Designation: D 4853 – 97 (Reapproved 2002)

Standard Guide for Reducing Test Variability[1]

This standard is issued under the fixed designation D 4853; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This guide serves as an aid to subcommittees writing and maintaining test methods. It helps to (*1*) determine if it is possible to reduce test variability, and, if so, (*2*) provide a systematic approach to the reduction.

1.2 This guide includes the following topics:

Topic Title	Section Number
Scope	1
Referenced Documents	2
Definitions	3
Significance and Use	4
Measures of Test Variability	5
Unnecessary Test Variability	6
Identifying Probable Causes of Test Variability	7
Determining the Causes of Test Variability	8
Averaging	9
Calibration	10
Keywords	11

1.3 The annexes include:

Topic Title	Annex Number
Statistical Test Selection	Annex A1
Frequency Distribution Identification	Annex A2
Design of Ruggedness Tests	Annex A3
Ruggedness Test Analysis:	
Unknown or Undefined Distribution—Small Sample	Annex A4
Unknown or Undefined Distribution—Large Sample	Annex A5
Binomial Distribution	Annex A6
Poisson Distribution	Annex A7
Normal Distribution	Annex A8
Design of a Randomized Block Experiment	Annex A9
Randomized Block Experiment Analysis:	
Unknown or Undefined Distribution—Small Sample	Annex A10
Unknown or Undefined Distribution—Large Sample	Annex A11
Binomial Distribution	Annex A12
Poisson Distribution	Annex A13
Normal Distribution	Annex A14
Averaging:	
No Compositing	Annex A15
Compositing	Annex A16

2. Referenced Documents

2.1 *ASTM Standards:*

D 123 Terminology Relating to Textiles[2]

D 1907 Test Method for Yarn Number by the Skein Method[2]

D 2256 Test Method for Tensile Properties of Yarns by the Single-Strand Method[2]

D 2904 Practice for Interlaboratory Testing of a Textile Test Method that Produces Normally Distributed Data[2]

D 2906 Practice for Statements on Precision and Bias for Textiles[2]

D 3512 Test Method for Pilling Resistance and Other Related Surface Changes of Textile Fabrics: Random Tumble Pilling Tester Method[3]

D 3659 Test Method for Flammability of Apparel Fabrics by Semi-Restraint Method[3]

D 4356 Practice for Establishing Consistent Test Method Tolerances[3,4]

D 4467 Practice for Interlaboratory Testing of a Test Method that Produces Non-Normally Distributed Data[3]

D 4686 Guide for Identification of Frequency Distributions[3]

D 4854 Guide for Estimating the Magnitude of Variability from Expected Sources in Sampling Plans[3]

E 456 Terminology Relating to Quality and Statistics[4]

E 1169 Guide for Conducting Ruggedness Tests[4]

2.2 *ASTM Adjuncts:*

TEX-PAC[5]

NOTE 1—Tex-Pac is a group of PC programs on floppy disks, available through ASTM Headquarters, 100 Barr Harbor Drive, West Conshohocken, PA 19428, USA. The analysis described in Annex A4 can be conducted using one of these programs.

3. Terminology

3.1 *Definitions:*

3.1.1 *average*, *n—for a series of observations*, the total divided by the number of observations. (Syn. *arithmetic average*, *arithmetic mean*, *mean*)

3.1.2 *block*, *n—in experimenting*, a group of units that is relatively homogeneous within itself, but may differ from other similar groups.

3.1.3 *degrees of freedom*, *n—for a set*, the number of values that can be assigned arbitrarily and still get the same value for each of one or more statistics calculated from the set of data.

---

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

Current edition approved Sept. 10, 1997. Published August 1998. Originally published as D4853 – 88. Last previous edition D4853 – 91.

[2] *Annual Book of ASTM Standards*, Vol 07.01.

[3] *Annual Book of ASTM Standards,* Vol 07.02.

[4] *Annual Book of ASTM Standards*, Vol 14.02.

[5] PC programs on floppy disks are available through ASTM. For a 3½ inch disk request PCN:12-429040-18, for a 5½ inch disk request PCN:12-429041-18.

3.1.4 *duplicate*, *n—in experimenting or testing*, one of two or more runs with the same specified experimental or test conditions but with each experimental or test condition not being established independently of all previous runs. (Compare *replicate*)

3.1.5 *duplicate*, *vt—in experimenting or testing*, to repeat a run so as to produce a duplicate. (Compare *replicate*)

3.1.6 *error of the first kind*, α, *n—in a statistical test*, the rejection of a statistical hypothesis when it is true. (Syn. *Type I error*)

3.1.7 *error of the second kind*, β, *n—in a statistical test*, the acceptance of a statistical hypothesis when it is false. (Syn. *Type II error*)

3.1.8 *experimental error*, *n*—variability attributable only to a test method itself.

