Designation: ~~D8146 – 18~~ D8146 – 22

An American National Standard

Standard Guide for Evaluating Test Method Capability and Fitness for Use[1]

This standard is issued under the fixed designation D8146; 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~~ Scope*

1.1 This guide covers techniques for evaluating the statistical capability and fitness for use of standard test methods used for measuring properties of petroleum products, liquid fuels, and lubricants. Specifically, this guide provides strategies for evaluating the capability of a test method to provide a sufficiently precise estimate of the intended parameter versus a given level or value of that parameter and for assessing, with sufficient confidence, the fitness for use of a test method for determining the acceptability of products versus specification, regulatory, or manufacturing limits.

1.2 The assessment of capability in this guide is applicable to test methods that always yield non-negative numerical results and have a hard, fixed zero lower limit or fixed upper limit (for example, 100 %). Detailed knowledge of a test method and professional judgement may be required in determining the applicability of this guide to certain test methods.

1.3 In accordance with this guide, the assessment of fitness for use of a test method versus specification, regulatory, or manufacturing limits is generally applicable to test methods that provide numerical results and have applicable reproducibility precision values.

1.4 This guide is intended for use by test method developers, specification setters, and laboratories selecting test methods to evaluate products for specific purposes or versus specific limits.

1.5 This guide is not applicable to test methods that measure temperature to determine properties such as, but not limited to, freeze point, pour point, flash point, and distillations. This guide is not intended for use by laboratories for evaluating laboratory test method execution capability, nor is it intended to evaluate manufacturing process capability.

1.6 The expressions of capability and fitness for use derived from use of this guide should not be the only criteria for selection and use of a given test method. Other factors such as the laboratory's experience with the test method, length of time to complete an analysis, typical product results relative to the regulatory or manufacturing specification limit, closeness of the results to zero, and cost factors may contribute more significantly to the decision to use a given test method for a specific application.

1.7 This guide draws on statistical approaches covered in other standards supporting petroleum products, liquid fuels, and lubricants, including Practices D3244, D6259, D6299, D6300, D6792 and ISO 4259.

1.8 *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] This guide is under the jurisdiction of ASTM Committee D02 on Petroleum Products, Liquid Fuels, and Lubricants and is the direct responsibility of Subcommittee D02.94 on Coordinating Subcommittee on Quality Assurance and Statistics.

Current edition approved ~~Oct. 1, 2018~~Nov. 1, 2022. Published ~~November 2018~~November 2022. Originally approved in 2018. Last previous edition approved in 2018 as D8146 – 18. DOI: ~~10.1520/D8146-18.~~10.1520/D8146-22.

***A Summary of Changes section appears at the end of this standard**

1.9 *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]

D3244 Practice for Utilization of Test Data to Determine Conformance with Specifications
D6259 Practice for Determination of a Pooled Limit of Quantitation for a Test Method
D6299 Practice for Applying Statistical Quality Assurance and Control Charting Techniques to Evaluate Analytical Measurement System Performance
D6300 Practice for Determination of Precision and Bias Data for Use in Test Methods for Petroleum Products, Liquid Fuels, and Lubricants
D6792 Practice for Quality Management Systems in Petroleum Products, Liquid Fuels, and Lubricants Testing Laboratories

2.2 *ISO Standard:*[3]

ISO 4259 Petroleum products—Determination and application of precision data in relation to methods of test

2.3 ASTM Standards referenced only in Appendix X2, Worked Examples, are listed in X2.3.

3. Terminology

3.1 *Definitions:*

3.1.1 *precision ratio (PR), n*—an estimate of relative magnitude of repeatability and reproducibility. The PR for a given standard test method can provide information on the relative significance between variation caused by different operators and laboratories compared to a single operator in a laboratory performing the standard test method. **D6792**

3.1.1.1 *Discussion*—

The PR for a published test method estimates the influence that non-site-specific variations have on the published precision; the PR metric essentially judges the adequacy of between-laboratory standardization, or the agreement between laboratories relative to within-lab repeatability precision.

3.1.2 *repeatability (r), n*—the quantitative expression for the random error associated with the difference between two independent results obtained under repeatability conditions that would be exceeded with an approximate probability of 5 % (1 case in 20 in the long run) in the normal and correct operation of the test method. **D6300**

3.1.2.1 *Discussion*—

This definition is applicable to precision values obtained following the practices described in D6300. Repeatability determined using different or modified practices may not have the same statistical significance as that inferred when following D6300.

3.1.3 *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. **D6300**

3.1.4 *reproducibility (R), n*—a quantitative expression for the random error associated with the difference between two independent results obtained under reproducibility conditions would be exceeded with an approximate probability of 5 % (1 case in 20 in the long run) in the normal and correct operation of the test method. **D6300**

3.1.4.1 *Discussion*—

This definition is applicable to precision values obtained following the practices described in D6300. Repeatability determined using different or modified practices may not have the same statistical significance as that inferred when following D6300.

