



# Standard Practice for Interlaboratory Testing of a Textile Test Method That Produces Non-Normally Distributed Data<sup>1</sup>

This standard is issued under the fixed designation D 4467; 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 D 2906. It also gives the information for determining the number of specimens per unit in the laboratory sample as required in Practice D 2905.

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 D 2904 and E 691 for applications that can be modeled by a normal distribution.

1.7 This practice includes the following sections:

	Sections
Scope	1
Referenced Documents	2
Terminology	3
Significance and Uses	4
General Considerations	5
Basic Statistical Design	6
Pilot-Scale Interlaboratory Test	7
Full-Scale Interlaboratory Test	8
Missing Data	9
Outlying Observations	10
Interpretation of Data	11

Plotting Results	12
Keywords	13
Pilot-Scale and Full-Scale Interlaboratory Tests	Annex A1
Calculation of Chi-Square	Annex A2

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:

- D 123 Terminology Relating to Textiles<sup>2</sup>
- D 2904 Practice for Interlaboratory Testing of a Textile Test Method that Produces Normally Distributed Data<sup>2</sup>
- D 2905 Practice for Statements on Number of Specimens for Textiles<sup>2</sup>
- D 2906 Practice for Statements on Precision and Bias for Textiles<sup>2</sup>
- D 4646 Test Method for 24-h Batch-Type Measurement of Contaminant Sorption by Soils and Sediments<sup>3</sup>
- D 4853 Guide for Reducing Test Variability<sup>4</sup>
- E 456 Terminology Relating to Quality and Statistics<sup>5</sup>
- E 691 Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method<sup>5</sup>
- E 1169 Guide for Conducting Ruggedness Tests<sup>5</sup>

## 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 D 123, and Terminology E 456.

### 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

<sup>1</sup> 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 D 4467 – 85. Last previous edition D 4467 – 85.

<sup>2</sup> *Annual Book of ASTM Standards*, Vol 07.01.

<sup>3</sup> *Annual Book of ASTM Standards*, Vol 11.04.

<sup>4</sup> *Annual Book of ASTM Standards*, Vol 07.02.

<sup>5</sup> *Annual Book of ASTM Standards*, Vol 14.02.

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 D 4853 or E 1169 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,
- $\tau$  = 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  $(1 - 1/\tau^2)$  of the total observations,  $n$ , will lie within the closed range  $\bar{x} \pm \tau\sigma$ , when  $\tau$  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 non-woven 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 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

**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	...
AVERAGE		3.5	3.5	3.0	1.0	4.0	5.0	5.0	5.0	...
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

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

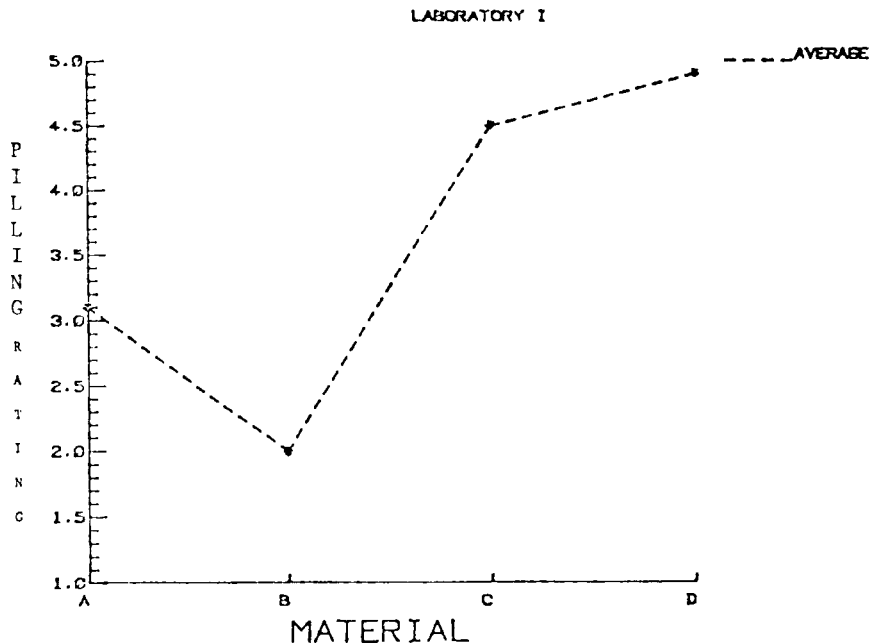


FIG. 1 Interlaboratory Test of Pilling-Resistance—Random Tumble Method (ASTM D 3512 – 82)