3.1.9 *factor*, *n—in experimenting*, a condition or circumstance that is being investigated to determine if it has an effect upon the result of testing the property of interest.

3.1.10 *interaction*, *n*—the condition that exists among factors when a test result obtained at one level of a factor is dependent on the level of one or more additional factors.

3.1.11 *mean*—See *average*.

3.1.12 *median*, *n—for a series of observations*, after arranging them in order of magnitude, the value that falls in the middle when the number of observations is odd or the arithmetic mean of the two middle observations when the number of observations is even.

3.1.13 *mode*, *n*—the value of the variate for which the relative frequency in a series of observations reaches a local maximum.

3.1.14 *randomized block experiment*, *n*—a kind of experiment which compares the average of *k* different treatments that appear in random order in each of *b* blocks.

3.1.15 *replicate*, *n—in experimenting or testing*, one of two or more runs with the same specified experimental or test conditions and with each experimental or test condition being established independently of all previous runs. (Compare *duplicate*)

3.1.16 *replicate*, *vt—in experimenting or testing*, to repeat a run so as to produce a replicate. (Compare *duplicate*)

3.1.17 *ruggedness test*, *n*—an experiment in which environmental or test conditions are deliberately varied to evaluate the effect of such variations.

3.1.18 *run*, *n—in experimenting or testing*, a single performance or determination using one of a combination of experimental or test conditions.

3.1.19 *standard deviation, s, n—of a sample*, a measure of the dispersion of variates observed in a sample expressed as the positive square root of the sample variance.

3.1.20 *treatment combination*, *n—in experimenting*, one set of experimental conditions.

3.1.21 *Type I error*—See *error of the first kind*.

3.1.22 *Type II error*—See *error of the second kind*.

3.1.23 *variance*, s^2, *n*—of a sample, a measure of the dispersion of variates observed in a sample expressed as a function of the squared deviations from the sample average.

3.1.24 For definitions of textile terms, refer to Terminology D 123. For definitions of other statistical terms, refer to Terminology E 456.

4. Significance and Use

4.1 This guide can be used at any point in the development or improvement of a test method, if it is desired to pursue reduction of its variability.

4.2 There are three circumstances in which a subcommittee responsible for a test method would want to reduce test variability:

4.2.1 During the development of a new test method, ruggedness testing might reveal factors which produce an unacceptable level of variability, but which can be satisfactorily controlled once the factors are identified.

4.2.2 Another is when analysis of data from an interlaboratory test of a test method shows significant differences between levels of factors or significant interactions which were not desired or expected. Such an occurrence is an indicator of lack of control which means that the precision of the test method is not predictable.

4.2.3 The third situation is when the method is in statistical control, but it is desired to improve its precision, perhaps because the precision is not good enough to detect practical differences with a reasonable number of specimens.

4.3 The techniques in this guide help to detect a statistical difference between test results. They do not directly answer questions about practical differences. A statistical difference is one which is not due to experimental error, that is, chance variation. Each statistical difference found by the use of this guide must be compared to a practical difference, the size of which is a matter of engineering judgment. For example, a change of one degree in temperature of water may show a statistically significant difference of 0.05 % in dimensional change, but 0.05 % may be of no importance in the use to which the test may be put.

5. Measures of Test Variability

5.1 There are a number of measures of test variability, but this guide concerns itself with only two: one is the probability, *p*, that a test result will fall within a particular interval; the other is the positive square root of the variance which is called the standard deviation, *s*. The standard deviation is sometimes expressed as a percent of the average which is called the coefficient of variation, *CV* %. Test variability due to lack of statistical control is unpredictable and therefore cannot be measured.

6. Unnecessary Test Variability

6.1 The following are some frequent causes of unnecessary test variability:

6.1.1 Inadequate instructions.

6.1.2 Miscalibration of instruments or standards.

6.1.3 Defective instruments.

6.1.4 Instrument differences.

6.1.5 Operator training.

6.1.6 Inattentive operator.

6.1.7 Reporting error.

6.1.8 False reporting.

6.1.9 Choice of measurement scale.

6.1.10 Measurement tolerances either unspecified or incorrect.

6.1.11 Inadequate specification of, or inadequate adherence to, tolerances of test method conditions. (For establishing consistent tolerances, see Practice D 4356.)

