Designation: D4467 – 94 (Reapproved 2001)

Standard Practice for Interlaboratory Testing of a Textile Test Method That Produces Non-Normally Distributed Data[1]

This standard is issued under the fixed designation D4467; 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 practice covers design and analysis of interlaboratory testing of a test procedure in the case where the resulting test data are discrete variates or are continuous variates not normally distributed. This practice applies to all such interlaboratory tests used to validate a test procedure.

1.2 Analysis of interlaboratory test results permits validation that the process of using the test method is in statistical control and provides the information required to write statements on precision and bias as directed in Practice D2906. It also gives the information for determining the number of specimens per unit in the laboratory sample as required in Practice D2905.

1.3 Precision statements for non-normally distributed data can be written as a function of the level of the property of interest without an interlaboratory test if the underlying distribution is known and statistical control can be assumed.

1.4 If the underlying distribution is unknown, the precision of the test method can only be approximated. There are no generally accepted methods of making approximations of this sort.

1.5 If statistical control cannot be assumed, then a meaningful precision statement cannot be written and the test method should not be used.

1.6 This practice is intended for use with data from test methods that cannot be properly modeled by a normal distribution. See Practices D2904 and E691 for applications that can be modeled by a normal distribution.

1.7 This practice includes the following sections:

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 whoever uses this standard to consult and establish appropriate safety and health practices and determine the applicability of regulatory limitations prior to use.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

D123 Terminology Relating to Textiles

D2904 Practice for Interlaboratory Testing of a Textile Test Method that Produces Normally Distributed Data

D2905 Practice for Statements on Number of Specimens for Textiles[3]

D2906 Practice for Statements on Precision and Bias for Textiles[3]

D4646 Test Method for 24-h Batch-Type Measurement of Contaminant Sorption by Soils and Sediments

D4853 Guide for Reducing Test Variability[3]

E456 Terminology Relating to Quality and Statistics

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

E1169 Practice for Conducting Ruggedness Tests

3. Terminology

3.1 *Definitions:*

3.1.1 *test method*, *n*—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.

3.1.2 For definitions of textile and statistical terms used in this practice and discussions of their use, refer to Terminology D123, and Terminology E456.

[1] This practice is under the jurisdiction of ASTM Committee D13 on Textiles and is the direct responsibility of Subcommittee D13.93 on Statistics.

Current edition approved June 15, 1994. Published August 1994. Originally published as D4467 – 85. Last previous edition D4467 – 85. DOI: 10.1520/D4467-94R01.

[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] Withdrawn. The last approved version of this historical standard is referenced on www.astm.org.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *assignable cause*—a factor which contributes to variation and is feasible to detect and identify.

3.2.2 *interlaboratory testing*—the evaluating of a test method in more than one laboratory by analyzing data obtained from one or more materials that are as homogeneous as practical.

3.2.3 *random cause*—one of many factors which contribute to variation but which are not feasible to detect and identify since they are random in origin and usually small in effect.

3.2.4 *state of statistical control*—a condition in which a process, including a measurement process, is subject only to random variation.

4. Significance and Use

4.1 The planning of interlaboratory tests requires a general knowledge of statistical principles. Interlaboratory tests should be planned, conducted, and analyzed after consultation with statisticians who are experienced in the design and analysis of experiments and who have some knowledge of the nature of the variability likely to be encountered in the test method.

4.2 The instructions of this practice are specifically applicable to the design and analysis of the following tests:

4.2.1 Pilot-scale interlaboratory tests and

4.2.2 Full-scale interlaboratory tests.

4.3 Procedures given in this practice are applicable to methods based on the measurement of the following types of variates:

4.3.1 Ratings (grades or scores), such as those resulting from comparisons with AATCC gray scales,

4.3.2 Percent of observations with a specific attribute,

4.3.3 Counts of attributes, such as number of nonconformities,

4.3.4 Any data not normally distributed which the analyst cannot or prefers not to transform, such as flammability data or percent extractables.

