Designation: ~~E2655 – 08~~ E2655 – 14 An American National Standard

Standard Guide for Reporting Uncertainty of Test Results and Use of the Term Measurement Uncertainty in ASTM Test Methods[1]

This standard is issued under the fixed designation E2655; 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.

1. Scope

1.1 This guide provides concepts necessary for understanding the term "uncertainty" when applied to a quantitative test result. Several measures of uncertainty can be applied to a given measurement result; the interpretation of some of the common forms is described.

1.2 This guide describes methods for expressing test result uncertainty and relates these to standard statistical methodology. Relationships between uncertainty and concepts of precision and bias are described.

1.3 This guide also presents concepts needed for a laboratory to identify and characterize components of method performance. Elements that an ASTM method can include to provide guidance to the user on estimating uncertainty for the method are described.

1.4 The system of units for this guide is not specified. Dimensional quantities in the guide are presented only as illustrations of calculation methods and are not binding on products or test methods treated.

1.5 *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 limitations prior to use.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

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

E122 Practice for Calculating Sample Size to Estimate, With Specified Precision, the Average for a Characteristic of a Lot or Process

~~E141 Practice for Acceptance of Evidence Based on the Results of Probability Sampling~~

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

E456 Terminology Relating to Quality and Statistics

E691 Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method

E1402 Guide for Sampling Design

E2554 Practice for Estimating and Monitoring the Uncertainty of Test Results of a Test Method Using Control Chart Techniques

E2586 Practice for Calculating and Using Basic Statistics

2.2 *Other Standard:*

~~ISO~~ISO/IEC 17025 General Requirements for the Competence of Testing and Calibration Laboratories[3]

3. Terminology

3.1 *Definitions:*

3.1.1 Additional statistical terms are defined in Terminology E456.

3.1.2 *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

[1] This guide 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, 2008~~Oct. 1, 2014. Published ~~November 2008~~October 2014. Originally approved in 2008. Last previous edition approved in 2008 as E2655 – 08. DOI: ~~10.1520/E2655-08.~~ 10.1520/E2655-14.

[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, http://www.ansi.org.

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. **E177**

3.1.3 *error of result, n*—a test result minus the accepted reference value of the characteristic.

3.1.4 *expanded uncertainty, U, n*—uncertainty reported as a multiple of the standard uncertainty.

3.1.5 *random error of result, n*—a component of the error that, in the course of a number of test results for the same characteristic, varies in an unpredictable way.

3.1.5.1 *Discussion—*

Uncertainty due to random error can be reduced by averaging multiple test results.

3.1.6 *sensitivity coefficient, n*—differential effect of the change in a factor on the test result.

3.1.7 *standard uncertainty, u, n*—uncertainty reported as the standard deviation of the estimated value of the quantity subject to measurement.

3.1.8 *systematic error of result, n*—a component of the error that, in the course of a number of test results for the same characteristic, remains constant or varies in a predictable way.

3.1.8.1 *Discussion—*

Systematic errors and their causes may be known or unknown. When causes are known, systematic error can sometimes be reduced by incorporating corrections into the calculation of the test result.

3.1.9 *uncertainty, n*—an indication of the magnitude of error associated with a value that takes into account both systematic errors and random errors associated with the measurement or test process.

3.1.10 *uncertainty budget, n*—a tabular listing of uncertainty components for a given measurement process giving the magnitudes of contributions to uncertainty of the result from those sources.

3.1.11 *uncertainty component, n*—a source of error in a test result to which is attached a standard uncertainty.

4. Significance and Use

4.1 Part A of the "Blue Book," Form and Style for ASTM Standards, introduces the statement of measurement uncertainty as an optional part of the report given for the result of applying a particular test method to a particular material.

4.2 Preparation of uncertainty estimates is a requirement for laboratory accreditation under ~~ISO~~ISO/IEC 17025. This guide describes some of the types of data that the laboratory can use as the basis for reporting uncertainty.

