T 1206 sp-91
OFFICIAL STANDARD – 1969
RECOMMENDED PRACTICE – 1986
REAFFIRMED – 1991

Precision statement for test methods

1. Scope

1.1 This recommended practice defines terms and describes how to estimate precision from available data.

1.2 A discussion of the application of the precision information is given in Appendix A.1, and some suggestions as to cause of poor precision and how to improve precision are given in Appendix A.2.

1.3 Two situations are considered: (a) repeatability or comparison of test results within a laboratory (same material, operator, apparatus and environmental conditions); and (b) reproducibility or comparison of test results among laboratories (same material, but different operator, apparatus and environmental conditions).

1.3.1 Other situations within a laboratory are not considered, e.g., data collected by all operators, or on all apparatus, or on different occasions when different environmental conditions might prevail. Because the agreement among the values obtained in these situations depends on the degree of supervision and control exercised in that laboratory, statements of precision about these situations are not appropriate in a test method for general use.

1.4 The precision information in a test method should not be used blindly. It is given as a guide and should be modified when required on the basis of information available about the particular circumstances of the application.

2. Definitions

2.1 *Precision*, the degree of agreement expected between two or more test results on the same property.

2.2 *Random error*, the chance variation encountered in all test work despite the closest control of variables.

2.3 *Variance*, a mathematical measure of the magnitude of random errors. It is equal to the sum of the squares of the random errors, divided by the number of degrees of freedom. (For the variance of values about their average, it is equal to the sum of the squares of the deviation of each value from the average, divided by the number of degrees of freedom, usually one less than the number of values).

2.4 *Degrees of freedom*, a number indicative of the effective number of values on which the evaluation of a variance is based.

2.5 *Standard deviation*, the square root of the variance.

2.6 *Test determination*, the process of carrying out the series of operations specified in a test method whereby the value for a single test specimen is obtained.

2.7 *Test unit*, a quantity of material sufficient to obtain a single adequate set of test results for all the properties to be measured.

2.8 *Test specimen*, a test unit, or a portion of a test unit, upon which a single or a multiple test determination is to be made.

2.9 *Test result*, the value obtained for one test unit of a sample of material by closely following all the instructions in the test method. This value may be, as specified in the instructions, that of a single test determination or it may be the average, median, standard deviation, or other combination of separate test determinations made on test

specimens taken from the test unit. Thus in a chemical method requiring duplicate test determinations, the test result will be the average of the duplicates.

2.10 *Probability interval*, upper and lower limits defining the range of values comprising a given fraction of a population of values. When each value is the difference between two test results whose expected difference is zero, only the positive limit (i.e., absolute value of the difference) need be stated. The following examples of such limits correspond to the two situations previously listed (1.3). They are especially appropriate for inclusion in the precision statement of a test method.

2.10.1 *Repeatability*, a limit within which agreement may be expected 95% of the time between two test results obtained under essentially the same conditions and from the same homogeneous sample of material. "Same conditions" means by the same operator on the same apparatus with environmental and other conditions held as constant as possible, such as for two test results obtained in immediate succession on the same day.

2.10.2 *Reproducibility*, a limit within which agreement may be expected 95% of the time between two test results obtained in different qualified laboratories for the same homogeneous sample of material. "Qualified laboratories" are those having the required apparatus and environmental conditions, and trained personnel who conscientiously follow the prescribed test method.

3. Estimating precision from available data

3.1 *Multiplier for TAPPI repeatability and reproducibility.* Repeatability and reproducibility have sometimes been equated to standard deviations. The 95% probability limit on the difference, to be used in TAPPI test methods, is 2.77 times the corresponding standard deviation. The factor 2.77 is 1.960 $(2)^{1/2}$. The 1.960 comes from the choice of the 95% probability limit and the usual assumption of approximately normal distribution of errors. The $(2)^{1/2}$ follows from the definition being in terms of the difference between two test results.

