Documents

2.1 ASTM Standards:²

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

E2282 Guide for Defining the Test Result of a Test Method

3. Terminology

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

3.1.1 **accepted reference value, n** —a value that serves as an agreed-upon reference for comparison, and which is derived as: (1) a theoretical or established value, based on scientific principles, (2) an assigned or certified value, based on experimental work of some national or international organization, or (3) a consensus or certified value, based on collaborative experimental work under the auspices of a scientific or engineering group.

¹ This practice 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. 1, 2010. Published November 2010. Originally approved in 1961. Last previous edition approved in 2008 as E177 – 08. DOI: 10.1520/E0177-10.

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.

3.1.1.1 **Discussion**—A national or international organization, referred to in 3.1.1 (2), generally maintains measurement standards to which the reference values obtained are traceable.

3.1.2 **accuracy, n** —the closeness of agreement between a test result and an accepted reference value.

3.1.2.1 **Discussion**—The term accuracy, when applied to a set of test results, involves a combination of a random component and of a common systematic error or bias component.

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

3.1.3.1 **Discussion**—Bias is the total systematic error as contrasted to random error. There may be one or more systematic error components contributing to the bias. A larger systematic difference from the accepted reference value is reflected by a larger bias value.

3.1.4 **characteristic, n** —a property of items in a sample or population which, when measured, counted or otherwise observed, helps to distinguish between the items. **E2282**

3.1.5 **intermediate precision, n** —the closeness of agreement between test results obtained under specified intermediate precision conditions.

3.1.5.1 **Discussion**—The specific measure and the specific conditions must be specified for each intermediate measure of precision; thus, “standard deviation of test results among operators in a laboratory,” or “day-to-day standard deviation within a laboratory for the same operator.”

3.1.5.2 **Discussion**—Because the training of operators, the agreement of different pieces of equipment in the same laboratory and the variation of environmental conditions with longer time intervals all depend on the degree of within-laboratory control, the intermediate measures of precision are likely to vary appreciably from laboratory to laboratory. Thus, intermediate precisions may be more characteristic of individual laboratories than of the test method.

3.1.6 **intermediate precision conditions, n** —conditions under which test results are obtained with the same test method using test units or test specimens taken at random from a single quantity of material that is as nearly homogeneous as possible, and with changing conditions such as operator, measuring equipment, location within the laboratory, and time.

3.1.7 *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**

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

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

3.1.9.1 *Discussion*—Precision depends on random errors and does not relate to the accepted reference value.

3.1.9.2 *Discussion*—The measure of precision usually is expressed in terms of imprecision and computed as a standard deviation of the test results. Less precision is reflected by a larger standard deviation.

3.1.9.3 *Discussion*—“Independent test results” means results obtained in a manner not influenced by any previous result on the same or similar test object. Quantitative measures of precision depend critically on the stipulated conditions. Repeatability and reproducibility conditions are particular sets of extreme stipulated conditions.

3.1.10 *repeatability, n*—precision under repeatability conditions.

3.1.10.1 *Discussion*—Repeatability is one of the concepts or categories of the precision of a test method.

3.1.10.2 *Discussion*—Measures of repeatability defined in this compilation are *repeatability standard deviation* and *repeatability limit*.

3.1.11 *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.

3.1.11.1 *Discussion*—See *precision*. The “same operator, same equipment” requirement means that for a particular step in the measurement process, the same combination of operator and equipment is used for every test result. Thus, one operator may prepare the test specimens, a second measure the dimensions and a third measure the mass in a test method for determining density.

3.1.11.2 *Discussion*—By “in the shortest practical period of time” is meant that the test results, at least for one material, are obtained in a time period not less than in normal testing and not so long as to permit significant change in test material, equipment or environment.

3.1.12 *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 %).

3.1.12.1 *Discussion*—The repeatability limit is 2.8 ($\approx 1.96\sqrt{2}$) times the repeatability standard deviation. This multiplier is independent of the size of the interlaboratory study.

3.1.12.2 *Discussion*—The approximation to 0.95 is reasonably good (say 0.90 to 0.98) when many laboratories (30 or more) are involved, but is likely to be poor when fewer than eight laboratories are studied.

3.1.13 *repeatability standard deviation (s_r), n*—the standard deviation of test results obtained under repeatability conditions.

3.1.13.1 *Discussion*—It is a measure of the dispersion of the distribution of test results under repeatability conditions.

3.1.13.2 *Discussion*—Similarly, “repeatability variance” and “repeatability coefficient of variation” could be defined and used as measures of the dispersion of test results under repeatability conditions.—In an interlaboratory study, this is the pooled standard deviation of test results obtained under repeatability conditions.