MATERIAL A

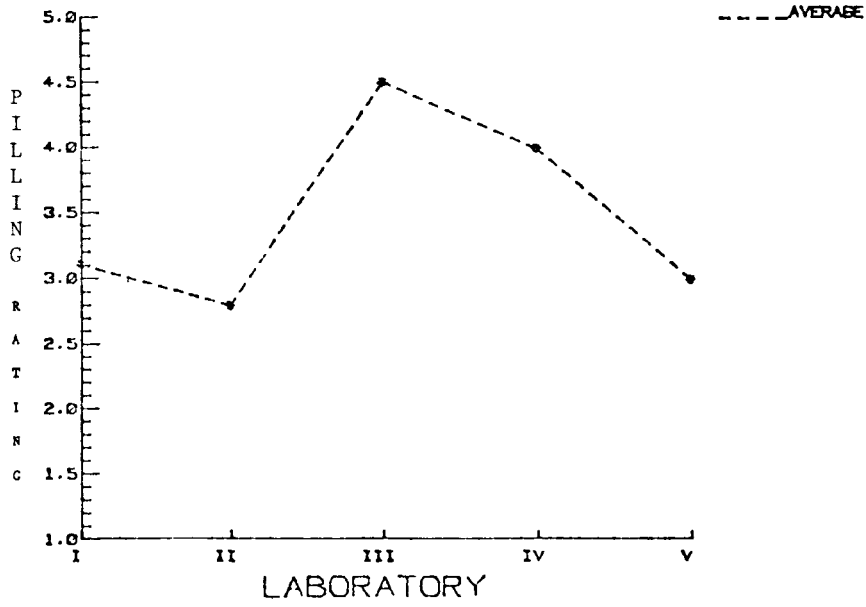


FIG. 2 Interlaboratory Test of Pilling Resistance—Random Tumble Method (ASTM D 3512 – 82)

LABORATORY I - OPERATOR a

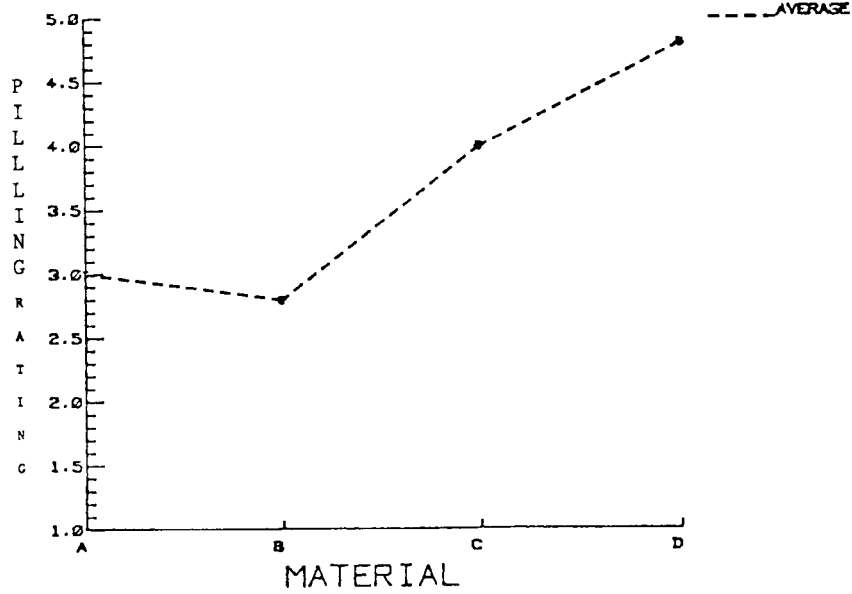


FIG. 3 Interlaboratory Test of Pilling Resistance—Random Tumble Method (ASTM D 3512 – 82)

LABORATORY I - OPERATOR X MATERIAL INTERACTION

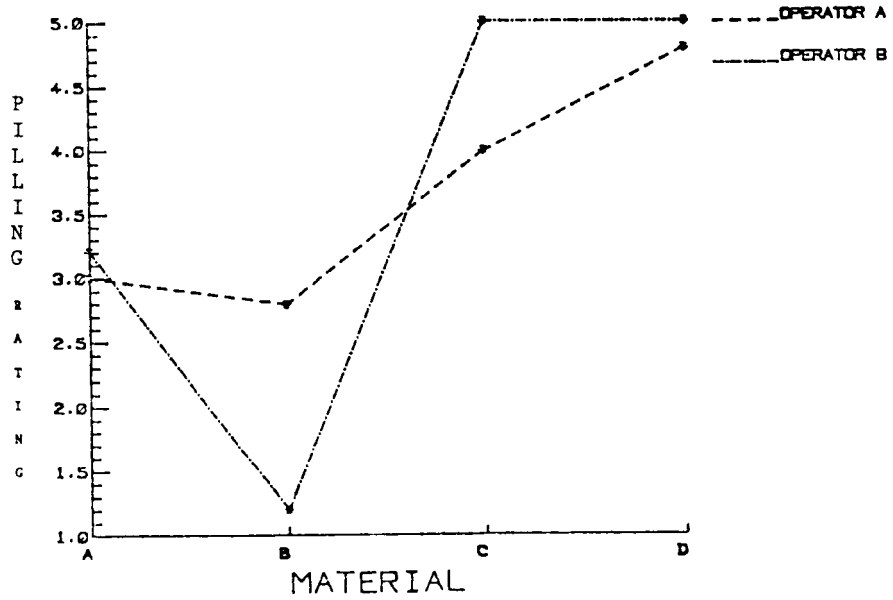


FIG. 4 Interlaboratory Test of Pilling Resistance—Random Tumble Method (ASTM D 3512 – 82)