3.2 *Repeatability from replicate measurements within a single laboratory.*

3.2.1 Normally *some* estimate of repeatability should be available from the laboratory in which the method was developed. If not, rough estimates can be quickly obtained from individual laboratories, possibly from data already existing in the laboratories.

3.2.2 Suppose k groups of measurements, each with approximately n test determinations are available; e.g., a laboratory regularly tests ten specimens ($n = 10$) and has three ($k = 3$) such groups of ten available (Table 1). From each group, compute an estimate s_e of the standard deviation among the test determinations, after first setting aside extreme test determinations using TAPPI T 1205 "Dealing With Suspect (Outlying) Test Determinations" for outliers.

NOTE 1: While the test determinations within a group must have been obtained under essentially the same conditions (see 2.10.1), each determination must otherwise be independent of the others; e.g., the determinations should not have been made on the same prepared portion of the chemical or physical specimen when the test method calls for separate selection or preparation of each test specimen.

3.2.3 If the proposed test method specifies that the average of determinations on m test specimens is to be considered a test result, then compute the standard deviation of a test result from each s_e. This is $s_r = s_e/(m)^{1/2}$. If, in the case of Table 1, $m = 5$, then $s_r = s_e/(5)^{1/2}$. Finally, compute an estimate of the repeatability by multiplying by 2.77, as explained in Section 3.1.

3.2.4 If the data for estimating the repeatability is obtained from different laboratories, calculate the estimates in the same manner as if obtained in a single laboratory.

3.2.5 In order to determine the percent repeatability for each material, multiply the repeatability by 100 and divide by the test average for that material.

3.2.6 If either the repeatability or the percent repeatability is approximately the same for all materials, as is the percent repeatability in Table 1, average the values and report the average as the repeatability of the test method, in this case 1.6%. (For further discussion of the variation of repeatability and reproducibility with the average test value, see 3.4.1 and 3.4.2).

3.3 *Reproducibility from test results obtained in several laboratories.*

3.3.1 During preliminary work, e.g., in developing the proposed test method in a preliminary round robin, each of several laboratories may have obtained test results on the same material (Table 2). Compute an estimate s_n of the standard deviation among the laboratory averages, after first setting aside extreme test results by using the test for outliers. If the test result in each laboratory were each averages of approximately n determinations ($n = 20$ in Table 2) but the proposed test method calls for basing a test result on m specimens (say, $m = 10$), then compute a corrected estimate s_R of the between-laboratory standard deviation as follows:

$$(s_R)^2 = (s_n)^2 + (1/m - 1/n)\,(s_e)^2$$

where s_e is an estimate of the standard deviation of repeated test determinations within a typical laboratory (3.2).

3.3.2 Finally (Table 2), compute an estimate of the reproducibility of the method by multiplying by the factor 2.77 (Section 3.1).

3.4 *Repeatability and reproducibility from an interlaboratory study designed in accordance with T 1200.*

3.4.1 Table 3A of TAPPI T 1200 "Interlaboratory Evaluation of Test Methods to Determine TAPPI Repeatability and Reproducibility" shows the standard deviations and percent coefficients of variation of repeatability and reproducibility arranged in order of the grand averages for the materials. If one of these, *s* or %CV is approximately constant throughout the range (i.e., from low to high values of the grand averages), then the repeatability or reproducibility may be expressed as a single number. For example, $\%CV_R$ is approximately constant (ignoring M1). Averaging these values (omitting M1), we obtain 5.224%, from which it could be said that the reproducibility of the method is approximately $R = 2.77 \times 5.224 = 14.5\%$.

3.4.2 If neither *s* nor %CV is approximately constant, the repeatability or reproducibility may be expressed by a combination of constant and proportional values, by a simple formula, or by a single compromise value selected by judgement. In a case where the variation with test value is such that no simple expression can represent the 95% limit, consider the variation as a range in the probability, and select a compromise value or simple expression for the limit so as to let the variation in the probability range upward from 95% or possible very slightly lower. Examples:

Thickness: Reproducibility = 0.005 mm or 3%, whichever is larger.

Tensile energy absorption: Reproducibility = 22%.

