Designation: E691 – 20

An American National Standard

Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method[1]

This standard is issued under the fixed designation E691; 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 This practice describes the techniques for planning, conducting, analyzing, and treating the results of an interlaboratory study (ILS) of a test method. The statistical techniques described in this practice provide adequate information for formulating the precision statement of a test method.

1.2 This practice does not concern itself with the development of test methods but rather with gathering the information needed for a test method precision statement after the development stage has been successfully completed. The data obtained in the interlaboratory study may indicate, however, that further effort is needed to improve the test method.

1.3 Since the primary purpose of this practice is the development of the information needed for a precision statement, the experimental design in this practice may not be optimum for evaluating materials, apparatus, or individual laboratories.

1.4 *Field of Application*—This practice is concerned exclusively with test methods which yield a single numerical figure as the test result, although the single figure may be the outcome of a calculation from a set of measurements.

1.4.1 This practice does not cover methods in which the measurement is a categorization; however, for many practical purposes categorical outcomes can be scored, such as zero-one scoring for binary measurements or as integers, ranks for example, for well-ordered categories and then the test result can be defined as an average, or other summary statistic, of several individual scores.

1.5 *This standard may involve hazardous materials, operations, and equipment. 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.*

1.6 *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]

E29 Practice for Using Significant Digits in Test Data to Determine Conformance with Specifications

E177 Practice for Use of the Terms Precision and Bias in ASTM Test Methods

E456 Terminology Relating to Quality and Statistics

E1169 Practice for Conducting Ruggedness Tests

E1402 Guide for Sampling Design

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 *accuracy, n*—the closeness of agreement between a test result and an accepted reference value. **E177**

3.1.2 *bias, n*—the difference between the expectation of the test results and an accepted reference value. **E177**

3.1.3 *interlaboratory study, (ILS) in ASTM, n*—a designed procedure for obtaining a precision statement for a test method, involving multiple laboratories, each generating replicate test results on one or more materials.

3.1.4 *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**

3.1.5 *precision, n*—the closeness of agreements between independent test results obtained under stipulated conditions. **E177**

[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 March 2021. Originally approved in 1979. Last previous edition approved in 2019 as E691 – 19$^{\varepsilon 1}$. DOI: 10.1520/E0691-20.

[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.

3.1.6 *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*). **E177**

3.1.7 *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. **E177**

3.1.8 *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 %). **E177**

3.1.9 *repeatability standard deviation, (s_r), n*—the standard deviation of test result obtained under repeatability conditions. **E177**

3.1.10 *reproducibility, n—precision* of test results from tests conducted on identical material by the same *test method* in different laboratories (see *reproducibility conditions*). **E177**

3.1.11 *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. **E177**

3.1.12 *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 %). **E177**

3.1.13 *reproducibility standard deviation (s_R), n*—the standard deviation of test results obtained under reproducibility conditions. **E177**

3.1.14 *ruggedness test, n*—a planned experiment in which environmental factors or test conditions are deliberately varied in order to evaluate the effects of such variation. **E1169**

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

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

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

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

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

3.1.20 *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**

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *average of the cell averages, $\bar{\bar{x}}$, n*—the average of the cell averages for a particular material.

3.2.2 *between-laboratory consistency statistic, h, n*—the ratio of the cell deviation to the standard deviation of the cell averages.

3.2.2.1 *Discussion*—This statistic is an indicator of how one laboratory's cell average compares with the average of the other laboratories for a particular material (see A1.2.2).

3.2.3 *between-laboratory standard deviation, s_L, n*—the sample standard deviation attributable to differences of test result means among laboratories.

3.2.4 *between-laboratory variance, s_L^2, n*—the sample variance component attributable to differences of test result means among laboratories.

3.2.4.1 *Discussion*—This statistic is estimated indirectly from the variance of cell averages and the repeatability variance. In situations where there is good agreement among laboratories the estimate of this variance component may be close to zero or be negative. In the latter case, the estimate is set to zero. (See Note 2 and A1.1.2.)

3.2.5 *cell, n*—the intersection of a row and column in a two-way classification table, in which the rows represent the laboratories and the columns represent the materials.

3.2.5.1 *Discussion*—The table holds the test results from an interlaboratory study, and each cell contains the test results from a particular laboratory on a particular material (see Section 7 and Table 1).

3.2.6 *cell average, $\bar{x}$, n*—the average of the test results in a particular cell.

3.2.7 *cell deviation, d, n*—the cell average minus the average of the cell averages.