BETWEEN LABORATORY DIFFERENCES

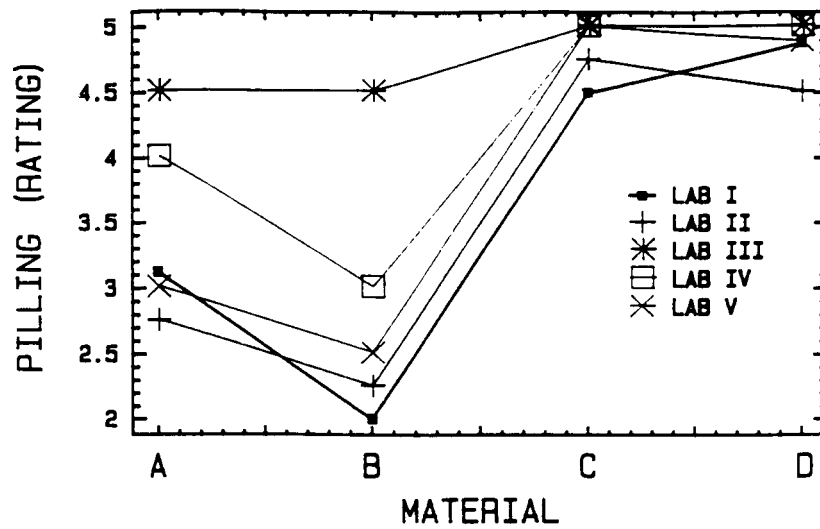


FIG. 5 Interlaboratory Test of Pilling Resistance—Random Tumble Method (ASTM D 3512 – 82)

BETWEEN MATERIAL DIFFERENCES

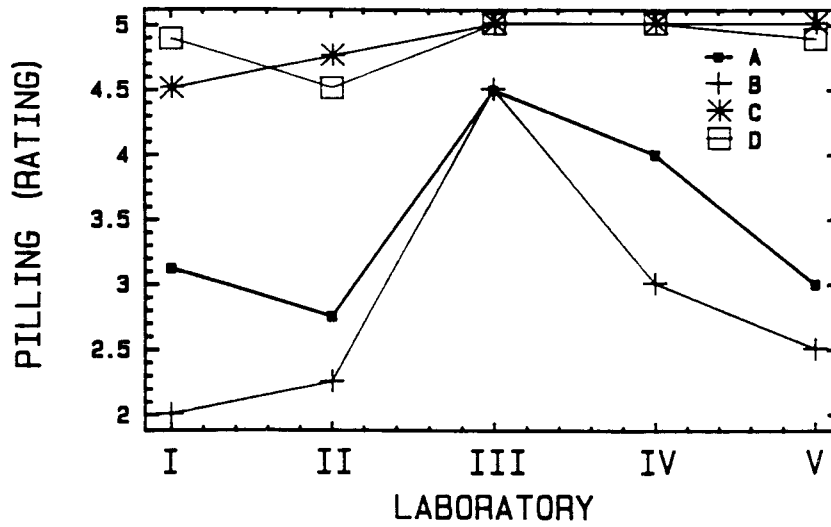


FIG. 6 Interlaboratory Test of Pilling Resistance—Random Tumble Method (ASTM D 3512 – 82)

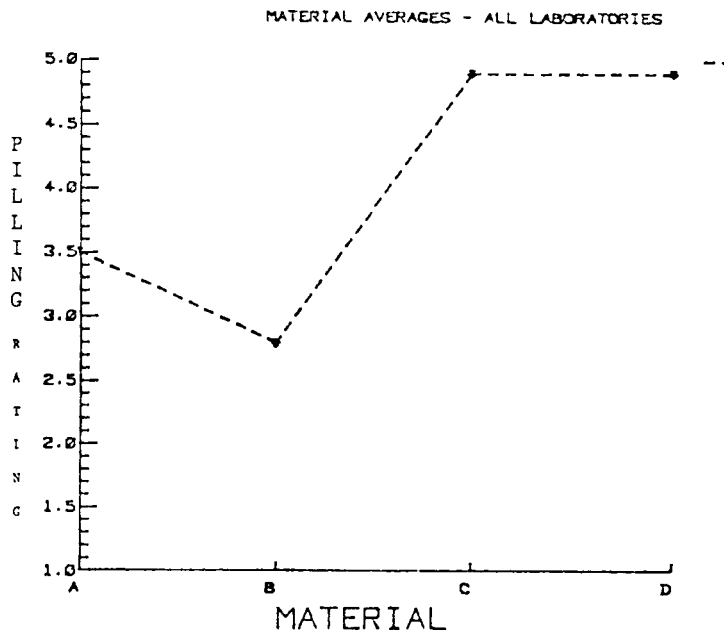


FIG. 7 Interlaboratory Test of Pilling Resistance—Random Tumble Method (ASTM D 3512 – 82)