4.4 Interlaboratory testing is a means of determining the consistency of results when the same material is tested under varying conditions such as: operators, laboratories, equipment, or environment. An interlaboratory test should do the following:

4.4.1 Show if the test method distinguishes between levels of the property being tested,

4.4.2 Show if the test method is in statistical control; statistical control being the presence of only random variation,

4.4.3 Detect operators, laboratories, and equipment out of statistical control.

4.5 An initial single-laboratory preliminary test of a test procedure is necessary, usually including ruggedness testing, to determine the feasibility of the method and to determine the method's sensitivity to variables which must be controlled. See Guides D4853 or E1169 for a discussion of ruggedness testing.

4.6 A pilot-scale interlaboratory test may be needed to identify sources of variation, to establish clarity of instructions of the proposed operating procedures, and to obtain estimates as to the number of test results per operator per material to be used in the initial full-scale interlaboratory test.

4.7 A full-scale interlaboratory test is usually made after a pilot-scale test. If the task group prefers, a full-scale test may be run without a previous pilot-scale test but with the understanding that unsatisfactory results would require another full-scale test.

4.8 Interlaboratory tests of the type discussed in this practice are used to locate and measure the sources of variability associated with a test method when the test method is used to evaluate a property of one or more materials, each of which is as homogeneous as practical with respect to that property. Such interlaboratory tests provide no information about the sources of variability associated with the sampling of the stream of product from a manufacturing process, a shipment, or material in inventory. Estimation of such sampling errors requires an entirely different type of experiment which is not specified presently in an ASTM Committee D-13 standard.

5. General Considerations

5.1 *Overview*—This section covers various aspects of allocating specimens to the participating laboratories.

5.2 *Sampling of Materials*—Select a source of samples of material in such a way that any one portion of the material, within which laboratories, operators, days, and other factors are to be compared, will be as homogeneous as possible with respect to the property being measured. Otherwise, increased replication will be required to reduce the size of the difference which can be detected.

5.3 *Randomization of Specimens*:

5.3.1 *Complete Randomization*—Randomize the selection of specimens for each laboratory sample; divide all the randomized specimens of a specific material, after labeling, into the required number of groups, each group corresponding to a specific laboratory.

5.3.2 *Stratification*—In some cases it is advantageous to follow a stratified pattern in the allocations of the specimens to laboratories. For example, if the specimens are bobbins of yarn from different spinning frames, it is better to allocate to each laboratory equal numbers of specimens from each spinning frame. In such cases, the specimens within each spinning frame are randomized separately rather than all of the specimens from all of the frames.

5.4 *Order of Tests*—In many situations, variability among replicate tests is greater when measurements are made at different times than when they are made together as part of a group. Sometimes trends are apparent among results obtained consecutively. Furthermore, some materials undergo measurable changes within relatively short storage periods. For these reasons, treat the dates of testing, as well as the order of tests carried out in a group as controlled, systematic variables.

5.5 *Selecting the Measure of Average Performance*—Data are summarized for presentation and analysis by use of some measure of typical performance. For textile testing, there are usually three choices:

5.5.1 *Arithmetic Average*—The arithmetic average is the measure of choice when the data are symmetrically distributed or are from a Poisson distribution.

5.5.2 *Median*—The median (midpoint, fiftieth percentile) is the preferred measure when the data are asymmetrically distributed. When the distribution is symmetrical, the arithmetic average and the median are equal.

5.5.3 *Proportion*—A proportion, which may be expressed as a fraction (decimal) or percent, is the measure to use when the data are counts of items having a particular attribute out of a specified number of items.

5.6 *Number of Replicate Specimens*—The number of specimens tested by each operator in each laboratory for each material may be calculated from previous information or from a pilot run. This number of specimens or replications (minimum of two) depends on the relative size of the random error and the smallest effect to be detectable. A replicate consists of one specimen of each condition and material to be tested in the statistical design.

5.6.1 *Symmetrical Non-Normal Distributions*—Calculate the number of observations required in each mean using Eq 1 (Note 1):

$$n = (\tau s/E)^2 = 16(s/E)^2 \quad (1)$$

where:

n = number of observations in each mean,
τ = 4 = specified value in Tchebychev's inequality (Note 2),
s = standard deviation for individual observations obtained from previously conducted studies, and
E = smallest difference it is of practical importance to detect, expressed in the same units of measure as the averages and standard deviation.