5. Concepts for Reporting Uncertainty of Test Results

5.1 Uncertainty is part of the relationship of a test result to the property of interest for the material tested. When a test procedure is applied to a material, the test result is a value for a characteristic of the material. The test result obtained will usually differ from the actual value for that material. Multiple causes can contribute to the error of result. Errors of sampling and effects of sample handling make the portion actually tested not identical to the material as a whole. Imperfections in the test apparatus and its calibration, environmental, and human factors also affect the result of testing. Nonetheless, after testing has been completed, the result obtained will be used for further purposes as if it were the actual value. Reporting measurement uncertainty for a test result is an attempt to estimate the approximate magnitude of all these sources of error. In common cases the measurement will be reported in the form $x \pm u$, in which x represents the test result and u represents the uncertainty associated with x.

5.2 Practice E177 describes precision and bias. Uncertainty is a closely related but not identical concept. The primary difference between concepts of precision and of uncertainty is the object that they address. Precision (repeatability and reproducibility) and bias are attributes of the test method. They are estimates of statistical variability of test results for a test method applied to a given material. Repeatability and intermediate precision measure variation within a laboratory. Reproducibility refers to interlaboratory variation. Uncertainty is an attribute of the particular test result for a test material. It is an estimate of the quality of that particular test result.

5.3 In the case of a quantity with a definition that does not depend on the measurement or test method ~~(e.g.,~~ (for example, concentration, pH, modulus, heat content), uncertainty measures how close it is believed the measured value comes to the quantity. For results of test methods where the target is only definable relative to the test method ~~(e.g.,~~ (for example, flash points, extractable components, sieve analysis), uncertainty of a test result must be interpreted as a measure of how closely an independent, equally competent test result would agree with that being reported.

5.4 In the simplest cases, uncertainty of a test result is numerically equivalent to test method precision. That is, if an unknown sample is tested, and the test precision is known to be sigma, then uncertainty of the result of test is sigma. The term uncertainty, however, is correct to apply where variation of repeated test results is not relevant, as in the following examples.

5.4.1 *Example*—The Newtonian constant of gravitation, G, is $6.6742 \times 10^{-11} \pm 0.0010 \times 10^{-11}$ $m^3kg^{-1}s^{-2}$ based on 2002 CODATA recommended values (**1**).[4] 0.0010×10^{-11} $m^3kg^{-1}s^{-2}$ is the standard uncertainty. The value and the uncertainty together represent the state of knowledge of this fundamental physical constant. It is not naturally thought of in terms of variation of repeated measurements. Both G and its uncertainty are derived from the analysis and comparison of a variety of measurement data using methods that are an elaboration of those presented in this guide.

5.4.2 *Example*—A length is measured but the result only reported to the nearest inch (for example, a measuring rod graduated in inches was used to obtain the measurement). Precision of the reported value, in the sense of variation of repeated measurements, is zero when all reported lengths are the same. In this case it is not possible to detect random variation in the series of repeated measurements. Uncertainty of the length is primarily composed of the systematic error of ±0.5 inch due to the resolution of the measurement apparatus.

5.5 The goal in reporting uncertainty is to take account of all potential causes of error in the test result. In many cases, uncertainty can be related to components of variability due to sampling and to testing. Both of these should be taken into account for the uncertainty of the measurement when the purpose of the result is to estimate the property for the entire lot of material from which the sample was taken. Uncertainty of the lot property value based on a single determination is then $\sqrt{s_1^2+s_2^2+u_{3^2}^2}$ where s_1 is an estimate of the sampling standard deviation, s_2 is an estimate of the standard deviation of the test method, and u_3 is standard uncertainty due to factors that affect all measurements under consideration.

5.6 A commonly cited definition (**2, 3**) defines uncertainty as "a parameter, associated with the measurement result, or test result, that characterizes the dispersion of values that could reasonably be attributed to the quantity subject to measurement or characteristic subject to test." This definition emphasizes uncertainty as an attribute of the particular result, as opposed to statistical variation of test results. The uncertainty parameter is a measure of spread (for example, the standard deviation) of a probability distribution used to represent the likelihood of values of the property.[5]

5.7 The methodology for uncertainty estimates has been classified as Type A and Type B as discussed in (**4**). Type A estimates of uncertainty include standard error estimates based on knowledge of the statistical character of observations, and based on statistical analysis of replicate measurements. Type B estimates of uncertainty include approximate values derived from experience with measurement processes similar to the one being considered, and estimates of standard uncertainty derived from the range of possible measurement values for a given material and an assumed distribution of values within that range. See Practice E122 for examples ~~(e.g.,~~ (for example, rectangular, triangular, normal) where a standard deviation is derived from a range without data from samples being available. Complex estimates of test result uncertainty are calculated by combining Type A and Type B component standard uncertainties for factors contributing to error (see Section 8).