LABORATORY AVERAGES - ALL MATERIALS

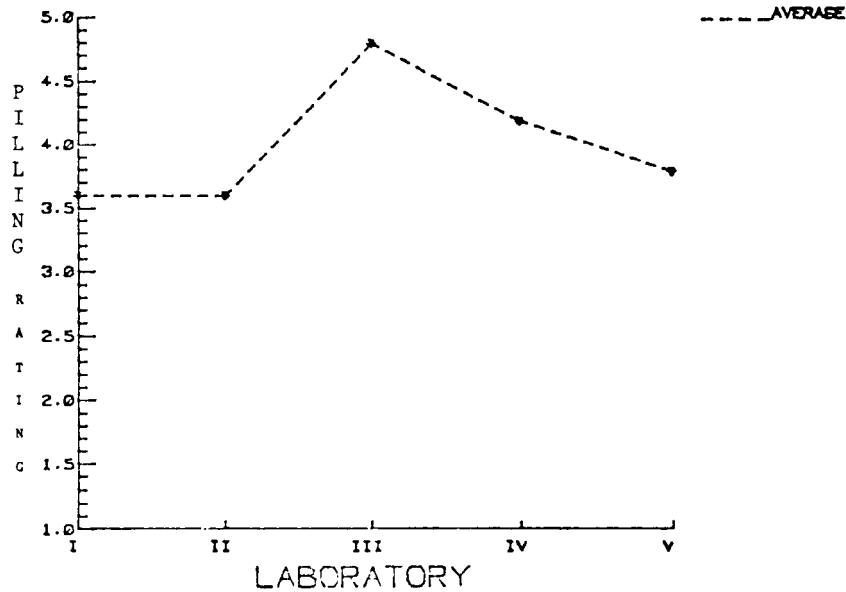


FIG. 8 Interlaboratory Test of Pilling Resistance—Random Tumble Method (ASTM D 3512 – 82)

ANNEXES

(Mandatory Information)

A1. PILOT-SCALE AND FULL-SCALE INTERLABORATORY TESTS

A1.1 After conducting the preliminary single-laboratory trial, a pilot-scale interlaboratory test may be needed. The methods of statistical analysis of the results from a pilot-scale test are the same as those used for analysis of the results from a large-scale test. 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.

A1.2 Complete factorial designs are used for full-scale interlaboratory tests. All laboratories test all materials; therefore, laboratories and materials are fully crossed factors. Operators and testing instruments are usually confined to their laboratories; therefore, operators and instruments are nested factors within laboratories. The design should provide for the same number of operators, number of instruments, and number of specimens from each material within each laboratory.

A1.3 Select laboratories and materials in accordance with Section 7 or 8, as is applicable.

A1.4 Summarize the results in a separate table for each laboratory showing averages obtained by each operator on each piece of equipment from each material. Provide averages for each operator and each piece of equipment for each material, and an overall average. The recommended summary format is shown in Table A1.1 for a laboratory with two operators and two testing machines.

A1.5 Summarize all the results in accordance with Table A1.2.

TABLE A1.1 Recommended Format for Summarizing Results from Each Laboratory

Interlaboratory Test of XXX Test Procedure—Averages for Operators and Machines Within Laboratory XX Tests Conducted on MM/DD/YY						
Operator	Machine	Material				Average
		A	B	...	Z	
a	1	x	x	...	x	x
	2	x	x	...	x	x
b	1	x	x	...	x	x
	2	x	x	...	x	x
a	...	x	x	...	x	x
b	...	x	x	...	x	x
...	1	x	x	...	x	x
...	2	x	x	...	x	x
Average		x	x	...	x	x

TABLE A1.2 Recommended Format for Summarizing Results from Pilot-Scale and Full-Scale Interlaboratory Tests

Interlaboratory Test of XXX Test Procedure—Averages for Materials by Laboratories Tests Conducted on MM/DD/YY					
Laboratory	Material				Average
	A	B	...	Z	
I	<sup>x</sup> A <sub>I</sub>	<sup>x</sup> B <sub>I</sub>	...	<sup>x</sup> Z <sub>I</sub>	<sup>x</sup> $\bar{x}$ <sub>I</sub>
II	<sup>x</sup> A <sub>II</sub>	<sup>x</sup> B <sub>II</sub>	...	<sup>x</sup> Z <sub>II</sub>	<sup>x</sup> $\bar{x}$ <sub>II</sub>
lig	<sup>lig</sup> A	<sup>lig</sup> B	...	<sup>lig</sup> Z	<sup>lig</sup> $\bar{x}$
N	<sup>x</sup> A <sub>N</sub>	<sup>x</sup> B <sub>N</sub>	...	<sup>x</sup> Z <sub>N</sub>	<sup>x</sup> $\bar{x}$ <sub>N</sub>
Average	<sup>x</sup> A	<sup>x</sup> B	...	<sup>x</sup> Z	<sup>x</sup> $\bar{x}$

**TABLE A1.3 Critical Values of the Calculated Friedman’s S Statistic at the 95 % Probability Level<sup>4</sup>**

n	k		
	3	4	5
2	...	6.0	...
3	6.0	7.4	8.5
4	6.5	7.8	8.8
5	6.4	7.8	8.9
6	7.0	7.6	...
7	7.1	7.8	...
8	6.2	7.6	...
9	6.2	...	...
10	6.2	...	...
11	6.5	...	...
12	6.5	...	...
13	6.6	...	...

**TABLE A1.4 Arrangement of Data for Testing Laboratory-by-Material Interaction**

Interlaboratory Test of Pilling Resistance Ratings (ASTM D 3512 – 82)—Average

Laboratory	Pilling Resistance Rating								Average
	Material								
	A		B		C		D		
	Sample		Sample		Sample		Sample		
	1	2	1	2	1	2	1	2	
I	2.75	3.50	2.00	2.00	4.50	4.50	4.75	5.00	3.63
II	2.50	3.00	2.50	2.00	5.00	4.50	5.00	4.00	3.56
III	4.25	4.75	4.00	5.00	5.00	5.00	5.00	5.00	4.75
IV	4.00	4.00	2.50	3.50	5.00	5.00	5.00	5.00	4.25
V	2.50	3.50	2.50	2.50	5.00	5.00	5.00	4.75	3.84
Average	3.20	3.75	2.70	3.00	4.90	4.80	4.95	4.75	4.01

