



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 U.S. 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, health, and environmental practices and determine the applicability of regulatory limitations prior to use.*

1.3 *This international standard was developed in accordance with internationally recognized principles on standardization established in the Decision on Principles for the Development of International Standards, Guides and Recommendations issued by the World Trade Organization Technical Barriers to Trade (TBT) Committee.*

2. Referenced Documents

2.1 ASTM Standards:²

- E456 Terminology Relating to Quality and Statistics
- E691 Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method
- E1488 Guide for Statistical Procedures to Use in Developing and Applying Test Methods
- E2282 Guide for Defining the Test Result of a Test Method
- E2586 Practice for Calculating and Using Basic Statistics
- E2587 Practice for Use of Control Charts in Statistical Process Control

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, 2020. Published October 2020. Originally approved in 1961. Last previous edition approved in 2019 as E177 – 19. DOI: 10.1520/E0177-20.

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. 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.

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 *coefficient of variation, CV, n*—for a nonnegative characteristic, the ratio of the standard deviation to the mean for a population or sample. **E2586**

3.1.6 *intermediate precision, n*—precision of test results from tests conducted on identical material by the same test method in a single laboratory at the same or various times with one or more known sources of variability controlled at multiple levels.

3.1.6.1 *Discussion*—Sources of variability may include, but are not limited to, operators, equipment, instruments, reagents,

environment, or the length of the time period over which the testing was conducted.

3.1.6.2 *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.”

3.1.6.3 *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.

3.1.7 *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.

3.1.8 *laboratory precision, n*—the precision of test results obtained within a laboratory for identical material under conditions that include all known sources of variation within the laboratory over extended periods of time, often months.

3.1.8.1 *Discussion*—*Laboratory precision* is also known as site precision.

3.1.9 *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**

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

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

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

3.1.11.2 *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.

3.1.11.3 *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.

3.1.12 *repeatability, n*—precision of test results from tests conducted within the shortest practical time period on identical material by the same test method in a single laboratory with all known sources of variability conditions controlled at the same levels (see *repeatability conditions*).

3.1.12.1 *Discussion*—Repeatability is one of the concepts or categories of the precision of a test method. Measures of repeatability defined in this compilation are repeatability standard deviation and repeatability limit.

3.1.12.2 *Discussion*—Sources of variability may include,

but are not limited to, operators, equipment, instruments, reagents, or environment.

3.1.13 *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.

3.1.13.1 *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.

3.1.13.2 *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.

3.1.14 *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 %).

3.1.14.1 *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.

3.1.14.2 *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.

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

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

3.1.15.2 *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.

3.1.16 *reproducibility, n*—precision of test results from tests conducted on identical material by the same test method in different laboratories (see *reproducibility conditions*).

3.1.16.1 *Discussion*—All sources of known variability are involved because of the different sources of variation occurring between laboratories.

3.1.17 *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.

3.1.17.1 *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.

3.1.18 *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 %).

3.1.18.1 *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).

3.1.18.2 *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.

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

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

3.1.19.2 *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).

3.1.20 *standard deviation, n—of a population, σ* , the square root of the average or expected value of the squared deviation of a variable from its mean; *—of a sample, s* , the square root of the sum of the squared deviations of the observed values in the sample divided by the sample size minus 1. **E2586**

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

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

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

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

3.1.25 *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**

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

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

3.1.27 *variance, σ^2, s^2, n* —square of the standard deviation of the population or sample. **E2586**

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).

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.