3.2.8 *cell standard deviation, s, n*—the standard deviation of the test results in a particular cell.

3.2.9 *repeatability variance, s_r^2, n*—the sample variance of test results obtained under repeatability conditions.

3.2.9.1 *Discussion*—This statistic is estimated for a material as the pooled within-laboratory variances over all of the laboratories in the ILS.

3.2.10 *reproducibility variance, s_R^2, n*—the sample variance of test results obtained under reproducibility conditions.

3.2.10.1 *Discussion*—This statistic is estimated as the sum of the two variance components due to between-laboratories, s_L^2, and within-laboratories, s_r^2.

3.2.11 *standard deviation of the cell averages, $s_{\bar{x}}$, n*—the standard deviation of the cell averages for a particular material.

3.2.12 *variance of the cell averages, $s_{\bar{x}}^2$, n*—the sample variance of the cell averages for a particular material.

3.2.13 *within-laboratory consistency statistic, k, n*—the ratio of the cell standard deviation to the repeatability standard deviation.

3.2.13.1 *Discussion*—This statistic is an indicator of how one laboratory's cell standard deviation under repeatability conditions compares with the repeatability standard deviation estimated from all laboratories for a particular material (see A1.2.3).

4. Significance and Use

4.1 ASTM regulations require precision statements in all test methods in terms of repeatability and reproducibility. This practice may be used in obtaining the needed information as simply as possible. This information may then be used to

prepare a precision statement in accordance with Practice E177. Knowledge of the test method precision is useful in commerce and in technical work when comparing test results against standard values (such as specification limits) or between data sources (different laboratories, instruments, etc.).

4.1.1 When a test method is applied to a large number of portions of a material that are as nearly alike as possible, the test results obtained will not all have the same value. A measure of the degree of agreement among these test results describes the precision of the test method for that material. Numerical measures of the variability between such test results provide inverse measures of the precision of the test method. Greater variability implies smaller (that is, poorer) precision and larger imprecision.

4.1.2 Precision is reported as a standard deviation, coefficient of variation (relative standard deviation), variance, or a precision limit (a data range indicating no statistically significant difference between test results).

4.1.3 This practice is designed only to estimate the precision of a test method. However, when accepted reference values are available for the property levels, the test result data obtained according to this practice may be used in estimating the bias of the test method. For a discussion of bias estimation and the relationships between precision, bias, and accuracy, see Practice E177.

4.2 The procedures presented in this practice consist of three basic steps: planning the interlaboratory study, guiding the testing phase of the study, and analyzing the test result data.

4.2.1 The planning phase includes forming the ILS task group, the study design, selection, and number of participating laboratories, selection of test materials, material certifications if applicable, and writing the ILS protocol. A well-developed test method is essential, so including a ruggedness test to determine control of test method conditions is highly recommended.

NOTE 1—In this practice, the term *test method* is used both for the actual measurement process and for the written description of the process, while the term *protocol* is used for the directions given to the laboratories for conducting the ILS.

4.2.2 The testing phase includes material preparation and distribution, liaison with the participating laboratories, and handling of test result data received from the laboratories.

4.2.3 The data analysis utilizes tabular, graphical, and statistical diagnostic tools for evaluating the consistency of the data so that unusual values may be detected and investigated, and also includes the calculation of the numerical measures of precision of the test method pertaining to repeatability and reproducibility.

4.3 The information in this practice is arranged as follows:

5. Concepts of Test Method Precision

5.1 *Repeatability and Reproducibility*—These two terms deal with the variability of test results obtained under specified laboratory conditions and represent the two extremes of test method precision. Repeatability concerns the variability between independent test results obtained within a single laboratory in the shortest practical period of time by a single operator with a specific set of test apparatus using test specimens (or test units) taken at random from a single quantity of homogeneous material obtained or prepared for the ILS. Reproducibility deals with the variability between single test results obtained in different laboratories, each of which has applied the test method to test specimens (or test units) taken at random from a single quantity of homogeneous material obtained or prepared for the ILS.

5.1.1 *Repeatability Conditions*—The single-operator, single-set-of-apparatus requirement means that for a particular step in the measurement process the same combination of operator and apparatus 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. "Shortest practical period of time" means that the test results, at least for one material, are obtained in a time not less than in normal testing and not so long as to permit significant changes in test material, equipment or environment.

5.1.2 *Reproducibility Conditions*—The factors that contribute to variability in a single laboratory, such as operator, equipment used, calibration of the equipment, and environment