**TABLE A1.5 Recommended Format for Arranging Data to Test for Interactions**

Interlaboratory Test of XXX Test Procedure  
Tests for Interactions from Laboratory Number XX

Material	Operator	Machine 1	Machine 2	Average
A	a	X	X	X
	b	X	X	X
B	a	X	X	X
	b	X	X	X
ijg	ijg	ijg	ijg	ijg
Z	a	X	X	X
	b	X	X	X
Average		$\bar{X}$	$\bar{X}$	$\bar{X}$

A1.6 Analyze the data using the Friedman Rank Sum Test.<sup>6</sup> This method is used to determine significance of: differences between operators within each laboratory, differences between machines within each laboratory, differences between laboratories, differences between materials, and any interactions.

A1.7 To test significance of differences between laboratories and between materials, arrange the data in a two-way layout in accordance with Table A1.2. If the difference between laboratories is being tested for significance, rank the results within each column, and then sum the ranks for each row. If the difference between materials is the one being tested, rank the results within each row, and then sum the ranks for each column.

<sup>6</sup> Hollander, Myles, and Wolf, Douglas, *Nonparametric Statistical Methods*, John Wiley & Sons, 1973, pp. 138–140, 366–371.

A1.8 Use Eq A1.1 to calculate the statistic, *S*.

$$S = \frac{12}{nk(k+1)} \sum_{i=1}^k R_i^2 - 3n(k+1) \tag{A1.1}$$

where:

*S* = Friedman Rank-Sum statistic for comparing laboratories (materials),

when comparing:

Symbol	Laboratories	Materials
<i>n</i>	number of materials	number of laboratories
<i>k</i>	number of laboratories	number of materials
<i>R</i>	sum of ranks for each of the laboratories	sum of ranks for each of the materials

A1.9 To determine if the difference between laboratories or materials is significant, compare the calculated *S* statistic with the values in a table of probabilities of Friedman’s *S* statistic,<sup>6</sup> or use Table A1.3.

A1.10 As the number *n* increases, the statistic *S* approaches  $\chi^2$  based on *k* – 1 degrees of freedom. Therefore, if the number of materials or laboratories exceeds the number shown in the table of probabilities of Friedman’s *S* statistic, then compare the calculated *S* statistic with the value shown in a  $\chi^2$  table for *k* – 1 degrees of freedom. The difference between laboratories or materials is significant if  $S \geq \chi^2$  at some preselected probability level.

A1.11 Apply this method to differences between operators within laboratories to differences between machines within laboratories. Calculate an *S* for each laboratory and sum them for all laboratories. If the resultant *S* is compared with values in a  $\chi^2$  table, the appropriate number of degrees of freedom is the sum of the degrees of freedom for each *S*. See Table A1.4.

A1.12 To determine the significance of two-way interactions,<sup>7</sup> arrange the data as shown in Table A1.5. The within-laboratory interactions to test include: operator-by-material, operator-by-machine, and machine-by-material. The only between-laboratory interaction to test is laboratory-by-material. The headings shown in Table A1.5 are an example of the headings to be used to test an operator-by-machine interaction. For further details on this type of analysis, see the indicated reference.<sup>7</sup>

A1.13 Tabulate the difference between corresponding values of the factor at each level and arrange them in a table as shown in Table A1.6. Table A1.6 has operator-by-material headings as an example. In the case of nested factors, arrange such a table for each laboratory.

A1.14 Assign ranks across each row and sum the ranks for each column.

A1.15 Calculate the *S* statistic using Eq A1.1. When testing for interactions of nested factors, calculate an *S* for each laboratory as directed in A1.11.

<sup>7</sup> Wilcoxon, Frank, “Some Rapid Approximate Statistical Procedures,” American Cyanamid Co., Stamford, CT, 1949, pp. 8–9.

**TABLE A1.6 Recommended Format for Arranging Differences Between Levels of Factors to Test for Interactions**

Interlaboratory Test of XXX Test Procedure Tests for Interactions from Laboratory Number XX			
Differences Between Results from Operators a and b			
Material	Operator a—Operator b		Average
	Machine 1	Machine 2	
A	X	X	X
B	X	X	X
lig	lig	lig	lig
Z	X	X	X
Average	$\bar{X}$	$\bar{X}$	$\bar{X}$

A1.16 If the interaction involves three levels of a factor, here called *A*, *B*, and *C*, it is possible to calculate *S* as the sum of two *S* values. One *S* is obtained by tabulating *A – B* for the different blocks. The other *S* is obtained by tabulating *A + B – 2C*. With four levels, the third *S* is obtained by tabulating *A + B + C – 3D*. These *S*'s are added to give the total *S*, and the degrees of freedom are added also. If interactions of multi-level, nested factors are being tested, after summing the *S*'s and degrees of freedom for each laboratory, sum these *S*'s and degrees of freedom and compare this result with the table value to determine significance.

A1.17 Examples are given to illustrate these procedures. Five laboratories participated in an interlaboratory study of pilling resistance using Practice D 3512–82. Four different materials were included in this study. Each laboratory had two operators each of whom tested three specimens from two samples of each of the materials. The raw data are not shown. The data listed in Table A1.7 are the averages of the three specimens per sample. The results averaged over operators, samples, and specimens are shown in Table A1.8.