NOTE 1—With a balanced design, half of the total observations in the experiment will be in each of the two sample means used to determine the possible effect of each factor being evaluated at two levels; one third of the total observations will be in each of the three sample means used to determine the possible effect of each factor being evaluated at three levels; and so on. The required value of n refers to such means.

NOTE 2—Tchebychev's inequality states that in all cases at least $(1 - 1/\tau)$ of the total observations, n, will lie within the closed range $\bar{x} \pm \tau\sigma$, when τ is not less than 1. For $\tau = 4$, at least 93.75 % of all observations will fall within $\bar{x} \pm 4\sigma$. For symmetrical distributions, the observed percentage is usually well above the minimum percentage specified by Tchebychev's inequality.

5.6.2 *Asymmetrical Distribution Except Poisson or Binomial*—Calculate the number of observations required in each mean using Eq 2 (Note 2):

$$n = (1.25\tau s/E)^2 = 25(s/E)^2 \quad (2)$$

where the terms in the equation are as defined in 5.6.1.

5.6.3 *Poisson Distributions*—Calculate the number of observations required in each mean using Eq 3 (Note 2):

$$n = a(t/E)^2 = 9a/E^2 \quad (3)$$

where:

t = 3 = specified value of Student's t,
a = total number of occurrences, and where the other terms in the equation are as defined in 5.6.1.

5.6.4 *Binomial Distributions*—Calculate the number of observations required in each mean using Eq 4 (Note 2):

$$n = p(1 - p)(t/E)^2 = 9p(1 - p)/E^2 \quad (4)$$

where:

t = 3 = specified value of Student's t,
p = proportion of the observations having a specific attribute, expressed as a decimal fraction, and

where the other terms in the equation are as defined in 5.6.1.

5.7 *Gain of Statistical Information*—More statistical information can be obtained from a small number of determinations on a large number of materials than from the same total number of determinations distributed over fewer materials. In the same way, a specific number of determinations per material will yield more information if they are spread over the largest number of laboratories possible. For the recommended minimum design, see 6.2. If experience with the pilot-scale interlaboratory test casts doubt on the adequacy of the starting design, estimate the number of determinations needed to detect the smallest differences of practical importance.

5.8 *Multiple Equipment (Instruments)*—When multiple instruments within a laboratory are used on an interlaboratory test, tests should be made on all equipment to establish the presence or absence of the equipment effects. All types of equipment allowed by a test method should be tested to allow greatest flexibility. If an equipment effect is present and cannot be eliminated by use of pertinent scientific principles, known standards should be run and appropriate within-laboratory quality control procedure should be used.

6. Basic Statistical Design

6.1 It is advisable to keep the design as simple as possible, yet to obtain estimates of within- and between-laboratory variation unconfounded with secondary effects. Provisions also should be made for estimates of significance of variation due to: materials-by-laboratories interactions, and operators-by-materials interactions.

6.2 Include in the basic statistical design the following:

6.2.1 A minimum of three materials spanning the range of interest for the property being measured,

6.2.2 At least ten laboratories unless the test method cannot be used in that many laboratories,

6.2.3 A recommended minimum of two operators per laboratory, and

6.2.4 At least two specimens of each material to be tested by each operator in a designated random order.

6.3 The laboratory report format is presented in Table 1.

6.4 Select materials to produce a wide range of expected results. The materials should include the applicable physical forms. For example, if woven fabric, knit fabric, and nonwoven fabric can all be tested by the method, these materials should each be represented over a wide range of values.

6.5 An illustrative example of a full-scale interlaboratory design and its analysis is shown in Annex A1.

7. Pilot-Scale Interlaboratory Test

7.1 Plan a pilot-scale interlaboratory test by preparing a definitive statement on the type of information the task group expects to obtain from the interlaboratory test, including the statistical analyses.

