Designation: D6085 – 97 (Reapproved 2016)

Standard Practice for Sampling in Rubber Testing—Terminology and Basic Concepts[1]

This standard is issued under the fixed designation D6085; 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 a standardized terminology and some basic concepts for testing and sampling across the broad range of chemical and physical testing operations characteristic of the rubber and carbon black manufacturing industries.

1.2 In addition to the basic concepts and terminology, a model for the test measurement process is given in Annex A1. This serves as a mathematical foundation for the terms and other testing concepts. It may also find use for further development of this practice to address more complex sampling operations.

1.3 This general topic requires a comprehensive treatment with a sequential or hierarchical development of terms with substantial background discussion. A number of ancillary terms are also given that make for a more self-contained document. This cannot be accommodated in Terminology D1566.

1.4 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health practices and determine the applicability of regulatory limitations prior to use.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

D1566 Terminology Relating to Rubber

D4483 Practice for Evaluating Precision for Test Method Standards in the Rubber and Carbon Black Manufacturing Industries

D5406 Practice for Rubber—Calculation of Producer's Process Performance Indexes

3. Terminology

3.1 *Definitions:*

3.2 Despite the adoption of standard test methods, test result variation influences the data generated in all testing programs. As outlined in Annex A1, there are two main categories: (*1*) variation inherent in the production process for a material or class of objects, and (*2*) variation due to the measurement operation itself. Each of these two sources may be further divided into two types of variation: (*1*) systematic or bias variation, and (*2*) random error variation. Both types can exist simultaneously for either of the main categories.

3.3 Random variation can be reduced to a low level by appropriate replication and sampling procedures, but bias variation cannot be so reduced. However, bias variation can be reduced or eliminated by comprehensive programs to sort out the causes of such perturbations and eliminate these causes.

3.4 *Elementary Testing Terms:*

3.4.1 *lot, n*—a specified mass of material or number of objects, generated by an identifiable process, with a recognized composition or property range.

3.4.1.1 *Discussion*—A lot is frequently generated by a common production process in a restricted time period and usually consists of a finite size or number. A lot may be a fractional part of a population (Interpretation 2 of *population*).

3.4.2 *material, n*—a specific entity that exists in bulk form (solid, powder, liquid).

3.4.2.1 *Discussion*—A material may or may not be homogeneous. Typical materials are individual rubbers, compounds, accelerators, carbon blacks, etc.

3.4.3 *object, n*—a discrete item or piece with a specified shape and size.

3.4.3.1 *Discussion*—Usually an object is an entity that is ready for testing. A typical object is an o-ring, dumbbell, pellet, or hose assembly.

3.4.4 *object class (or class of objects), n*—a number of objects, with a recognized property range, generated by a common process, the objects are usually characterized by the value(s) of a unique property.

3.4.4.1 *Discussion*—In any testing program the phrase "a recognized property range" implies that the tester is aware of the approximate value of this range. At one extreme, this recognized range may denote "essentially identical property values;" at the opposite extreme this recognized range may

---

[1] This practice is under the jurisdiction of Committee D11 on Rubber and is the direct responsibility of Subcommittee D11.16 on Application of Statistical Methods.

Current edition approved June 1, 2016. Published July 2016. Originally approved in 1997. Last previous edition approved in 2011 as D6085 – 97 (2011). DOI: 10.1520/D6085-97R16.

[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.

denote "widely varying property values." The range applies to both materials and object classes, and testing programs will differ in regard to the extent of variation under consideration.

3.4.5 *population, n*—a generic term used in testing operations that may refer to any one of the following: (*1*) a single object or a very limited mass of material, (*2*) a finite (but large) number of objects or large mass of material, or (*3*) a hypothetical infinite number of objects or mass of material; all three interpretations imply that the objects or material are generated by some identifiable process and have recognized property range.

3.4.5.1 *Discussion*—Testing programs may vary from a very limited focus of attention (Interpretation 1) to an extreme or broad focus of attention (Interpretation 3). The focus of attention is determined by the testing and the sampling program.