A1.18 To test for significance of the difference between laboratories, rank the results within each column and sum the ranks for each laboratory. These rankings and their sums are shown in Table A1.9.

A1.19 Calculate *S* for the difference between laboratories using *k* = 5 laboratories, *n* = 4 materials, and Eq A1.1.

$$S = \frac{12}{(4) \cdot (5)(5 + 1)} \cdot (8^2 + 6.5^2 + 18.5^2 + 16.5^2 + 10.5^2) - (3) \cdot (4)(5 + 1) + 11.1$$

A1.20 *Solution:*

A1.20.1 *Solution Using Table A1.3*—Table A1.3 shows that for *k* = 5 and *n* = 4, an *S* equal to or greater than 8.8 is significant at the 95 % confidence level. That is, if there is no difference between the laboratory means, 95 % of the time *S* will be less than 8.8. Since *S* = 11.1, the calculated *S* is significant, so the difference between laboratories is significant.

A1.20.2 *Solution Using Reference Text*—The calculated *S* statistic can be compared with the values in a table of probabilities of Friedman's *S* statistic.<sup>4</sup> For *k* = 5 and *n* = 4, the probability of an *S* being as high as 11.1 is only 0.9 %; therefore, it is concluded that the difference between laboratories is significant.

A1.21 The only between-laboratory interaction to test is laboratory-by-material. Arrange the data as shown in Table

**TABLE A1.7 Arrangement of Data for Testing Operator-by-Material Interaction**

Interlaboratory Test of Pilling Resistance Ratings (ASTM D 3512 – 82)  
Average Pilling Resistance Rating

Laboratory I						
Operator	Sam- ple	Material				Aver- age
		A	B	C	D	
a	1	2.5	2.5	4.0	4.5	3.38
	2	3.5	3.0	4.0	5.0	3.88
b	1	3.0	1.5	5.0	5.0	3.62
	2	3.5	1.0	5.0	5.0	3.62
Average		3.12	2.00	4.50	4.88	3.62
Laboratory II						
Operator	Sam- ple	Material				Aver- age
		A	B	C	D	
a	1	2.0	2.5	5.0	5.0	3.62
	2	3.0	2.5	4.0	4.0	3.38
b	1	3.0	2.5	5.0	5.0	3.88
	2	3.0	1.5	5.0	4.0	3.38
Average		2.75	2.25	4.75	4.50	3.56
Laboratory III						
Operator	Sam- ple	Material				Aver- age
		A	B	C	D	
a	1	4.0	4.0	5.0	5.0	4.50
	2	4.5	5.0	5.0	5.0	4.88
b	1	4.5	4.0	5.0	5.0	4.62
	2	5.0	5.0	5.0	5.0	5.00
Average		4.50	4.50	5.00	5.00	4.75
Laboratory IV						
Operator	Sam- ple	Material				Aver- age
		A	B	C	D	
a	1	3.5	2.5	5.0	5.0	4.00
	2	3.5	4.0	5.0	5.0	4.38
b	1	4.5	2.5	5.0	5.0	4.25
	2	4.5	3.0	5.0	5.0	4.38
Average		4.00	3.00	5.00	5.00	4.25
Laboratory V						
Operator	Sam- ple	Material				Aver- age
		A	B	C	D	
a	1	3.0	2.0	5.0	5.0	3.75
	2	3.5	2.0	5.0	5.0	3.88
b	1	2.0	3.0	5.0	5.0	3.75
	2	3.5	3.0	5.0	4.5	4.00
Average		3.00	2.50	5.00	4.88	3.84

A1.4. Note that when analyzing a two-way interaction, a third factor is needed for blocking; in this case, replicate is used for this purpose.

A1.22 Since this analysis includes multi-level factors, calculate the different *S*'s as directed in A1.16 and sum them.

A1.23 Table A1.10 shows the tabulations *A – B*, *A + B – 2C*, and *A + B + C – 3D* and the rankings needed to calculate each of the three *S*'s. Using Eq A1.1, Calculate each *S*.

$$S_{A-B} = \frac{12}{(2) \cdot (5)(5 + 1)} (9^2 + 5^2 + 4^2 + 7^2 + 5^2) - (3) \cdot (2)(5 + 1) = 3.20$$

$$S_{A+B-2C} = \frac{12}{(2) \cdot (5)(5 + 1)} (6^2 + 3^2 + 10^2 + 8^2 + 3^2)$$

**TABLE A1.8 Average Pilling Resistance Ratings<sup>A</sup>**

Interlaboratory Test of Pilling Resistance Ratings (ASTM D 3512 – 82)— Averages by Materials and Laboratories					
Laboratory	Material				Average
	A	B	C	D	
I	3.12	2.00	4.50	4.88	3.62
II	2.75	2.25	4.75	4.50	3.56
III	4.50	4.50	5.00	5.00	4.75
IV	4.00	3.00	5.00	5.00	4.25
V	3.00	2.50	5.00	4.88	3.84
Average	3.47	2.85	4.85	4.85	4.01

<sup>A</sup> ASTM Rating System 5—no pilling  
4—slight pilling  
3—moderate pilling  
2—severe pilling  
1—very severe pilling

