Designation: D 3244 – 97 (Reapproved 2002)

An American National Standard

Standard Practice for Utilization of Test Data to Determine Conformance with Specifications[1]

This standard is issued under the fixed designation D 3244; 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.

INTRODUCTION

The properties of commercial petroleum products are measured by standardized laboratory test methods to check their conformance to specifications. Two or more measurements of the same property of a specific sample by any given test method usually will not give precisely the same answer. Therefore, the test methods generally include a paragraph on the precision of results. This precision is an expression of the reliability of the value of the measured property.

Many difficulties that arise in interpreting specifications are due to test imprecision. Because of this, a true value of a property can never be determined exactly; and it is necessary to infer from measured values the range within which the "true value" is likely to lie. The main purpose of this practice is to indicate how test imprecision should be interpreted relative to specification values.

1. Scope

1.1 This practice covers guidelines with which two parties, usually a supplier and a receiver, can compare and combine independently obtained test results whenever there is a product quality dispute.

1.2 This practice defines a technique for comparing an assigned test value with a specification limit.

1.3 This practice applies only to those test methods which specifically state that the repeatability and reproducibility values conform to the definitions herein.

2. Referenced Documents

2.1 *ASTM Standards:*

D 1319 Test Method for Hydrocarbon Types in Liquid Petroleum Products by Fluorescent Indicator Adsorption[2]

D 4057 Practice for Manual Sampling of Petroleum and Petroleum Products[3]

D 4177 Practice for Automatic Sampling of Petroleum and Petroleum Products[3]

E 29 Practice for Using Significant Digits in Test Data to Determine Conformance with Specifications[4]

2.2 *ISO Standard:*[5]

ISO 4259 Determination and Application of Precision Data in Relation to Methods of Test

3. Terminology

3.1 *Definitions:*

3.1.1 *acceptance limit (AL)*, *n*—a numerical value that defines the point between acceptable and unacceptable quality.

3.1.1.1 *Discussion*—The AL is not necessarily the specification limit. It is a value that takes into account the specification value, the test method precision, and the confidence level desired for defining minimum acceptable quality relative to the specification value.

3.1.2 *assigned test value (ATV)*, *n*—the average of all results obtained in the several laboratories which are considered acceptable based on the reproducibility of the test method.

3.1.3 *determination*, *n*—the process of carrying out the series of operations specified in the test method whereby a single value is obtained.

3.1.4 *dispute*, *n*—when there is a question as to product quality because a test value obtained falls outside the acceptance limit.

3.1.5 *operator*, *n*—a person who normally and regularly carries out a particular test.

3.1.6 *precision*, *n*—the degree of agreement between two or more results on the same property of identical test material. In

[1] This practice is under the jurisdiction of ASTM Committee D02 on Petroleum Products and Lubricants and is the direct responsibility of Subcommittee D02.94 on Quality Assurance and Statistics.

Current edition approved Dec. 10, 2002. Published March 2003. Originally published as an appendix to the *1968 Annual Book of ASTM Standards*, Part 18. Last previous edition approved in 1997 as D 3244 – 97.

[2] *Annual Book of ASTM Standards*, Vol 05.01.

[3] *Annual Book of ASTM Standards*, Vol 05.02.

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

[5] Available from American National Standards Institute, 25 W. 43nd St., 4th floor, New York, NY 10036.

this practice, precision statements are framed in terms of the repeatability and reproducibility of the test method.

3.1.7 *receiver*, *n*—any individual or organization who receives or accepts the product delivered by the supplier.

3.1.8 *repeatability (r)*, *n*—quantitative expression of the random error associated with a single operator in a given laboratory obtaining replicate results with the same apparatus under constant operating conditions on identical test material within a short period of time. It is defined (3.1.8.1) as that difference between two such single results as would be exceeded in the long run in only 1 case in 20 in the normal and correct operation of the test method (3.1.8.3). (This is known as the 95 % confidence level.)

3.1.8.1 *Discussion*—The repeatability and reproducibility values should have been determined according to the methods described in ASTM Research Report RR:D02-1007, Manual on Determining Precision data for ASTM Methods of Petroleum Products and Lubricants[3] or ISO 4259.