3.4.6 *test (or testing), n*—a technical procedure performed on a material or object class using specified equipment that produces data unique to the material or object class; the data are used to evaluate or model selected properties or characteristics.

3.4.6.1 *Discussion*—Testing can be conducted on a narrow or broad basis, depending on the decisions to be made for any program. Sampling and replication plans (see 3.5) need to be specified for a complete program description. Testing may be divided into two major categories:

3.4.6.2 *global testing, n*—testing that is conducted at two or more locations or laboratories for the purpose of comparing materials or object classes at each location for selected characteristic properties.

3.4.6.3 *Discussion*—Typical global testing applications are producer-user testing and interlaboratory comparisons such as precision evaluation. These may be conducted on a worldwide basis.

3.4.6.4 *local testing, n*—testing that is conducted at one location or laboratory for the purpose of comparing a number of materials or object classes for some selected characteristic properties.

3.4.6.5 *Discussion*—Quality control and internal development programs are typical local testing.

3.5 *Sampling and Related Testing Terms:*

3.5.1 *random sample, n*—one of a sequence of samples (or sub-samples) taken on a random basis from a lot or population.

3.5.2 *replicate, n*—a generic term used in testing operations that denotes one of a selected number of fractional parts or objects taken from a sample; each fractional part or object is tested.

3.5.3 *replication, n*—the act of selecting replicates.

3.5.3.1 *Discussion*—The primary purpose of replication is the reduction of random test measurement variation. See Annex A1 for additional discussion on types of replication.

3.5.4 *sample, n*—a small fractional part of a material or a specified number of objects that are selected from a lot for testing, inspection, or specific observations of particular characteristics.

3.5.4.1 *Discussion*—A sample may be obtained from the combination of a series of incremental parts, each part obtained in one step of a particular sampling process (see *sub-sample*).

3.5.5 *sampling, n*—the act of selecting samples of any type.

3.5.5.1 *Discussion*—The goals of sampling are (*1*) an accurate or representative evaluation of the characteristic properties of the lot or population, and (*2*) an accurate estimate of the variation of properties within the lot or population.

3.5.6 *sub-sample, n*—one of a sequence of intermediate fractional parts or intermediate sets of objects taken from a lot or population that usually will be combined in a prescribed protocol to form a sample.

3.5.7 *systematic sample, n*—one of a sequence of samples (or sub-samples), each taken from a particular location or region of a lot or population according to a prescribed plan; the usual goal is to determine if the lot or population is homogeneous.

3.5.7.1 *Discussion*—Non-homogeneity may be of a random nature or it may be systematic (variation), which is frequently referred to as "stratification."

3.5.8 *test piece, n*—see *test specimen*, the preferred term.

3.5.9 *test portion, n*—see *test sample*, the preferred term.

3.5.10 *test sample, n*—that part of a sample of any type taken (usually with a prescribed blending or other protocol) for chemical or other analytical testing.

3.5.10.1 *Discussion*—A test sample may be a replicate or it may be subdivided into *n* replicates; these *n* replicates are frequently called "laboratory samples." Test samples are typically drawn to represent materials such as rubbers, pigments, chemicals, etc.

3.5.11 *test specimen, n*—an object (appropriately shaped and prepared) taken from a sample for physical or mechanical testing.

3.5.11.1 *Discussion*—A test specimen is usually a replicate. Test specimens are typically prepared by a rubber processing (and curing) operation to evaluate characteristic properties of development compounds and the compounds used in manufactured rubber products. They may also be manufactured rubber parts such as o-rings, motor mounts, etc.

3.6 *Testing Variation Terms*—Annex A1 defines a number of test data variation terms (see A1.4). Some additional terms, as given below, apply to typical development testing as well as quality control and quality assurance.