Smoothness: Reproducibility = (18 - 4 log Bekk) %.

Tensile breaking strength: Reproducibility = 9.0% (actual range 8.3 - 9.1 %).

4. Recommended comprehensive statement of precision

4.1 *Essential parts of a precision statement.*

4.1.1 Number of test determinations in a test result.

4.1.2 Identification of the source of the precision information, including:

4.1.2.1 Specific reference to published data or to TAPPI archives where raw data and analysis may be filed.

4.1.2.2 If an interlaboratory study, the number of materials, the number of laboratories, and the number of test results or test determinations per material per laboratory.

4.1.2.3 If other than an interlaboratory study, similar information that will allow the reader to judge the reliability of the precision information.

4.1.3 The types of materials and ranges of test values to which the precision information applies.

4.1.4 A description of any deviation from complete replication of the test method for each test result.

4.1.5 The repeatability and reproducibility as estimated in Section 3.

4.1.6 Reference to T 1206 for definitions of repeatability and reproducibility.

4.2 Example:

> **11. Precision**
>
> 11.1 The values of repeatability and reproducibility provided below have been calculated for test results each of which is the average of 10 replicate test determinations. The values are based on data obtained (insert specific reference to literature or TAPPI archives) in which the range of test results was 32 to 275 Sheffield units. On the average, 32 laboratories participated for each of 18 materials, each laboratory obtaining two test results per material.
> 11.2 Repeatability (within a laboratory) = 8.2 Sheffield units. (The range of repeatability values for all of the materials in the study was 4.4 to 14.0 Sheffield units, with the central 90% of these values between 5.4 and 11.9).
> 11.3 Reproducibility (between laboratories) = 25 Sheffield units. (The range of reproducibility values for all of the materials in the study was 13 to 46 Sheffield units, with the central 90% of these values between 16 and 36).
> 11.4 Repeatability and reproducibility represent the agreement which is expected 95% of the time when two test results are compared. Refer to T 1206 for complete definitions of these terms.

5. Additional information

Effective date of issue: March 1, 1991

Appendixes

A.1 Application of the indices of precision

A.1.1 *Repeatability* refers to the precision associated with the most restricted system of causes of variability it is reasonable and useful to consider. It involves the same operator, same apparatus, usually the same day, and test specimens from the same sample of material. Repeatability may, therefore, be used as a value against which to judge (1) the ability of an operator to repeat his measurements, (2) the stability of an apparatus or environment for repeated measurements, and (3) the homogeneity of a lot of material. The repeatability of a test method places a lower practical value on the limit for ranges of the process control chart. If the effect of a small change in the manufacturing process or preparation of a material is to be evaluated, repeatability is the appropriate measure of the ability to detect this effect.

A.1.2 When operators, apparatuses, or environments are to be compared in the same laboratory to determine their degree of agreement, the goal is to approach the value obtained for repeatability, but a realistic figure will be somewhere *between repeatability and reproducibility*. Whether the degree of agreement in any case may be expected to be closer to repeatability or to reproducibility will depend to a marked extent on the degree of supervision and control maintained within the laboratory. Control chart limits for averages taken by several operators or over several days would similarly be governed by the degree of supervision and control. Presumably, the specifications for the test method reduce the effect of the causes of variability insofar as is practicable on an interlaboratory basis, but specific study and better control of them is recommended and feasible within a laboratory.

A.1.3 *Reproducibility* may be used as a value against which to judge (1) the ability of two laboratories to obtain the same test results on test specimens from the same homogeneous sample of material and (2) the closeness to which a laboratory may determine compliance with specifications. Confidence limits for a published test result, as well as the minimum tolerance permissible for a realistic specification should be derived from reproducibility (and not repeatability). The problem of determining compliance with a specification is essentially the same as that of obtaining agreement between laboratories, the buyer's and seller's. However, to assure compliance the seller must allow for the reproducibility limit.