**TABLE A1.9 Interlaboratory Test of Pilling Resistance**

Ratings (ASTM D 3512 – 82) Rankings of Pilling Resistance Ratings <sup>A</sup> —Averages by Materials and Laboratories					
Laboratory	Material				Rankings Sum
	A	B	C	D	
I	3	1	1	3	8
II	1	2	2	1.5	6.5
III	5	5	4	4.5	18.5
IV	4	4	4	4.5	16.5
V	2	3	4	1.5	10.5

<sup>A</sup> The lowest pilling resistance rating is given a ranking of one in this example. This is common practice, but is not mathematically significant as long as ascendancy or decendency of rankings is consistently applied.

$$- (3) \cdot (2)(5 + 1) = 7.60$$

$$S_{A+B+C-3D} = \frac{12}{(2) \cdot (5)(5 + 1)} (3^2 + 5.5^2 + 10^2 + 7.5^2 + 4^2) - (3) \cdot (2)(5 + 1) = 6.30$$

A1.24 The three  $S$ 's are summed to determine the  $S$  for laboratory-by-material interaction as follows:

$$S_{\text{lab.} \times \text{mat.}} = 3.20 + 7.60 + 6.30 = 17.10$$

A1.25 Calculate the degrees of freedom for each  $S$  using Eq A1.2 as follows:

$$df_s = (n - 1)(k - 1) \quad (\text{A1.2})$$

where:

$df_s$  = degrees of freedom for the Friedman Rank-Sum statistic being calculated,

$n$  = number of blocks (replications in this example), and

$k$  = number of columns (laboratories in this example).

A1.26 In this example, each of the three  $S$ 's has the same number of degrees of freedoms, four. The total degrees of freedom for  $S_{\text{lab.} \times \text{mat.}}$  is twelve.

A1.27 To determine if the calculated  $S_{\text{lab.} \times \text{mat.}}$  is significant, compare it with the value in a table of  $\chi^2$  at the 95 % probability level for twelve degrees of freedom.

A1.28 The calculated  $S_{\text{lab.} \times \text{mat.}}$ , 17.10 is less than the table value 21.026. Therefore, there is not a significant laboratory-by-material interaction in this interlaboratory test.

A1.29 An example of testing for the significance of interactions of factors nested within laboratories is given by calculating the Friedman Rank-Sum statistic for operator-by-material in the interlaboratory test of crease appearance.

A1.30 Arrange the data as shown in Table A1.7, displaying results from each of the participating laboratories.

A1.31 Tabulate the differences in results obtained by operators from each of the samples. Table A1.11 shows these differences and the ranks needed to calculate the Friedman Rank-Sum statistic for this example.

A1.32 Use Eq A1.1 to calculate an  $S$  for each laboratory.

$$S_I = \frac{12}{(2) \cdot (4)(4 + 1)} (5^2 + 8^2 + 2^2 + 5^2) - (3) \cdot (2)(4 + 1) = 4.8$$

$$S_{II} = \frac{12}{(2) \cdot (4)(4 + 1)} (3.5^2 + 7^2 + 4^2 + 5.5^2) - (3) \cdot (2)(4 + 1) = 2.25$$

$$S_{III} = \frac{12}{(2) \cdot (4)(4 + 1)} (2^2 + 6^2 + 6^2 + 6^2) - (3) \cdot (2)(4 + 1) = 3.60$$

$$S_{IV} = \frac{12}{(2) \cdot (4)(4 + 1)} (2^2 + 7^2 + 5.5^2 + 5.5^2) - (3) \cdot (2)(4 + 1) = 4.05$$

$$S_V = \frac{12}{(2) \cdot (4)(4 + 1)} (6.5^2 + 2^2 + 5^2 + 6.5^2) - (3) \cdot (2)(4 + 1) = 4.05$$

A1.33 The  $S$ 's are summed to determine the overall  $S$  for operator-by-material interaction.

$$S_{\text{op.} \times \text{mat.}} = 4.8 + 2.25 + 3.60 + 4.05 + 4.05 = 18.75$$

A1.34 Using Eq A1.2, the degrees of freedom for each of the five laboratories are three. The total degrees of freedom for  $S_{\text{op.} \times \text{mat.}}$  is fifteen.

A1.35 Comparing the calculated  $S_{\text{op.} \times \text{mat.}}$  with the value in a  $\chi^2$  table at the 95 % confidence level for fifteen degrees of freedom results in a conclusion that there is no significant interaction of operators and materials. The  $S_{\text{op.} \times \text{mat.}}$ , 18.75, is less than the table value of 24.996.

**TABLE A1.10 Tabulation of Differences to Calculate the Friedman Rank-Sum Statistic to Determine the Significance of the Laboratory-by-Material Interaction**

Material A—Material B Pilling Resistance										
Sample Number	Laboratory									
	I		II		III		IV		V	
	A – B	Rank	A – B	Rank	A – B	Rank	A – B	Rank	A – B	Rank
1	0.75	4	0.00	1.5	0.25	3	1.5	5	0	1.5
2	1.50	5	1.00	3.5	-0.25	1	0.5	2	1.0	3.5
Sum of Ranks	...	9	...	5.0	...	4	...	7	...	5.0

Material A + Material B – 2 Material C Pilling Resistance										
Sample Number	Laboratory									
	I		II		III		IV		V	
	A + B – 2C	Rank	A + B – 2C	Rank	A + B – 2C	Rank	A + B – 2C	Rank	A + B – 2C	Rank
1	-4.25	3	-5.00	1.5	-1.75	5	-3.50	4	-5.00	1.5
2	-3.50	3	-4.00	1.5	-0.25	5	-2.50	4	-4.00	1.5
Sum of Ranks	...	6	...	3	...	10	...	8	...	3