7.2 Conduct a pilot study using two or three materials of established values (low, medium, and high values of the

TABLE 1 Interlaboratory Test of Pilling Resistance: Random Tumble Method (ASTM D3512 – 82)

Pilling Ratings Laboratory I										
Sample	Specimen	Material								Overall
		A		B		C		D		
		operator		operator		operator		operator		
		a	b	a	b	a	b	a	b	
1	1	2.5	3.0	3.0	1.5	3.5	5.0	4.5	5.0	...
	2	2.5	3.5	2.5	1.5	4.5	5.0	5.0	5.0	...
	3	2.5	2.5	2.0	1.5	4.0	5.0	4.0	5.0	...
AVERAGE		2.5	3.0	2.5	1.5	4.0	5.0	4.5	5.0	...
2	1	3.0	3.5	3.0	1.0	4.0	5.0	5.0	5.0	...
	2	4.0	3.5	3.0	1.0	4.0	5.0	5.0	5.0	...
	3	3.5	3.5	3.0	1.0	4.0	5.0	5.0	5.0	...
		3.5	3.5	3.0	1.0	4.0	5.0	5.0	5.0	...
AVERAGE										
Averages										
Operator a/Material		3.00	...	2.75	...	4.00	...	4.75	...	3.62
Operator b/Material		...	3.25	...	1.25	...	5.00	...	5.00	3.62
Material		3.12		2.00		4.50		4.88		3.62

property under evaluation) in preferably three to four laboratories. A recommended minimum of two operators per laboratory should each test a minimum of two specimens per material.

7.3 Based on the data on a single-laboratory preliminary test, prepare the design plan and circulate it to all task group members and all other competent authorities for review and criticism. Also include examples of suggested materials that cover the range of property to be measured and that represent all classes of the material for which the method will be used. Revise the plan for the pilot-scale test as required by this review.

7.4 Conduct a pilot-scale interlaboratory test using the design plan.

7.5 Analyze the data from the plan described in 7.3 as directed in Annex A1.

7.6 On the basis of the data analysis from the pilot run, and comments from the cooperating laboratories, revise instructions and procedures to minimize operator and instrument variation to the extent practicable.

8. Full-Scale Interlaboratory Tests

8.1 After a thorough review of procedural instructions and evaluations of pilot run data as specified in Section 7, canvass the potential participating laboratories to ascertain the number and extent of participation in a full-scale test. If practicable, secure at least ten laboratories unless the test method cannot be used in that many laboratories. Have each laboratory test a series of materials, using two operators per laboratory and two or more specimens per operator per material.

8.2 Prepare a definitive statement of the type of information the task group expects to obtain from the interlaboratory test, including the statistical analyses.

8.3 Obtain adequate quantities of a series of homogeneous materials covering the general range of values normally expected to be encountered for the test method. For distribution to each participating laboratory, divide the available quantity of homogeneous material into sampling units (specimens), and select the appropriate number for each laboratory by simple random sampling. From each material, allocate enough samples to provide for all participating laboratories and a sufficient number of additional samples for replacement of lost or spoiled samples. Label each specimen by means of a code symbol and record the coded identification of the specimens for further reference. Store and maintain reserve specimens in such a manner that the characteristic being studied does not change with time. If specimens are to be prepared and distributed, observe the same precautions. See 5.3 for sampling procedures.

8.4 Analyze the data from the plan described in 8.2 as directed in Annex A1.

9. Missing Data

9.1 Occasionally, when conducting interlaboratory tests, accidents may result in the loss of data. In such an event use reserve samples or specimens, if at all possible. If reserves are not available, a valid analysis of the data with missing items can be made by use of the theory behind the methods of calculation. Consult a statistician for calculation procedures when data are missing.

10. Outlying Observations

10.1 Retain all test data. Data should be excluded from reporting only when assignable causes for deletion of a test value are present. Examples of assignable causes are: the operator observed some instrument malfunction, specimen preparation error, or other circumstance that should logically result in the termination of the test procedure at that specific point. In cases where there is no assignable cause for an apparent outlier, the test value should be reported. In cases where there is an assignable cause, test a reserve and report the assignable cause that justified the use of the reserve specimen.