3.6.1 *common cause variation, n*—that residual variation inherent in any process that (*1*) is operating in a state of statistical control, and (*2*) is operating at some recognized or ascertained level of technological competence (Practice D5406).

3.6.2 *special cause variation, n*—that variation attributable to certain specific or assignable sources that have been (or may be) discovered through an investigation of the process (Practice D5406).

3.6.2.1 *Discussion*—Both common and special cause variation may contain bias as well as random variation components.

A synonym for common cause variation is *non-assignable variation*; a synonym for special cause variation is *assignable variation*.

3.6.3 *total process variation, n*—a range, along the measured property scale, defined as six times the standard deviation (determined under specified process conditions); the variation may contain either common or combined common and special cause sources (Practice D5406).

3.6.3.1 *Discussion*—Although this is defined in terms of quality control operations, it applies equally to general testing. For this general testing, the "process" and the "standard deviation" refer to the total (testing and sampling) operation.

4. Significance and Use

4.1 This practice provides a standardized terminology and a review of some basic concepts for elementary sampling and testing. These operations are important for: (*1*) local or intralaboratory testing that generally applies to internal quality control, process capability or performance evaluation, as well as development programs, and (*2*) global or interlaboratory testing that applies to producer-user and other testing programs that have industry-wide or worldwide scope. It is recognized that certain test methods may require more detailed and specialized sampling operations than the generic ones as described herein. These specialized sampling procedures should utilize, as far as possible, the basic concepts as given in this practice.

4.2 This practice provides a list of terms and concepts that may be used in numerous test methods used by Committees D11 and D24. It improves communication among those that conduct testing in the rubber and carbon black manufacturing industries and those that use such testing and test results for technical decisions.

5. Elementary Sampling Concepts

5.1 The purpose of testing is the production of test data to be used to make technical decisions. Test data are the end result of a three-step process: (*1*) test program organization, (*2*) sampling, and (*3*) testing. Test data quality depend on the organization and the sampling for any program.

5.2 *Types of Sampling Plans:*

5.2.1 *Intuitive Sampling*—This is a plan organized on the basis of the developed skill and judgment of the sampler. General information of the lot or population as well as past sampling and testing experience are used to make sampling decisions. The decisions made on the data generated by such a plan are based on a combination of the skill and experience of the tester buttressed by limited statistical conclusions. Strict probabilistic conclusions usually are not warranted.

5.2.2 *Statistical Sampling*—This is based on strict statistical sampling and such a plan provides the basis for authentic probabilistic conclusions. Hypothesis testing may be conducted, inferences drawn, and predictions made about future system or process behavior. Usually a large number of samples are needed if the significance of small differences is of importance. Conclusions from this type of sampling usually are not controversial. The statistical model chosen is important. When the number of samples required is large and this imposes a testing burden, hybrid plans using some simplifying intuitive assumptions are frequently employed.

5.2.3 *Protocol Sampling*—These are specified plans used for decision purposes in a given situation. Regulations (of the protocol) usually specify the type, size, frequency, and period of sampling in addition to the test methods to be used and other important sampling issues. The protocol may be based on a combination of intuitive and statistical considerations. Testing for conformance with producer-user specifications for commercial transactions is typical for this approach.

5.3 *Ensuring the Quality of Samples and Testing:*

5.3.1 Only the most elementary sampling issues are addressed in this section. For more detailed information, standard texts on sampling and sampling theory should be consulted. Well defined sampling operations must be conducted to ensure that high quality samples are used for any testing program. Quality is ensured when the samples are drawn in accordance with a prescribed procedure. This ensures that they accurately represent the lot or population. Such issues as sample homogeneity (or unintended stratification) and sample stability (conditioning or storage changes in the sample prior to testing) must be addressed. The sampling procedures, holding time, and other handling operations should be well documented. The test methods used for any program should be stable or in a state of statistical control and have demonstrated sensitivity as well as good precision in regard to the measured parameter.