3.1.13.3 *Discussion*—The repeatability standard deviation, usually considered a property of the test method, will generally be smaller than the within-laboratory standard deviation. (See *within-laboratory standard deviation*.)

3.1.14 *reproducibility, n*—precision under reproducibility conditions.

3.1.15 *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.

3.1.15.1 *Discussion*—*Identical material* means either the same test units or test specimens are tested by all the laboratories as for a nondestructive test or test units or test specimens are taken at random from a single quantity of material that is as nearly homogeneous as possible.

A different laboratory of necessity means a different operator, different equipment, and different location and under different supervisory control.

3.1.16 *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 %).

3.1.16.1 *Discussion*—The reproducibility limit is 2.8 ($\approx 1.96\sqrt{2}$) times the reproducibility standard deviation. The multiplier is independent of the size of the interlaboratory study (that is, of the number of laboratories participating).

3.1.16.2 *Discussion*—The approximation to 0.95 is reasonably good (say 0.90 to 0.98) when many laboratories (30 or more) are involved but is likely to be poor when fewer than eight laboratories are studied.

3.1.17 *reproducibility standard deviation (s_R), n*—the standard deviation of test results obtained under reproducibility conditions.

3.1.17.1 *Discussion*—Other measures of the dispersion of test results obtained under reproducibility conditions are the “reproducibility variance” and the “reproducibility coefficient of variation.”

3.1.17.2 *Discussion*—The reproducibility standard deviation includes, in addition to between-laboratory variability, the repeatability standard deviation and a contribution from the interaction of laboratory factors (that is, differences between operators, equipment and environments) with material factors (that is, the differences between properties of the materials other than that property of interest).

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

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

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

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

3.1.22 *test specimen, n*—the portion of a test unit needed to obtain a single test determination. **E2282**

3.1.23 *test unit, n*—the total quantity of material (containing one or more test specimens) needed to obtain a test result as specified in the test method. See *test result*. **E2282**

3.1.24 *trueness, n*—the closeness of agreement between the population mean of the measurements or test results and the accepted reference value.

3.1.24.1 *Discussion*—“Population mean” is, conceptually, the average value of an indefinitely large number of test results

3.1.25 *within-laboratory standard deviation, n*—the standard deviation of test results obtained within a laboratory for a single material under conditions that may include such elements as different operators, equipment, and longer time intervals.

3.1.25.1 *Discussion*—Because the training of operators, the agreement of different pieces of equipment in the same laboratory and the variation of environmental conditions with longer time intervals depend on the degree of within-laboratory control, the within-laboratory standard deviation is likely to vary appreciably from laboratory to laboratory.

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 of precision is not intended to exhibit values that can be exactly duplicated in every user’s laboratory. Instead, the statement provides guidelines as to the magnitude of variability that can be expected between test results when the method is used in one, or in two or more, reasonably competent laboratories. For a discussion of precision, see 8.1.

4.3 *Bias*—A statement of 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 9.1).

4.3.1 An alternative term for bias is trueness, which has a positive connotation, in that greater bias is associated with less favorable trueness. Trueness is the systematic component of accuracy.

4.4 *Accuracy*—The term “accuracy,” used in earlier editions of Practice E177, embraces both precision and bias (see 9.3 and Note 3).

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 7.1.

5.3 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 (7.1), and might also reference a sampling procedure appropriate for the intended use of the method.

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

6. Measurement Terminology

6.1 A *test result* is the value obtained by carrying out the complete protocol of the test method once, being as simple as the result of a single direct visual observation on a test specimen or the result of a complex series of automated procedures with the test result calculation performed by a computer.

6.2 The following terms are used to describe partial results of the test method: *observed value*, and *test determination*, which are more fully described in Guide E2282.

6.2.1 An *observed value* is interpreted as the most elemental single reading obtained in the process of making an observation. As examples, an observation may involve a zero-adjusted micrometer reading of the thickness of a test strip at one position along the strip or the weight of a subsample taken from a powder sample.

6.2.2 A *test determination* summarizes or combines one or more observed values. For example, (1) the measurement of the bulk density of a powder may involve the observation of the mass and the tamped volume of the sample specimen, and the calculated bulk density as the ratio mass/volume is a test determination; (2) the test determination of the thickness of a test specimen strip may involve averaging micrometer caliper observations taken at several points along the strip.

6.2.3 A test result summarizes or combines one or more test determinations. For example, (1) a test method on bulk density might require that the test determination of density for each of five subsamples of the powder sample be averaged to calculate the test result; (2) a test method may involve multiple automated operations, combined with a calibration procedure, with many observed values and test determinations, and the test result calculated and printed out by a computer.