5.8 *Forms of Uncertainty Expression:*

5.8.1 *Standard Uncertainty*—The uncertainty is reported as the standard deviation of the reported value. The report $x \pm u$ implies that the value should be between $x - u$ and $x + u$ with approximate probability two-thirds, where x is the test result.

5.8.2 *Relative Standard Uncertainty*—The uncertainty is reported as a fraction of the reported value. For a measured value and a standard uncertainty, $x \pm u$, the relative standard uncertainty is u/x. This method of expressing uncertainty may be useful when standard uncertainty is proportional to the value over a wide range. However, for a particular result, reporting the value and standard uncertainty is preferred.

5.8.3 *Expanded Uncertainty*—The uncertainty is reported as $x \pm U$, where the value of U is a multiple of the standard uncertainty u. The most common multiple used is 2, which is approximately equal to the 1.96 factor for a 95 % two-sided confidence interval for the mean of a normal distribution (see 5.8.4).

5.8.4 *Confidence Intervals*—A confidence interval for a parameter (the actual value of the material property subject to measurement) consists of upper and lower limits generated from sample data by a method that ensures the limits bracket the parameter value with a stated probability 1-α, referred to as the confidence coefficient.

5.8.4.1 From statistical theory, a 95 % confidence interval for the mean of a normal distribution, given n independent observations $x_1, x_2,\ldots, x_n$ drawn from the distribution, is $\bar{x}\pm ts/\sqrt{n}$ where $\bar{x}$ is the sample mean, s is the standard deviation of the observations, and t is the 0.975 percentile of the Student's t distribution with n-1 degrees of freedom. Because Student's t distribution approaches the Normal as n increases, the value of t approaches 1.96 as n increases. This is the basis for using the factor 2 for expanded uncertainty.

5.8.4.2 Practice E2586 defines confidence intervals and provides additional detail on their interpretation.

5.8.5 *Measurement Uncertainty*—Measurement uncertainty is uncertainty reported for a test result without taking into account sampling variation or heterogeneity of the material of interest. The report of measurement uncertainty then refers specifically to the particular sample presented for analysis.

5.8.6 *Reporting Uncertainty with a Bias Component*—Good measurement practice requires that biases due to environmental and other factors should be corrected in the reported result when there is a sound basis for correction and the error in the correction

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

[5] A probability distribution representing the likelihood of property values given data is known in statistical theory as the Bayes posterior distribution of the property value.

terms themselves is not greater than the bias. Such corrections are part of the calculation of the result within the test method. The symmetrical form of reporting a measurement with standard uncertainty, $x \pm u$, is adequate for measurements where bias is absent or corrected. If the measurement process has a bias for which there is an estimate of magnitude and it is not corrected in the reported value x, a form of reporting should be used making clear both bias and random components. A typical form to highlight the asymmetry caused by bias is $x - u_l / + u_h$, where u_l = bias – standard uncertainty and u_h = bias + standard uncertainty.

5.8.7 Bias estimates are often subjective or based on weak information. When bias is present, but magnitude and direction are unknown, the uncertainty of the bias is an important part of uncertainty as a whole and should be combined with random components. The overall root mean square uncertainty is then ~~$u=\sqrt{u_{bias}^2+\sigma^2}$.~~$u=\sqrt{u_{bias}^2+\sigma^2}$.