5.3.2 Annex A1, paragraph A1.4, addresses the issue of production process and measurement variation or variance and how to evaluate these components. Table 1 illustrates four testing scenarios for sampling variance, S^2 (sampling), and measurement variance, S^2 (measurement). The importance or significance of either component is determined in large part by the magnitude of the expected difference, *d*, for a simple comparison of the measured parameter values for two different or potentially different lots or populations.

5.3.3 Type 1 is encountered when the expected difference, *d*, is large. For this situation the variance of neither component is critical. Type 2 is typical of a less precise test where sampling variation is low. If the test is in control, and the variance known, the number of measurements needed can be calculated for any desired level of confidence for *d*.

5.3.4 Type 3 is characteristic of a relatively precise test measurement where several samples are required to give a good estimate of lot or population properties needed to calculate *d*. A defined sampling program is required. Only a few measurements (one, two) need be made on any sample.

TABLE 1 Four Testing Variation Component Scenarios

Type of Scenario	Component[A]	
	S^2 (sampling)	S^2 (measurement)
1	Not significant	Not significant
2	Not significant	Significant
3	Significant	Not significant
4	Significant	Significant

[A] Not Significant = no significant or large variation component. Significant = a significant or large variation component.

5.3.5 Type 4 is the most complex since both components are important or significant. This is unfortunately frequently encountered in much testing. A specified sampling plan with multiple samples is required as well as multiple measurements on each sample. Substantial background knowledge in addition to a formal analysis of variance for such a test situation is often required for an efficient evaluation of *d*.

6. Keywords

6.1 lot; population; replicate; replication; sample; sampling; testing

ANNEX

(Mandatory Information)

A1. STATISTICAL MODEL FOR TESTING OPERATIONS

A1.1 Background

A1.1.1 The purpose of this annex is to present some of the more fundamental concepts and definitions used and implied when test measurement variation is discussed. Using a mathematical model format improves understanding and more clearly demonstrates how the various concepts relate to each other. In the annex, some of the words or terms are given a specific definition; some are informally defined in the text in the way they are used.

A1.1.2 In the real world of measurement, all measurement values are perturbed to some degree by a "system-of-causes" that produces error or variation in operation of the instruments or machines used for the testing procedure. The word "machine" is used in a broad context, that is, any device that generates test data. There are two general variation categories for any system:

A1.1.2.1 *Production Variation*—Deviations in certain properties that are (*1*) inherent in the process that produces or generates the different classes of objects or materials being tested, or (*2*) acquired deviations (storage or conditioning effects) after such processes are complete.

A1.1.2.2 *Measurement Variation*—Deviations in the operation of instruments or machines that evaluate certain properties for any class of objects or material; these deviations perturb the observed values for these properties.

A1.1.3 The system-of-causes is defined by the scope and organization of any testing program and by the replication and sampling operations that are part of the program. These systems can vary from simple to very complex.

A1.1.4 The production process is broadly defined and can be (*1*) the ordinary operation of a manufacturing facility, (*2*) a naturally occurring process, or (*3*) some smaller scale processing or other procedure that generates a material or class of objects for testing.

A1.1.5 This annex is drafted to apply to both objects and materials. Objects may be discrete items such as o-rings or test specimens generated by a particular preparation process. Materials may be tested in a direct manner, such as the moisture content of a rubber or rubber chemical, or in an indirect manner, such as the quality of a carbon black via a physical property in a standard rubber formulation. In the case of direct testing for a bulk material, an appropriate sample taken from the lot is tested. In the case of indirect testing, the material tested is usually combined with other materials in a specified way and the composite is tested. This composite testing may involve objects or test specimens for the measurement process.

