~~Designation:E177–06b~~ Designation: E 177 – 08

An American National Standard

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.

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]

E 178 Practice for Dealing With Outlying Observations

E 456 Terminology Relating to Quality and Statistics

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

E 2282 Guide for Defining the Test Result of a Test Method

3. Terminology

3.1 *Definitions:* Terminology E 456 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.

3.1.1.1 *Discussion*—A national or international organization, referred to in (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. **E 2282**

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

~~3.1.4.1~~

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."

---

[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 ~~Nov. 15, 2006.~~Oct. 1, 2008. Published ~~January 2007.~~October 2008. Originally approved in 1961. Last previous edition approved in 2006 as E 177 – 06~~a~~b.

[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.

~~3.1.4.2~~

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.

~~3.1.5~~

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.

~~3.1.6~~

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. **E 2282**

3.1.8 *observed value*, *n*—the value obtained by making an observation. **E 2282**

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

~~3.1.6.1~~

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

~~3.1.6.2~~

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.

~~3.1.6.3~~

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.

~~3.1.7~~

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

~~3.1.7.1~~

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

~~3.1.7.2~~

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

~~3.1.8~~

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.

~~3.1.8.1~~

3.1.11.1 *Discussion*—See *precision*, ~~3.1.6.3.~~ 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.

~~3.1.8.2~~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.

~~3.1.9~~

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 %).

~~3.1.9.1~~

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.

~~3.1.9.2~~

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.

~~3.1.10~~

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

~~3.1.10.1~~

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

~~3.1.10.2~~

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.

~~3.1.10.3~~

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.*)

~~3.1.11~~

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

~~3.1.12~~

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.

~~3.1.12.1~~

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.

~~3.1.13~~

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 %).

~~3.1.13.1~~

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).

~~3.1.13.2~~

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.

~~3.1.14~~

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

~~3.1.14.1~~

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."

~~3.1.14.2~~

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).

~~3.1.15~~

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. **E 2282**

3.1.19 *test method*, *n*—a definitive procedure that produces a test result. **E 2282**

3.1.20 *test observation*, *n*—see **observation.** **E 2282**

3.1.21 *test result*, *n*—the value of a characteristic obtained by carrying out a specified test method. **E 2282**

3.1.22 *test specimen*, *n*—the portion of a test unit needed to obtain a single test determination. **E 2282**

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.* **E 2282**

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

~~3.1.15.1~~

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

~~3.1.16~~

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.

~~3.1.16.1~~

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 Section 15.

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 Section 16).

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 E 177, embraces both precision and bias (see Section 17 and Note 3).

4.5 A Table of Contents is shown below, listing the concepts in this standard.

TABLE OF CONTENTS

	Section
GENERAL CONCEPTS	
Test Method	5
Measurement Terminology	6
SOURCES OF VARIABILITY	
Experimental Realization of a Test Method	7
Operator	8
Apparatus	9
Environment	10
Sample	11
Time	12
STATISTICAL CONCEPTS	
Accepted Reference Value	13
Statistical Control	14
Precision	15
Bias	16
Accuracy	17
Variation of Precision and Bias with Material	18
Variation of Precision and Bias with Sources of Variability	19
COMBINATIONS OF SOURCES OF VARIABILITY	
Repeatability and Laboratory Bias	20
Other Within-a-Single Laboratory Precisions	21
Reproducibility and Bias of the Test Method	22
Range of Materials	23
METHODS OF EXPRESSING PRECISION AND BIAS	
Indexes of Precision	24
Preferred Indexes of Precision for ASTM Test Methods	25
Preferred Statements of Bias for ASTM Test Methods	26
Elements of a Statement of Precision and Bias	27
STATEMENTS OF PRECISION AND BIAS	
Examples of Statements of Precision and Bias	28

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 .

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 (Section), 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 E 2282.

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.

SOURCES OF VARIABILITY

7. Experimental Realization of a Test Method

7.1 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.

7.2 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 8-12.

8. Operator

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

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

8.2.1 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 E 691).

8.3 *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.

9. Apparatus

9.1 *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.

9.2 *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.

10. Environment

10.1 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 14.2). 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.

11. Sample (Test Specimens)

11.1 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 15.4 and Practice E 691) 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.

12. Time

12.1 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 Sections 20 and 21).

STATISTICAL CONCEPTS

13. Accepted Reference Value

13.1 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 (Sections 7-12). The variability may include systematic as well as random components. The systematic components may be evaluated (Section 16) 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 1—When the accepted reference value is a theoretical value, it is sometimes referred to as the "true" value, but this usage is not recommended.

14. Statistical Control

14.1 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.

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

14.2.1 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.

14.3 The presence of outliers (Practice E 178) 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.

14.4 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 Section 16).

15. Precision

15.1 The *precision* of a measurement process, and hence the stated precision of the test method from which the process is