11. Interpretation of Data

11.1 If the difference between laboratories is significant as determined by using Annex A1, examine and decide which laboratory or laboratories contributed to the significant laboratory difference. On the basis of this information, ascertain actual test conditions and instrument setups that may have contributed to these significantly different laboratories.

11.2 A significant laboratory-by-material interaction means that materials may be ranked in significantly different response magnitudes or different orders by different laboratories. Since a significant laboratory-by-material interaction might arise from poorly written instructions, reevaluate procedural instructions and instrument set ups. After such evaluation, it is likely that the interlaboratory test will need to be repeated in order to obtain the objective of determining the precision of the test method.

11.3 Where significant between-operator-within-laboratory differences occur, reevaluate procedural instructions and examine operator techniques to find differences in preparation or in procedures, or both. The task group must determine if the interlaboratory test should be repeated.

12. Plotting Results

12.1 Graphs aid in presenting the results, but conclusions about the significance of differences should be based on the analyses made as directed in Annex A1. Plots of interest include the following:

12.1.1 On a separate graph for each laboratory, plot the averages for each material. An example is shown in Fig. 1.

12.1.2 On a separate graph for each material, plot the averages for each laboratory where an average can be calculated. An example is shown in Fig. 2.

12.1.3 On a separate graph for each operator within each laboratory, plot the averages for each material. An example is shown in Fig. 3.

12.1.4 On a separate graph for each laboratory having more than one operator reporting results, plot the averages for each operator from each material. An example is shown in Fig. 4.

12.1.5 On one graph, representing each laboratory with a separate line, plot the averages for each material. An example is shown in Fig. 5.

12.1.6 On one graph, representing each material with a separate line, plot the averages for each laboratory. An example is shown in Fig. 6.

12.1.7 On one graph, combining results from all laboratories, plot the averages for each material. An example is shown in Fig. 7.

12.1.8 On one graph, combining results from all materials, plot the averages for each laboratory. An example is shown in Fig. 8.

13. Keywords

13.1 discrete data; interlaboratory testing; non-normally distributed data; precision; statistics

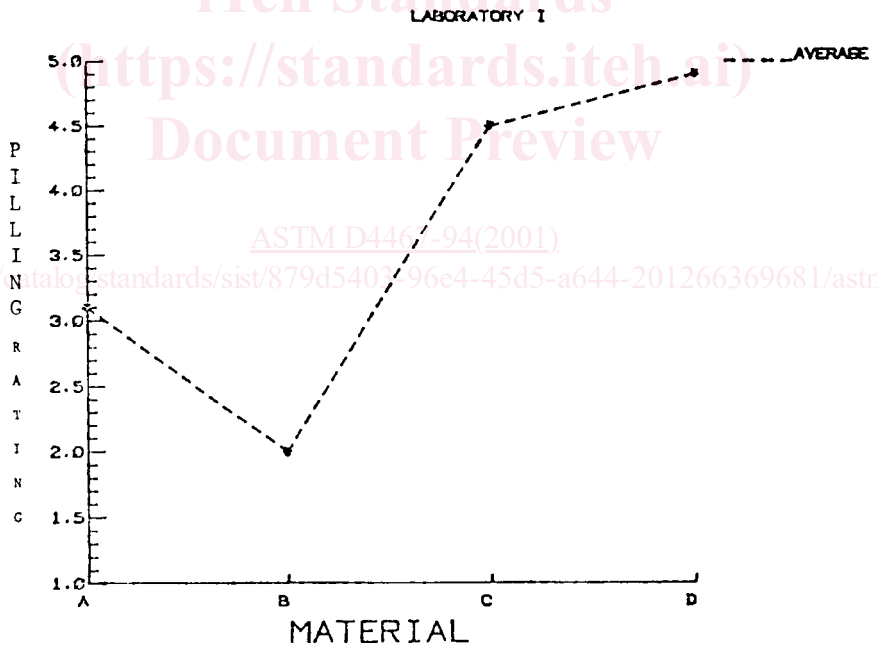

FIG. 1 Interlaboratory Test of Pilling-Resistance—Random Tumble Method (ASTM D3512 – 82)