A.2 Some causes and cures of poor precision

A.2.1 Generally, the value given for repeatability is an "average" for homogeneous material and is applicable to a reasonably-competent, appropriately-trained operator, conscientiously following the prescribed test method, and working with average but reasonably well-maintained apparatus and environmental control equipment. Experience has shown that the ability to repeat measurements does not vary considerably from laboratory to laboratory, except where substandard apparatus or work is involved, or where the environment is changing rapidly. Therefore when replicate measurements within a laboratory are in poorer agreement than would be expected from the stated repeatability of the method, one should look for careless work, non-standard apparatus, poor maintenance of apparatus or environment, or a highly variable sample. When the replicate measurements are in better agreement than could be expected, one should look for measurement scale that is too crude, data that are rounded-off excessively, or for subconscious or conscious manipulation of the data. On rare occasions, unusually good agreement among replicate measurements may be due to a very homogeneous sample or unusually great care in maintenance of apparatus and environment and in measurement.

A.2.2 When two laboratories fail to agree within the reproducibility value, the cause may be failure on the part of one or both to follow the prescribed procedure. Too often an operator, attempting to improve within-laboratory control or simply doing it in a way which he thinks is better, may modify a procedure to the point that he is no longer complying with the test method. Experience has shown that those laboratories that independently maintain rigid within-laboratory control over operator, apparatus and environment, do not necessarily show better agreement than those laboratories that conscientiously follow the prescribed test method.

A.2.3 If the stated repeatability of test results is not adequate for comparisons in which repeatability is the appropriate index of precision, it *may* be helpful to require that a test result be based on more test determinations. However, this additional within-laboratory replication may be of little value when reproducibility is the proper index.

A.2.4 If the stated reproducibility of the test method is inadequate, better agreement between two laboratories may sometimes be achieved through the use of a reference material. Improvement in reproducibility may be limited,

however, unless the reference material is a homogeneous sample of the same material as that to be evaluated, e.g., from previously accepted production. Further improvement in reproducibility may sometimes be achieved by bringing the two laboratories under the same supervisory control; e.g., by regular exchange visits between supervisors or operators to study each other's procedures in detail. Rarely, however, can the agreement between laboratories be made to equal the repeatability.

Table 1. Replicate measurement within a single laboratory.

		Tearing strength in grams		
	Replication	*Material A*	*Material B*	*Material C*
	1	32.0	45.5	92.0
	2	32.8	47.0	92.6
	3	31.8	45.8	100.7*
	4	32.2	46.5	92.8
	5	32.5	46.7	94.0
	6	31.5	46.0	93.4
	7	31.8	45.9	95.8
	8	32.3	47.4	93.0
	9	31.7	46.5	94.0
	10	32.2	46.2	95.0
Average, $\bar{x}$		32.08	46.35	93.62
Standard deviation, s_e		0.397	0.583	1.210
Std. dev. of test result, $s_r = s_e/(5)^{1/2}$		0.178	0.260	0.540
Repeatability, $r = 2.77\ s_r$		0.492	0.723	1.50
Repeatability in % = 100 $\bar{r}/\bar{x}$		1.53	1.56	1.60
Average repeatability in percent				1.6%

*Rejected at 1% level using test for outliers (T 1205).

Table 2. Test results from several laboratories.

Laboratory	*Average tearing strength, g*	*Standard deviation, s_e*
1	46.73	0.53
2	44.55	0.68
3	45.90	0.57
4	46.22	0.53
5	47.14	0.64
6	46.19	0.76
7	46.18	0.48
8	46.97	0.50
Average	46.23	0.582
Between-laboratory std. dev., s_n	0.807	
$(s_R)^2 = (s_n)^2 + (1/m - 1/n)\ (s_e)^2$		
$= (0.807)^2 + (1/10 - 1/20)\ (0.582)^2 = 0.668$		
Corrected between-lab. std. dev., s_R	0.817	
Reproducibility (between labs), $R = 2.77\ s_R$	2.26 g or 4.9%	

Your comments and suggestions on this procedure are earnestly requested and should be sent to the TAPPI Technical Divisions Administrator. ■