6.3 Precision statements for ASTM test methods are applicable to comparisons between test results, not test determinations nor observations, unless specifically and clearly indicated otherwise.

7. Sources of Variability

7.1 *Experimental Realization of a Test Method:*

7.1.1 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.

7.1.2 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 7.2-7.6.

7.2 Operator:

7.2.1 *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.

7.2.2 *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.

7.2.2.1 If requirements are not explicitly stated in the test method (see 5.4), they must be included in the instructions for the interlaboratory study (see Practice E691).

7.2.3 *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.

7.3 Apparatus:

7.3.1 *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.

7.3.2 *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.

7.4 Environment:

7.4.1 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 (see 7.7.2). 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.

7.5 Sample (Test Specimens):

7.5.1 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. In interlaboratory evaluation of test methods to determine testing variability, special attention is required in the selection of the material sample (see 8.1.4 and Practice E691) 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.

7.6 Time:

7.6.1 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. 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 (see 8.2 and 8.3).

7.7 Statistical Control:

7.7.1 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.

7.7.2 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 (1),³ but may be helped by the use of control charts (see Part 3, MNL 7) (2, 3).

7.7.2.1 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.

7.7.3 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.

7.7.4 The discussion in succeeding sections assumes that the measurement process is in a state of statistical control for some specified set of conditions. If measurements are all to be made in a given laboratory, for example, any systematic deviation from the expected value pertinent to that laboratory will show up as a bias for measurements made under the prescribed conditions (see 9.1).

8. Precision

8.1 Precision:

8.1.1 The *precision* of a measurement process, and hence the stated precision of the test method from which the process is generated, is a generic concept related to the closeness of agreement between test results obtained under prescribed like conditions from the measurement process being evaluated. The measurement process must be in a state of statistical control; else the precision of the process has no meaning. The greater the dispersion or scatter of the test results, the poorer the precision. (It is assumed that the least count of the scale of the test apparatus is not so poor as to result in absolute agreement among observations and hence among test results.) Measures of dispersion, usually used in statements about precision, are, in fact, direct measures of imprecision. Although 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 10.1).

8.1.2 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.

8.1.3 The precision of the measurement process will depend on what sources (7.1-7.6) of variability are purposely included

and may also depend on the test level (see 10.4.9). 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.

8.1.4 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 E691 for further information).

8.2 Repeatability and Laboratory Bias :

8.2.1 *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 8.2.1.1-8.2.1.3, the last situation being most reliable; that is, the estimate of this precision is improved progressively by pooling additional information.

NOTE 1—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.

8.2.1.1 *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.

8.2.1.2 *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 8.2.1.1 must be repeated on different days by the same and different operators. Then the precision estimates, obtained as in 8.2.1.1, 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

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

the test results pooled in order to obtain an overall single-operator-day-equipment precision for that laboratory.

8.2.1.3 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 **8.2.1.2** 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 **E691**, 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.

8.2.2 Repeatability Conditions—While other conditions (**8.3**) have sometimes been used for obtaining repeated test results in the determination of repeatability, the preferred conditions (illustrated above in **8.2.1-8.2.1.3**) 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 **E691**, 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 1** above.

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

8.2.4 Bias of a Particular Laboratory, relative to the other laboratories may be calculated by averaging test values obtained as described in **8.2.1.2** for that laboratory and comparing the result with the average of all test values obtained as described in **8.2.1.3**. The *bias of the test method* may be calculated by comparing the latter average with the accepted reference value (**9.2**), or it may be determined as described in **8.4.3**. Once the bias is known, the method should be modified to correct for it (see **9.1.5**).

8.3 Other Within-A-Single Laboratory Precisions:

8.3.1 Single-Operator-Apparatus, Multi-Day Precision—A single operator using one set of equipment obtains replicate test results as in **8.2**, but one on each of two or more days. Since the time interval is greater than in **8.2**, 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 **8.2.1.3**, 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.

8.3.2 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 **10.5.7**). If information on multi-operator precision is needed by a laboratory, it should be studied by that laboratory.

8.4 Reproducibility and Bias of the Test Method:

8.4.1 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.

8.4.1.1 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 **E691**, 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.

8.4.2 Reproducibility, as used in **8.4.1** and **8.4.1.1**, 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 than that defined in Terminology **E456** because it assumes the simpler interlaboratory study of **8.2.1.3** and Practice **E691** where only one operator-day-apparatus combination is involved in each laboratory.