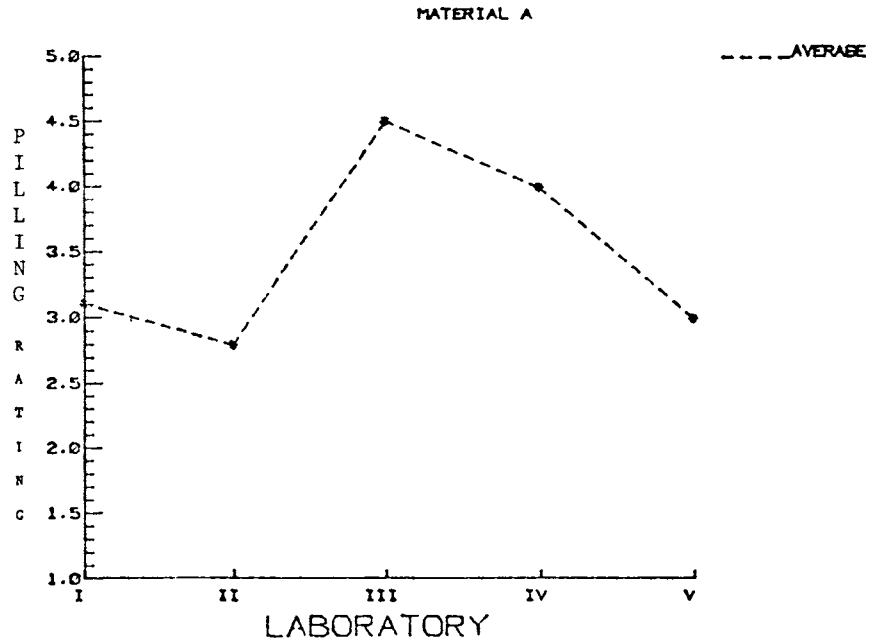

FIG. 2 Interlaboratory Test of Pilling Resistance—Random Tumble Method (ASTM D3512 – 82)

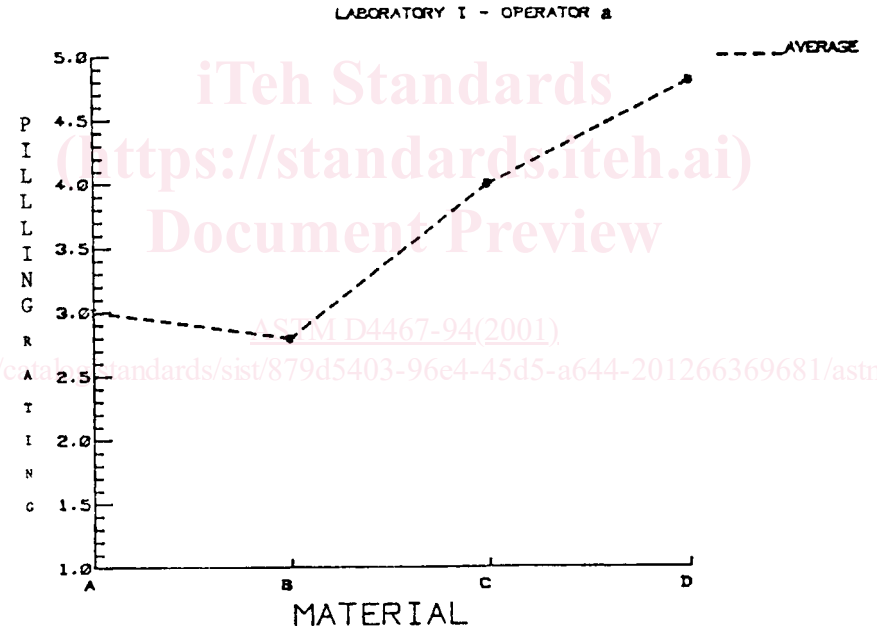

FIG. 3 Interlaboratory Test of Pilling Resistance—Random Tumble Method (ASTM D3512 – 82)