A1.2 General Model

A1.2.1 For any established "system-of-causes," each measurement, *y* (*i*), can be represented as a linear additive combination of fixed or variable (mathematical) terms as indicated by Eq A1.1. Each of these terms is an individual component of variation and the sum of all components is equal to the total variation observed in the individual measurement procedure. The equation applies to any brief time period of testing for a standardized test procedure. All participants test a number of classes of objects (each class having a number of individual objects) or different materials, drawn from a lot of some specified uniformity, employ the same type of apparatus, use skilled operators, and conduct testing in a typical laboratory or test location.

$$y\,(i) = \mu(o) + \mu(j) + \sum(b) + \sum(e) + \sum(\beta) + \sum(\varepsilon) \quad \text{(A1.1)}$$

where:

$y\,(i)$ = a measurement value, at time (*i*), using specified equipment and operators, at laboratory or location (*q*),

$\mu(o)$ = a general or constant term (mean value) unique to the type of test being used,

$\mu(j)$ = a constant term (mean value) unique to material or object class (*j*),

$\sum(b)$ = the (algebraic) sum of some number of individual *bias deviations* in the process that produced material or object class (*j*),

$\sum(e)$ = the (algebraic) sum of some number of individual *random deviations* in the process that produced material or object class (*j*),

$\sum(\beta)$ = the (algebraic) sum of some number of individual *bias deviations,* for measurement (i), generated by the measurement system, and
$\sum(\varepsilon)$ = the (algebraic) sum of some number of individual *random deviations*, for measurement (i), generated by the measurement system.

A1.2.2 Eq A1.1 indicates that there are three main generic variation components: (*1*) constant terms (population mean values), (*2*) bias deviation terms, and (*3*) random deviation terms. These three are discussed in detail in succeeding sections.

A1.3 Specific Model Format

A1.3.1 A more useful format is obtained when Eq A1.1 is expressed using Eq A1.2, where the summations are replaced by a series of typical individual terms appropriate to interlaboratory or different location testing on a number of different object classes or materials, for a particular time period sufficient to complete the testing. This permits greater insight into the model and how it relates to real testing situations.

$$y\,(i) = \mu(o) + \mu(j) + \sum b + \sum e + \beta(L) + \beta(E) + \beta(OP) + \varepsilon(E) + \varepsilon(OP) \quad \text{(A1.2)}$$

where:

$\beta(L)$ = a bias deviation term unique to laboratory or location (q),
$\beta(E)$ = a bias deviation term unique to the specific instrument or machine,
$\beta(OP)$ = a bias deviation term unique to the operator(s) conducting the test,
$\varepsilon(E)$ = a random deviation in the use of the specific instrument or machine, and
$\varepsilon(OP)$ = a random deviation inherent in operator's technique.

Other types of testing perturbations not included in Eq A1.2 may exist, for example, bias and random components due to temperature and the time of the year that testing is conducted.

A1.3.2 *The* $\mu(o) + \mu(j)$ *Terms*—In the absence of bias or random deviations of any kind, a number of materials or object classes would have individual measured test values given by the sum of the two terms, $\mu(o) + \mu(j)$. The term $\mu(o)$ would be unique to whatever test was being employed and each material or object class would be characterized by the value of $\mu(j)$, which would produce a varying value for the sum $[\mu(o) + \mu(j)]$ across the number of materials or object classes in the test program. The sum would be the "true" test value, that is, without any error or variation of any sort.

A1.3.3 *The Production Terms* $\sum(b) + \sum(e)$—There will always be some bias and random variation in the materials or object classes produced by the process that generates them. This usually unknown number of bias and random variations is designated by $\sum(b) + \sum(e)$. The goal of some sampling plans may be the evaluation of either of these sources of variation. In other testing operations the goal may be to reduce such variation to the lowest possible level. Appropriate sampling plans can usually reduce the random components. It is often more difficult to reduce the bias components.

A1.3.4 *The Measurement Bias* (β) *Terms*—The classic statistical definition of a bias is "the difference between the average measured test result and the accepted reference value (true value); it measures in an inverse manner the accuracy of a test" (see Practice D4483). Bias deviations are non-random components and for a series of extended measurements (a long run) the value of bias terms may be either fixed or variable as well as positive or negative, depending on the system-of-causes. The variable bias terms are typically a non-random finite distribution which, in the long run, give a non-zero average. Biases or bias deviations are what make one laboratory, location, or test instrument different in comparison to other laboratories, locations, or instruments.