3.1.5 *reproducibility conditions, n*—conditions where independent test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment. Note that different laboratory by necessity means a different operator, different equipment, and different location and under different supervisory control. **D6300**

3.1.6 *site precision (R'), n*—the value below which the absolute difference between two individual test results obtained under site

[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.

precision conditions may be expected to occur with a probability of approximately 0.95 (95 %). It is calculated as 2.77 times the standard deviation ($\sigma_{R'}$) of results obtained under site precision conditions. **D6299**

3.1.6.1 *Discussion—*

Site precision is often estimated in the setup of a quality control chart, and the TPI can be used to evaluate the performance of the test method in the laboratory as described in D6792.

3.1.7 *site precision conditions, n—for a single analytical measurement system,* conditions under which test results are obtained by one or more operators in a single site location practicing the same test method on a single measurement system, ~~which may comprise multiple instruments,~~ using test specimens taken at random from the same sample of material over an extended period of time spanning at least a ~~15-day~~ 20 day interval.

3.1.7.1 *Discussion—*

For an in-depth discussion, see Practice D6299. **D6299**

3.1.8 *test performance index (TPI), n*—an approximate measure of a laboratory's testing capability, defined as the ratio of test method reproducibility to site precision. **D6792**

3.1.8.1 *Discussion—*

The term capability as used in this definition refers to ability of the laboratory performing the test method relative to the published reproducibility and not to the capability of the test method with respect to application at a given level or value.

3.1.8.2 *Discussion—*

The TPI term is similar to precision ration (PR = R/r), but the use of site precision instead of repeatability provides a direct evaluation of the laboratory's performance of the test method versus reproducibility. Since there is a general expectation that R > R' > r, then generally it would be expected that TPI < PR.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *analytical performance value (APV), n*—a metric that assesses the precision of a test method relative to applicable specification limits. APV is the ratio of the repeatability or reproducibility of a test method to specific limits, expressed as a percent.

3.2.1.1 *Discussion—*

The APV is the percent relative repeatability or reproducibility.

3.2.2 *capability, n*—as applied to this guide, an estimate of the ability of a test method to provide a precise estimate of the intended parameter at a given level or value of that parameter.

3.2.2.1 *Discussion—*

APV is the metric used to express capability.

3.2.3 *fit for use, n*—as applied to this guide, an expression of the adequacy of a test method to provide a precise estimate of the intended parameter with a desired level of confidence at the level of a corresponding specification, regulatory, or manufacturing limit to support compliance decision.

3.2.3.1 *Discussion—*

Fit for use is the effectiveness of an analytical measurement system to adequately meet the precision requirements for the intended application relative to specific limits.

3.2.3.2 *Discussion—*

Also known as *fitness for use*.

4. Significance and Use

4.1 This guide provides the developers of test methods with systematic approaches to evaluate the inherent adequacy (capability) of the precision and the adequacy of between-laboratory standardization as determined by interlaboratory testing in accordance with Practice D6300 for a new or modified test method.

4.1.1 The tools presented in this guide can be used to assess the adequacy of the repeatability and reproducibility determined for a given test method relative to their intended applications. For those test methods with less than optimal precision, the responsible subcommittees would have an opportunity to investigate, suggest improvements in the procedure, and design an interlaboratory study as necessary. It is important to note that non-optimal precision does not necessarily imply that the test method is invalid, but that the test method may be less capable for an intended application.

4.2 This guide is applicable to test method users to evaluate the relative performance of different test methods that purport to measure the same property as applied to a specific application of the test methods.

4.3 This guide can be used to evaluate the fitness for use of a test method reproducibility relative to its intended use in specifications.

4.3.1 When test method results are intended to be compared against specification limits, this guide provides the tools to assess whether the test method's reproducibility is expected to be fit for use relative to the stated limits.

4.4 This guide is most effective when used in conjunction with sufficient knowledge of the chemistry, instrumental technology, metrology principles, sample characteristics, and a fundamental understanding of statistical meaning and implications associated with repeatability and reproducibility of the test method being evaluated.

4.5 This guide does not cover the qualitative and business factors that might also be considered in judging a test method's capability or fitness for use versus a specific application.

5. Test Method Capability

5.1 Test method capability can be evaluated using Analytical Performance Values (APV), Precision Ratio (PR), and Test Performance Index (TPI).

5.2 *APV:*

5.2.1 Determine the Analytical Performance Value (APV) by dividing the test method precision, either reproducibility (R) or repeatability (r), by the test result or target level and multiplying by 100. Express the result to the nearest percent as appropriate. For test methods that do not have a published R, then a laboratory's site precision (R') can be used to estimate the APV_R. Report APV with reference to the target value, for example APV_R = 20 % (at 50 mg/kg).

$$\text{APV-Reproducibility } (APV_R) = (R/X) \times 100 \quad (1)$$

$$\text{APV-Repeatability } (APV_r) = (r/X) \times 100 \quad (2)$$

Where X is the target level, specification, regulatory limit, a test result, or average from multiple test results.