Material A + Material B + Material C – 3 Material D Pilling Resistance										
Sample Number	Laboratory									
	I		II		III		IV		V	
	A + B + C – 3D	Rank	A + B + C – 3D	Rank	A + B + C – 3D	Rank	A + B + C – 3D	Rank	A + B + C – 3D	Rank
1	-5.00	2	-5.00	2	-1.75	5	-3.50	4	-5.00	2
2	-5.00	1	-2.50	3.5	-0.25	5	-2.50	3.5	-3.25	2
Sum of Ranks	...	3	...	5.5	...	10	...	7.5	...	4

**TABLE A1.11 Tabulation of Differences Needed to Calculate the Friedman Rank-Sum Statistic to Determine the Significance of Operator-by-Material Interaction**

Differences in Pilling Resistance Ratings Operator a Minus Operator b										
Sample	Laboratory I									
	Material									
	A		B		C		D			
	a – b	rank	a – b	rank	a – b	rank	a – b	rank	a – b	rank
1	-0.5	2.5	1.0	4	-1.0	1	-0.5	2.5		
2	0.0	2.5	2.0	4	-1.0	1	0.0	2.5		
Sum of Ranks	...	5	...	8	...	2	...	5		

Sample	Laboratory II									
	Material									
	A		B		C		D			
	a – b	rank	a – b	rank	a – b	rank	a – b	rank	a – b	rank
1	-1.0	1	0.0	3	0.0	3	0.0	3		
2	0.0	2.5	1.0	4	-1.0	1	0.0	2.5		
Sum of Ranks	...	3.5	...	7	...	4	...	5.5		

Sample	Laboratory III									
	Material									
	A		B		C		D			
	a – b	rank	a – b	rank	a – b	rank	a – b	rank	a – b	rank
1	-0.5	1	0.0	3	0.0	3	0.0	3		
2	-0.5	1	0.0	3	0.0	3	0.0	3		
Sum of Ranks	...	2	...	6	...	6	...	6		

Sample	Laboratory IV									
	Material									
	A		B		C		D			
	a – b	rank	a – b	rank	a – b	rank	a – b	rank	a – b	rank
1	-1.0	1	0.0	3	0.0	3	0.0	3		
2	0.0	1	1.0	4	0.0	2.5	0.0	2.5		
Sum of Ranks	...	2	...	7	...	5.5	...	5.5		

Sample	Laboratory V									
	Material									
	A		B		C		D			
	a – b	rank	a – b	rank	a – b	rank	a – b	rank	a – b	rank
1	1.0	4	-1.0	1	0.0	2.5	0.0	2.5		
2	0.0	2.5	-1.0	1	0.0	2.5	0.5	4		
Sum of Ranks	...	6.5	...	2	...	5	...	6.5		

**A2. CALCULATION OF  $\chi^2$  TO DETERMINE SIGNIFICANCE OF DIFFERENCES**

A2.1  $\chi^2$  is used to determine significance of differences between distributions. It can be used whether the distribution is

expressed as relative frequency or frequency counts. It can be used to analyze results from distributions of any shape.  $\chi^2$  can be calculated for one set of results to be compared with a standard or a known distribution, for comparing two sets of results with each other, or for comparing several sets of results as from an interlaboratory study of a test method.

A2.2 For methods of calculation and tables of significant  $\chi^2$  values, refer to the appropriate textbooks on statistics.

A2.3 The following cautions should be observed when employing this method of analysis:

A2.3.1 Such analytical techniques require a greater number of observations than do parametric techniques to find signifi-

cant differences. If the data can be normalized by appropriate transformations, it is recommended that this be done in order to utilize the techniques of analysis such as those contained in Test Method D 4686.

A2.3.2 No class in the observed distribution should be such that the frequency of expected observations is zero. An actual frequency observed for a specific class or cell may be zero.

A2.3.3 If any data classes contain fewer than five expected observations, a Yates' correction for continuity<sup>8</sup> should be used.

---

<sup>8</sup> Fisher, R. A., *Statistical Methods for Researchers*, Biological Monographs and Manuals, Oliver and Boyd, London, 11 ed., 1950, pp. 94–97.

*The American Society for Testing and Materials takes no position respecting the validity of any patent rights asserted in connection with any item mentioned in this standard. Users of this standard are expressly advised that determination of the validity of any such patent rights, and the risk of infringement of such rights, are entirely their own responsibility.*

*This standard is subject to revision at any time by the responsible technical committee and must be reviewed every five years and if not revised, either reapproved or withdrawn. Your comments are invited either for revision of this standard or for additional standards and should be addressed to ASTM Headquarters. Your comments will receive careful consideration at a meeting of the responsible technical committee, which you may attend. If you feel that your comments have not received a fair hearing you should make your views known to the ASTM Committee on Standards, at the address shown below.*

*This standard is copyrighted by ASTM, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959, United States. Individual reprints (single or multiple copies) of this standard may be obtained by contacting ASTM at the above address or at 610-832-9585 (phone), 610-832-9555 (fax), or service@astm.org (e-mail); or through the ASTM website (www.astm.org).*