8.4.3 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 **8.4.1** for that material with the accepted reference value (see **9.2**) for that material. If no accepted reference value is available, bias cannot be calculated (however, see **10.3.2**). For a valid determination of

bias, the results of the test method must indicate a state of statistical control (see 7.7).

9. Bias

9.1 Bias:

9.1.1 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.

9.1.2 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 10.4.9).

9.1.3 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 9.2.1 (see 8.2.3 and 8.4.3).

9.1.4 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.

9.1.5 The concept of bias may also be used to describe the systematic difference between two operators, two test sites (see 8.2.3), 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.

9.2 Accepted Reference Value:

9.2.1 A measurement process is generated by the application of a test method. Variability can be introduced unintentionally into the measurement process through the impact of many sources, such as heterogeneity of the material, state of maintenance and calibration of equipment, and environmental fluctuations (7.1-7.6). The variability may include systematic as well as random components. The systematic components may be evaluated (9.1) if an accepted reference value is available. An *accepted reference value* is a value that serves as an agreed-upon reference for comparison. It may be:

- (1) a theoretical or established value based on scientific principles;
- (2) an assigned value based on experimental work of some national or international organization such as the U.S. National Institute of Standards and Technology;
- (3) a consensus value based on collaborative experimental work under the auspices of a scientific or engineering group; or
- (4) for an isolated application, when no value for (1), (2), or (3) exists, an agreed upon value obtained using an accepted reference method.

NOTE 2—When the accepted reference value is a theoretical value, it is sometimes referred to as the “true” value, but this usage is not recommended.

9.3 Accuracy:

9.3.1 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.

9.3.2 There are two schools of thought on defining the *accuracy of a measuring process* (2, 4). In either case, the measurement process must be in a state of statistical control, otherwise the accuracy of the process has no meaning:

9.3.2.1 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.

9.3.2.2 The closeness of agreement between the accepted reference value and the individual test results (5, 6).

9.3.3 In 9.3.2.1 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 9.3.2.2 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.

9.4 Variation of Precision and Bias with Sources of Variability:

9.4.1 The precision and bias of test results obtained by repeated applications of a test method depend upon what combinations of the sources of variability (7.1-7.6) 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 8.2-8.4.

10. Statements of Precision and Bias

10.1 Indexes of Precision:

10.1.1 *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 9.4-8.4) 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.

10.1.2 *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 E691), where standard deviation has its usual meaning (for example, see Terminology E456). 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 E691 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.

10.1.3 Possible Indexes of Precision:

10.1.3.1 “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).

10.1.3.2 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 $\sqrt{2} s$) = 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 8.2 and 8.4, these limits are known as the repeatability and reproducibility limits.

10.1.3.3 Multiplier for 95 % Limit:

(1) The multiplier 1.960 or 2.0 used in 10.1.3.1 and 10.1.3.2 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.

(2) It may be thought that the use of the multiplier 1.960 (or approximately 2.0) in 10.1.3.1 and 10.1.3.2 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 (7). 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.

10.1.4 Indexes in Percent—In some instances (see 10.2.5) 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.

10.1.5 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.

10.2 Preferred Indexes of Precision for ASTM Test Methods:

10.2.1 Preferred Types of Precision and Preferred Indexes—The types of precision described in 8.2.1.3 and 8.4.1, 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 10.1.3.2), namely, 2.8 s or 2.8 CV %. Also the corresponding standard deviation (s) or percent coefficient of variation (CV %) shall be indicated.

10.2.2 Recommended Terminology for Preferred Indexes:

r = 95 % repeatability limit (1)

R = 95 % reproducibility limit (1)

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

r = 95 % repeatability limit (within a laboratory) (2)

R = 95 % reproducibility limit (between laboratories) (2)

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

s_r = repeatability standard deviation (within a laboratory) (3)

s_R = reproducibility standard deviation (between laboratories) (3)

and for the coefficients of variation:

CV % _{r} = repeatability coefficient of variation in percent (within a laboratory) (4)

CV % _{R} = reproducibility coefficient of variation in percent (between laboratories) (4)

where:

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

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

depending on how the indexes vary with the test level (see 10.2.5). 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.

10.2.3 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.

10.2.4 Quantitative estimates of repeatability and reproducibility may be obtained from an interlaboratory study conducted as directed in Practice E691.

10.2.5 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.

10.2.5.1 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.

10.2.5.2 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.

10.2.5.3 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.

10.2.5.4 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 [10.5.3](#)), 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 [10.5.6](#)) 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 ([10.4.9](#)) and the description may eventually allow identification of the cause.

10.3 Preferred Statements of Bias for ASTM Test Methods:

10.3.1 Some information may be available concerning the bias or part of the bias of a test method as determined from an interlaboratory study ([8.4.3](#) and [8.2.3](#)) or from known effects of environmental or other deviations as determined in ruggedness tests (see [10.1.3.2](#)). 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.