A1.3.4.1 Bias terms that are fixed under one "system-of-causes" may be variable under another "system-of-causes" and vice-versa. As an example, consider the bias terms $\beta(L)$ and $\beta(E)$ which apply to most types of testing. For a particular laboratory (with one test machine) both of these bias terms would be constant or fixed. For a number of test machines, all of the same design in a given laboratory, $\beta(L)$ would be fixed, but $\beta(E)$ would be variable, each machine potentially having a unique value. For a measurement system consisting of a number of typical laboratories, each with one machine, both $\beta(L)$ and $\beta(E)$ would be variable for the multilaboratory "system-of-causes," but both $\beta(L)$ and $\beta(E)$ would be fixed or constant for the "system-of-causes" in each laboratory.

A1.3.5 *The Measurement Random* (ε) *Terms*—These are the components that are frequently called error. Random deviations are positive or negative values that have an expected mean (average) of zero over the long run. The distribution of these terms is assumed to be approximately normal, but in practice it is usually sufficient if the distribution is unimodal. The value of each random term influences the measured y(i) value on an individual measurement basis. However, in the long run when $y(i)$ values are averaged over a substantial number of measurements, the influence of the random terms may be greatly diminished or eliminated depending on the sampling and replication plan, since each term averages out to zero (or approximately zero) and the average $y(i)$ is essentially unperturbed. In ordinary testing the magnitude of the individual bias and random components or deviations are not known. Their collective effect influences each measured $y(i)$ value and this collective effect is what is normally evaluated in variance testing.

A1.3.6 *Test Replication*—There are three general types of sample replication procedures that apply to testing, where the word" item" refers to an object or a test sample (part) of a bulk material.

Type 1—sample replication (m) = using the same test item with 1 to m repeated tests
Type 2—sample replication (n,1) = using n test items, each item being tested one time
Type 3—sample replication (n,m) = using n test items, each item being tested m times

For Type 1, the sample size is 1, with m replicates; for Types 2 and 3 the sample size is n, also with m replicates. The scope of the sampling and replication plan needs to be clearly defined for any testing program. Replication Types 1 (with m tests) and

3 may be used for nondestructive testing, while Type 2 is the only type available for multi-sample destructive testing. Type 3 testing reduces the influence of the production random variation as well as the random measurement variation.

A1.3.6.1 Replicated testing of any type with only a few replicates (where *n* and *m* jointly or each equal less than 10) gives a test result average value, *Y* (*n,m* < 10), as indicated by Eq A1.3, where the appearance of $\sum(e)$ and $\sum(\varepsilon)$ indicates that these sums are not equal to zero. Usually $\sum(\varepsilon)$ and $\sum(e)$ are much less than $\sum(b)$ and $\sum(\beta)$.

$$Y(n,m<10) = \mu(o)+\mu(j)+\sum(b)+\sum(e)+\sum(\beta)+\sum(\varepsilon) \tag{A1.3}$$

A1.3.6.2 Highly replicated testing (10 or more measurements for both *n* and *m*) reduces the perturbation of the random deviations to near zero and thus the test result average, *Y*(*n,m* > 10), is given by Eq A1.4, which is perturbed by only bias components.

$$Y(n,m>10) = \mu(o)+\mu(j)+\sum(b)+\sum(\beta) \tag{A1.4}$$

A1.3.6.3 Eq A1.4 shows that ordinary highly replicated testing (usually Type 3) does not approximate the "true value" for any candidate if any production or measurement system bias deviations exist. The tester ordinarily does not know of the potential sources of this inherent process and measurement bias variation and no individual assignment of variation components can be made. These terms remain in their generalized format.