5.2.2 From a capability perspective, numerically smaller APV are preferred versus larger ones. The APV value depends on the intended use of the test method. As APV decreases (r or R decrease relative to the target level) the test method would appear to be more capable for the intended application. Conversely, as the APV increases (r or R increase to approach or exceed the target level) the test method would appear to be less capable for the intended application.

5.2.2.1 It is up to the user to determine at what point higher APV values indicate a potential issue with the use of the test method for the intended application.

5.2.2.2 Other factors such as availability of alternative test methods, proximity of the target to zero or to a detection limit, critical or non-critical nature of specification parameter, or other factors may influence this determination.

5.2.3 It is important to evaluate APV_r and APV_R over the entire range of application of the test method. The precision performance of some test methods tends to degrade when used at or close to their lower limits. (See Appendix X2.)

5.2.4 It is useful to compare the APV_r and APV_R of new or modified test methods to that for existing similar test methods to determine the significance of any improvements in precision relative to the intended applications.

5.3 *APVr at Lower Limits:*

5.3.1 When evaluated at the lower limit of the test method, the APV_r metric assesses the capability of the test method to be repeated in an individual laboratory relative to that lower level. As stated above, when test methods have r values that are less than or much less than the target level (that is, APV_r is low), then the assessment of capability is more favorable. Examples are shown in Appendix X2.

5.3.2 It is generally expected that APV_r should be less than about 28 % for a capable test method at its lowest limit.

NOTE 1—The 28 % APV_r limit is related to limit of quantitation determinations as described in D6259.

5.4 *Precision Ratio (PR):*

5.4.1 The Precision Ratio (PR) is the ratio of the test method's reproducibility (R) to its repeatability (r), calculated to the nearest integer or to the nearest 0.1 when PR < 1.

$$\text{Precision Ratio (PR)} = R / r \tag{3}$$

5.4.2 The adequacy of the between-lab standardization of a test method generally can be evaluated using the PR as shown in Table 1.

5.4.3 Test methods with PR > 4 generally indicate a significantly large difference between repeatability and reproducibility, which means that the between-laboratory bias is the dominant contributor towards reproducibility. The implication is that the test method is not adequately standardized and further work on the test method should be considered.

5.4.4 When PR is small and in the range of 1 to 2 (that is, as r approaches R) then it could imply that the contribution of between-laboratory bias towards reproducibility is small or insignificant relative to the within-lab repeatability. When PR < 1 (that is, as r exceeds R) then it could imply that the repeatability is so large that it masks the between-laboratory common causes. For this scenario, further work on improving the test method repeatability should be considered.

5.4.4.1 To put smaller (< 2) PR values into proper perspective, they should be considered along with the corresponding APV_r values, which then accounts the intended application of the test method. When PR is < 2 (that is, as r approaches R) and the APV_r is < 28 % (that is, as r is much smaller relative to the target value) then R also is small relative to the target value and the test method may be useful at the target value. When PR < 2 and the APV_r is >> 28 % (that is, r is approaching the target value), then R is also approaching the target value and the test method may have limited usefulness at the level of intended use.

5.4.5 It is useful to compare the PR obtained for a new or modified test method to that for existing similar test methods to determine whether there are significant improvements.

5.5 *PR and TPI:*

5.5.1 The PR can be used in conjunction with the Test Performance Index (TPI) as discussed in D6792. PR and TPI are used to suggest minimum QC sample testing frequencies, and the PR is used to recommend ranges for determining acceptable versus unacceptable TPI values obtained from the site precision standard deviation ($\sigma_{R'}$) generated for control charts.

$$\text{Test Performance Index} = \text{Test Method Reproducibility} / \text{Site Precision} \tag{4}$$

Where site precision = 2.77 × $\sigma_{R'}$

6. Test Method Fit for Use

6.1 Test method fitness for use is evaluated by comparing the test method reproducibility to the span or range of the specification, regulatory, or manufacturing limits.

TABLE 1 Evaluation of Test Method Reproducibility Using PR and Not Considering the Level of Intended Use of the Test Method

PR	Adequacy of Reproducibility
≤ 1	Within-laboratory common cause variability is so large that it may be masking the between-laboratory common causes.
1 to 2	Between-laboratory common cause variability is small or insignificant relative to the within-laboratory common cause variability.
2 to 4	Standardization conditions appear to adequately control between-laboratory common cause variability.
4 to 10	Standardization conditions to control between-laboratory common cause variability appear to be less than optimal.
>10	Between-laboratory bias appears to be the dominant contributor towards reproducibility and that the test method does not appear to be sufficiently standardized.