



Designation: E 177 – 04^{ε1}

An American National Standard

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

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 reappraisal. A superscript epsilon (ϵ) indicates an editorial change since the last revision or reappraisal.

^{ε1} NOTE—Editorial corrections were made in 27.3.3 and 28.2 in April 2005.

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 1169 Guide for Conducting Ruggedness Tests

2.2 Other Document:

ISO 3534 Statistics—Vocabulary and Symbols³

3. Terminology

3.1 Definitions:

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.

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

3.1.4.1 *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.”

3.1.4.2 *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 *intermediate precision conditions, n* —conditions under which test results are obtained with the same test method using test units or test specimens (10.3) 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 *precision, n* —the closeness of agreement between independent test results obtained under stipulated conditions.

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. 1, 2004. Published December 2004. Originally approved in 1961. Last previous edition approved in 2002 as E 177 – 90a (2002).

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 Available from American National Standards Institute (ANSI), 25 W. 43rd St., 4th Floor, New York, NY 10036.

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

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

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

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

3.1.8 *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 *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 *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 *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 *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 *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 *repeatability standard deviation, n*—the standard deviation of test results obtained under repeatability conditions.

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

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

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

3.1.14.1 *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 *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 *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.

3.1.15.1 *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 18.

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 19).

4.4 **Accuracy**—The term “accuracy,” used in earlier editions of Practice E 177, embraces both precision and bias (see Section 20 and Note 4).

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
Observation	7
Test Determination	8
Test Result	9
SOURCES OF VARIABILITY	
Experimental Realization of a Test Method	10
Operator	11
Apparatus	12
Environment	13
Sample	14
Time	15
STATISTICAL CONCEPTS	
Accepted Reference Value	16
Statistical Control	17
Precision	18
Bias	19
Accuracy	20
Variation of Precision and Bias with Material	21
Variation of Precision and Bias with Sources of Variability	22
COMBINATIONS OF SOURCES OF VARIABILITY	
Repeatability and Laboratory Bias	23
Other Within-a-Single Laboratory Precisions	24
Reproducibility and Bias of the Test Method	25
Range of Materials	26
METHODS OF EXPRESSING PRECISION AND BIAS	
Indexes of Precision	27
Preferred Indexes of Precision for ASTM Test Methods	28
Preferred Statements of Bias for ASTM Test Methods	29
Elements of a Statement of Precision and Bias	30
STATEMENTS OF PRECISION AND BIAS	
Examples of Statements of Precision and Bias	31

GENERAL CONCEPTS

5. Test Method

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

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

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

NOTE 1—A screening procedure or ruggedness test is a procedure for investigation of the effects of variations in environmental and other pertinent factors on the test results obtained from a test in order to determine how control of such factors should be specified in the written description of the method. For example, temperature of the laboratory or of a heating device used in the test may have a significant effect in some cases and less in others. In a screening procedure, deliberate variations in temperature would be introduced to establish the limits of significant effect, (1, 2, 3).⁴

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

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

6. Measurement Terminology

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

6.2 For clarity, it is necessary to select certain of these terms for specific use. However, the word “measurement” will be used in a generic sense to cover observation (or observed

⁴ The boldface numbers in parentheses refer to the list of references at the end of this standard.

value), test determination and test result. The use of the word “test” by itself is discouraged.

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

7. Observation

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

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

8. Test Determination

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

8.2 Examples:

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

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

9. Test Result

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

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

9.3 Examples:

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

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

9.3.3 The test method for a tensile strength test of paper may specify that a tensile strength determination be performed on

each of ten specimens and that the ten tensile test determinations be averaged to get the test result.

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

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

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

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

SOURCES OF VARIABILITY

10. Experimental Realization of a Test Method

10.1 A realization of a test method refers to an actual application of the test method to produce a test result as specified by the test method. The realization involves an *interpretation* of the written document by a *specific test operator*, who uses a *specific unit and version of the specified test apparatus*, in the *particular environment* of his testing laboratory, to evaluate a *specified number of test specimens* of the material to be tested. Another realization of the test method may involve a change in one or more of the above emphasized experimental factors. The test result obtained by another realization of the test method will usually differ from the test result obtained from the first realization. Even when none of the experimental factors is intentionally changed, small changes usually occur. The outcome of these changes may be seen as variability among the test results.

10.2 Each of the above experimental factors and all others, known and unknown, that can change the realization of a test method, are potential sources of variability in test results. Some of the more common factors are discussed in Section 11-15.