10.3.2 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 [8.4.3](#)), 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 [10.5.9](#)) (6).

NOTE 3—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.

10.4 Elements of a Statement of Precision and Bias:

10.4.1 The precision and bias section of a test method should include, as a minimum, the elements specified in [10.4.2-10.4.5](#) and in [10.4.7](#):

10.4.2 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 [E691](#)) 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.

10.4.3 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.

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

10.4.5 A statement of the precision between test results expressed in terms of the 95 % repeatability limit and the 95 % reproducibility limit (see [10.2.2](#)), including any variation of these statistics with test level or material (see [10.2.5](#) and [10.5.6](#)). Report the repeatability and reproducibility standard deviations (or percent coefficients of variation) among test results as indicated in [10.2.2](#). Finally, state that repeatability and reproducibility are used as directed in this practice.

10.4.6 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).

10.4.7 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.

10.4.8 Range of Materials:

10.4.8.1 The estimates of precision and bias described in [8.2-8.4](#) 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 [10.4.9](#)) and are likely to be compared by means of the test method.

10.4.8.2 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.

10.4.9 Variation of Precision and Bias with Material:

10.4.9.1 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.

10.4.9.2 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 10.5.3).

10.5 Examples of Statements of Precision and Bias:

10.5.1 *Example Statements of Precision 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 E691. (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 E691, 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 E177. 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.

10.5.2 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 E691 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 E177. 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.

10.5.3 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 E691), 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 E177. The respective standard deviations among test results may be obtained by dividing the above limit values by 2.8.

10.5.4 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 (8) 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 5.3 % 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 16.2 % 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 9.2 to 25.5 %.

Ex.4.5 *Definitions and Standard Deviations*—The above terms repeatability limit and reproducibility limit are used as specified in Practice E177. The respective percent coefficients of variation among test results may be obtained by dividing the above numbers by 2.8.

10.5.5 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 E177. The respective standard deviations and percent coefficients of variation among test results may be obtained by dividing the above limit values by 2.8.

10.5.6 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

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

Ex.6A.1 *Interlaboratory Test Program*—An interlaboratory study of glucose in serum was conducted in accordance with Practice E691 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 E177.

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 E691 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 E177.

10.5.7 If multi-operator precision (8.2.1) 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 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 4—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.

10.5.8 An example of a bias statement when bias has been removed through comparison with a reference method is given in 10.5.2 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.

10.5.9 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.

10.5.10 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).

11. Keywords

11.1 accepted reference value; accuracy; bias; interlaboratory study; precision; precision conditions; repeatability; reproducibility; standard deviation

REFERENCES

- (1) Shewhart, Walter A., *Statistical Method from the Viewpoint of Quality Control*, The Graduate School of the Department of Agricultural, Washington, DC, 1939.
- (2) 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.
- (3) Manual on Presentation of Data and Control Chart Analysis, *MNL7A*, ASTM 2002.
- (4) Murphy, R. B., “On the Meaning of Precision and Accuracy”, *Materials Research and Standards*, ASTM, April 1961, pp. 264–267.
- (5) Eisenhart, Churchill, “The Reliability of Measured Values—Part I: Fundamental Concepts”, *Photogrammetric Engineering*, June 1952, pp. 542–561.
- (6) 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.
- (7) 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.
- (8) TAPPI Collaborative Reference Program, Reports 25 through 51, Aug. 1973 through Jan. 1978, Technical Association of the Pulp and Paper Industry.

ASTM International takes no position respecting the validity of any patent rights asserted in connection with any item mentioned in this standard. Users of this standard are expressly advised that determination of the validity of any such patent rights, and the risk of infringement of such rights, are entirely their own responsibility.

This standard is subject to revision at any time by the responsible technical committee and must be reviewed every five years and if not revised, either reapproved or withdrawn. Your comments are invited either for revision of this standard or for additional standards and should be addressed to ASTM International Headquarters. Your comments will receive careful consideration at a meeting of the responsible technical committee, which you may attend. If you feel that your comments have not received a fair hearing you should make your views known to the ASTM Committee on Standards, at the address shown below.

This standard is copyrighted by ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959, United States. Individual reprints (single or multiple copies) of this standard may be obtained by contacting ASTM at the above address or at 610-832-9585 (phone), 610-832-9555 (fax), or service@astm.org (e-mail); or through the ASTM website (www.astm.org). Permission rights to photocopy the standard may also be secured from the ASTM website (www.astm.org/COPYRIGHT/).