6.1.12 Incorrect identification of materials submitted for testing.

6.1.13 Damaged materials.

7. Identifying Probable Causes of Test Variability

7.1 Sometimes the causes of test variability will appear to be obvious. These should be investigated as probable causes, but the temptation should be avoided to ignore other possible causes.

7.2 The list contained in Section 6 should be reviewed to see if any of these items could be producing the observed test variability.

7.3 To aid in selecting the items to investigate in depth, plot frequency distributions and statistical quality control charts **(1)**.[6] Make these plots for all the data and then for each level of the factors which may be causes of, or associated with, test variability.

7.4 In examining the patterns of the plots, there may be some hints about which factors are not consistent among their levels in their effect on test variability. These are the factors to pursue.

8. Determining the Causes of Test Variability

8.1 *Use of Statistical Tests*:

8.1.1 This section includes two statistical techniques to use to investigate the significance of the factors identified as directed in Section 7: ruggedness tests and randomized block experiments analyses. In using these techniques, it is advantageous to choose a model to describe the distribution from which the data come. Methods for identifying the distributions are contained in Annex A2 and Guide D 4686. For additional information about distribution identification, see Shapiro **(2)**.

8.1.2 In order to assure being able to draw conclusions from ruggedness testing and components of variance analysis, it is essential to have sufficient available data. Not infrequently, the quantity of data is so small as to preclude significant differences being found if they exist.

8.2 *Ruggedness Tests*:

8.2.1 Use ruggedness testing to determine the method's sensitivity to variables which may need to be controlled to obtain an acceptable precision. Ruggedness tests are designed using only two levels of each of one or more factors being examined. For additional information see Guide E 1169.

8.2.2 Prepare a definitive statement of the type of information the task group expects to obtain from the ruggedness test. Include an example of the statistical analysis to be used, using hypothetical data.

8.2.3 Design, run, and analyze the ruggedness test as directed in Annex A3-Annex A8.

8.2.4 From a summary table obtained as directed in A3.6, the factors to which the test method is sensitive may become apparent. Some sensitivity is to be expected; it is usually desirable for a test method to detect differences between fabrics of different constructions, fiber contents, or finishes. Some sensitivities may be expected, but may be controllable; temperature is frequently such a factor.

8.2.5 If analysis shows that any test conditions have a significant effect, modify the test procedure to require the degree of control required to eliminate any significant effects.

8.3 *Randomized Block Experiment*:

8.3.1 When it is desired to investigate a test method's sensitivity to a factor at more than two levels, use a randomized block experiment. Such factors might be: specimen chambers within a machine, operators, shifts, or extractors. Analysis of the randomized block experiment will help to determine how much the factor levels contribute to the total variation of the test method results. Comparison of the factor level variation from factor to factor will identify the sources of large sums of squares in the total variation of the test results.

8.3.2 Prepare a definitive statement of the type of information the task group expects to obtain from the measurement of sums of squares. Include an example of the statistical analysis to be used, using hypothetical data.

8.3.3 Design, run, and analyze the randomized block experiment as directed in Annex A9-Annex A14.

8.3.4 From a summary table of results with blocks as rows and factor levels as columns, such as in A14.2.1, the levels of a factor to which the test method is sensitive may become apparent. Some sensitivity to level changes may be expected, but may be controllable; different operators are frequently such levels.

8.3.5 If the analysis shows any significant effects associated with changes in level of a factor, revise the test procedure to obviate the necessity for a level change. If this is not possible, give a warning, and explain how to minimize the effect of necessary level changes.

9. Averaging

9.1 *Variation*—Averages have less variation than individual measurements. The more measurements include in an average, the less its variation. Thus, the variation of test results can be reduced by averaging, but averaging will not improve the precision of a test method as measured by the variance of specimen selection and testing (Note 2).

NOTE 2—This section is applicable to all sampling plans producing variables data regardless of the kind of frequency distribution of these data because no estimations are made of any probabilities.

9.2 *Sampling Plans with No Composites*—Some test methods specify a sampling plan as part of the procedure. Selective increases or reductions in the number of lot and laboratory samples, and specimens specified can sometimes be made which will reduce test result variation and also reduce cost (Note 3). To investigate the possibility of making sampling plan revisions which will reduce variation proceed as directed in either Annex A15 or Annex A16.

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