5.9 The repeatability and reproducibility values published for an ASTM method are derived from an interlaboratory study following Practice E691 or a similar procedure. Repeatability and reproducibility values given for ASTM test methods are intended to estimate the variability of test results for competent laboratories (see Practice E177). Reproducibility measures variability of test results on identical samples derived independently by different laboratories. This reproducibility is a good guide to the uncertainty level that it is possible to achieve for measured values obtained using the method. It may be useful to a user of test results from the method in the absence of a more definite uncertainty estimate. However, a laboratory generating test results using the test method should derive the value to quote for its test results based on its own methodology and experience, which are not necessarily equivalent to the laboratories that participated in the original interlaboratory study. This is particularly true when the laboratory uses a highly refined measurement method that no other or very few other laboratories can replicate.

5.9.1 Variability of samples, when the quantity is a property of a heterogeneous material, is part of uncertainty for the measurement. This component of variability is not usually included in reproducibility because interlaboratory evaluation of test methods uses test materials that are as uniform as possible.

5.10 Certified reference values for standard materials that cannot be made to a known value are often obtained by interlaboratory testing. The average of test results for participating laboratories becomes the "consensus" accepted reference value. The standard uncertainty of the consensus value is ~~$s/\sqrt{n}$,~~$s/\sqrt{n}$, where s is the standard deviation of results reported by the n laboratories.

5.11 Practice E29 describes evaluation of conformance of a material with a specification by comparing the test result with specification limits. Some proposals (**5**) use uncertainty values in an alternative procedure for evaluating conformance with specifications. Compliance of the material with specifications is demonstrated if the entire expanded uncertainty interval is contained within the specification range. Noncompliance is demonstrated when the entire uncertainty interval is outside the specification range. Where the uncertainty interval straddles a specification limit, the test result is indecisive.

5.11.1 If this method for evaluating conformance is used, the test method shall include an explicit procedure for calculating the uncertainty interval.

6. Uncertainty for Estimates Based on Probability Samples

6.1 Classical statistical methods for estimation apply directly to the estimation of uncertainty provided the underlying distribution assumptions are met. Probability sampling (see ~~Practice~~Guide ~~E141~~E1402) is a procedure ~~for ensuring~~to ensure that statistical methods are applicable and provide valid estimates of their uncertainty. Measurement tasks to which probability samples apply include determining the proportion of items in a specified set having a qualitative observable characteristic, the average of a quantitative characteristic which may be non-uniform over a prescribed area, or the aggregate of a property for a lot of material which may be non-uniform. The examples considered illustrate some aspects of uncertainty.

6.2 ~~*Simple Statistical Applications:*~~ *Uncertainty for Average Values:*

6.2.1 ~~The~~ When the value to be reported is an average of n measurements each of which has standard deviation σ, bias is presumed to be absent, and the measurements are mutually ~~independent. Then~~independent, then uncertainty of the average value is ~~$\sigma/\sqrt{n}$.~~$\sigma/\sqrt{n}$.

6.2.2 ~~The~~ When the value to be reported is an average of measurements that are not ~~independent.~~ independent, then the average can have a residual uncertainty that cannot be reduced by increasing the number of the measurements. This situation occurs when some ~~uncertainty~~ components of error are shared among all measurements. If standard deviations are respectively σ_1 and σ_2 for the shared components and unshared (independent for different measurements) components, the uncertainty of the average of n such correlated measurements is ~~$\sqrt{\sigma_1^2+\frac{\sigma_2^2}{n}}$.~~$\sqrt{\sigma_1^2+\frac{\sigma_2^2}{n}}$ ~~This example shows that, when measurements in a group share some sources of error, the average has a residual uncertainty that cannot be removed by increasing the number of the measurements in the group.~~.

~~6.2.3 The measurement is made by comparison against a reference material similar in kind to the sample. The data for measurements of this type are responses x and y for sample and reference material respectively, and an accepted reference value Y of the reference having standard uncertainty u_Y. The measurement result is then calculated as $X=(x/y)\times Y$. Validity of this result depends critically on response being directly proportional to the quantity, which must be demonstrated for the method. Uncertainty of the result depends on variability of responses and on the uncertainty of the reference value. Then the standard uncertainty of the determination based on responses $x_1, x_2,\ldots, x_n$ of sample and $y_1, y_2,\ldots, y_m$ of reference material is:~~