A1.3.7 *New Term, M(j)*—With highly replicated testing programs (both production and test measurement replication) the average values obtained in any program are estimates or very close approximations to the value of a new combined term as given by Eq A1.5.

$$M(j) = \left[\mu(o)+\sum(b)+\sum(\beta)\right]+\mu(j) \tag{A1.5}$$

M(*j*) is the mean value for the material or class of objects being tested, for laboratory or location (*q*), for the specific equipment and operators used during the existing time period and it contains bias components or potential bias components for all of these conditions. If all biases are fixed for any given program, the three terms in the bracket can be considered as a constant and the average test value varies across the number of materials or object classes because of the varying value of μ(*j*). If there are variable biases, then both μ(*j*) and the biases influence the average value for any candidate.

A1.4 Evaluating Process and Measurement Variance

A1.4.1 Eq A1.1 may be used to illustrate how the variance of individual measurements, *y* (*i*), may be related to the terms or components of the equation. Recall that μ(*o*) and μ(*j*) are constants, $\sum(b)$ and $\sum(e)$ refer to the sum of bias and random components, respectively, for the production process and $\sum(\beta)$ and $\sum(\varepsilon)$ refer to the sum of bias and random components, respectively, for the test measurement operation. The magnitude of the individual components are ordinarily not known and the equation can be simplified by combining the bias and random components for both sources.

$$y(i) = \mu(o)+\mu(j)+\sum(b,e)+\sum(\beta,\varepsilon) \tag{A1.6}$$

where:

$\sum(b,e)$ = sum of bias and random components for the production process, and
$\sum(\beta,\varepsilon)$ = sum of bias and random components for the measurement procedure.

A1.4.2 The variance of any individual measurement *y* (*i*), designated by Var [*y* (*i*)], is given by Eq A1.7.

$$\mathrm{Var}\left[y\ (i)\right] = \left[\sum \mathrm{Var}\ (b,e)\right]+\left[\sum \mathrm{Var}\ (\beta,\varepsilon)\right] \tag{A1.7}$$

where:

[$\sum$ Var (*b,e*)] = a variance that is the sum of individual bias and random variances for the production process, and
[$\sum$ Var (β,ε)] = a variance that is the sum of individual bias and random variances for the measurement procedure.

Eq A1.7 can be written in simplified format as indicated in Eq A1.8, using the conventional symbol S^2 for the variance.

$$S^2\,(\mathrm{tot}) = S^2\,(p)+S^2\,(m) \tag{A1.8}$$

where:

S^2 (tot) = total variance among the materials or object classes in a test program,
S^2 (*p*) = variance due to the production process, and
S^2 (*m*) = variance due to the measurement operation.

For the measurement situation where testing is nondestructive and any sample may be tested more than one time, all three variance components can be evaluated. Table A1.1 for a typical testing scenario will help in illustrating this. There are (*k*) materials or object classes tested, each has four samples (*n* = 4) and each sample has two replicates (*m* = 2). Each pair of *y* (*ij*) values constitutes a cell in the table.

A1.4.3 There are $(k) \times 8$ individual test values and the variance for all of these values is S^2 (tot). The variance S^2 (*m*) is evaluated by taking the variance for each cell in the table (each cell has 1 DF) and pooling this across all cells for all materials or classes. The variance S^2 (*p*) is evaluated by difference as given in Eq A1.9.

$$S^2\ (p) = S^2\,(\mathrm{tot}) - S^2\,(m) \tag{A1.9}$$

This approach to production process and test measurement variance evaluation assumes that the replicate (within cell) testing variance is equal in the long run for all materials or object classes. The value of S^2 (*p*) as obtained from this analysis is a collective value and represents the influence of bias and random variation for all materials or object classes.

TABLE A1.1 Illustration of Typical Testing Scenario

Candidate Material (or Object Class)	Sample Number			
	1	2	3	4
A	y11, y12	y21, y22	y31, y32	y41, y42
B	etc.	...	...	...
...	...	...	...	...
k	...	...	...	...