3.1.8.2 *Discussion*—Not all standards organizations define repeatability and reproducibility in precisely these same terms, and attention should always be paid to definitions before comparing precision values quoted.

3.1.8.3 *Discussion*—This difference is related to the repeatability or the reproducibility standard deviation but is not the standard deviation.

3.1.9 *reproducibility (R)*, *n*—quantitative expression of the random error associated with operators working in different laboratories, each obtaining single results on identical test material when applying the same method. It is defined (3.1.8.1) as that difference between two such single and independent results as would be exceeded in the long run in only 1 case in 20 in the normal and correct operation of the test method. See 3.1.8.3.

3.1.10 *result*, *n*—the value obtained by following the complete set of instructions of a test method. It may be obtained from a single determination or several determinations, depending on the instruction of the test method.

3.1.11 *supplier*, *n*—any individual or organization responsible for the quality of a product just before it is taken over by the receiver.

3.1.12 *test sample*, *n*—a portion of the product taken at the place where the product is exchanged, that is, where the responsibility for the product quality passes from the supplier to the receiver. Actually, this is rarely possible and a suitable sampling location should be mutually agreed upon.

3.1.13 *true value (μ)*, *n*—for practical purposes, the value towards which the average of single results obtained by *N* laboratories tends, when *N* becomes very large (3.1.13.1). Consequently, such a true value is associated with the particular test method employed.

3.1.13.1 *Discussion*—It is recognized that there are cases where a true value not equal to the method average can exist. As used in this practice, the method average value is intended to mean "true value" even if the method is biased.

4. Significance and Use

4.1 This practice provides a means whereby the parties to a transaction can resolve potential disputes over those product properties which can be tested and expressed numerically.

4.1.1 This practice can be used to ensure that such properties are correctly stated on labels or in other descriptions of the product.

4.1.2 This practice can be implemented in those cases where a supplier uses a commercial testing laboratory to sample and test a product prior to releasing the product to a shipper (intermediate receiver) and the ultimate receiver also uses a commercial testing laboratory to sample and test the product upon arrival at the destination. The assigned test value (ATV) would still be determined according to 8.3.

4.2 This practice can also assist in the determination of proper tolerances which will ensure that the actual value of a property is sufficiently close to the specification value so that the product is acceptable to the receiver. Such tolerances are bounded by an *acceptance limit* (AL). If the value determined by testing (*assigned test value*, (ATV)) falls on the AL or on the acceptable side of the AL, the product can be accepted; otherwise it must be rejected.

4.3 Both parties must agree in advance on setting the AL and on how the ATV is to be calculated.

4.3.1 This agreement should include a decision as to whether the test values are to be determined by the absolute or rounding-off method of Practice E 29, as therein defined.

4.3.1.1 If the rounding-off method is to be used, the number of significant digits to be retained must also be agreed upon.

4.3.1.2 These decisions must also be made in the case where only one party is involved, as in the case of a label.

4.4 This practice is designed to be suitable for reference in contracts governing the transfer of petroleum products and lubricants from a supplier to a receiver.

4.5 Application of this practice requires the designation of each limit of each property of a specification as "critical" or "noncritical" at a desired probability level, as defined in this practice.

4.6 As a prerequisite for acceptance for lab test results to be used in this practice, the following conditions shall be satisfied:

4.6.1 Long-term standard deviation for the appropriate test method(s) from each lab, as substantiated by in-house quality control programs, on material typical of the product in dispute, shall be statistically equivalent or better than the published method standard deviation under reproducibility conditions.

4.6.2 Each lab shall be able to demonstrate, by way of results from interlaboratory exchange programs, a lack of a statistically significant bias relative to exchange averages for the appropriate test method(s).

4.6.3 In the event that the long-term standard deviation for any party's laboratory is not statistically equivalent, then, for the purpose of establishing the assigned test value (ATV), each laboratory's test result(s) shall be inversely weighted in accordance with laboratory's demonstrated variance(s).

4.7 It is recommended that this practice be conducted under the guidance of a qualified statistician.

5. Sampling

5.1 Sampling should be carried out in accordance with standard sampling procedures for petroleum products (Practice D 4057 and Practice D 4177). Obtain enough sample to allow all required determinations to be made. Divide the sample into three secondary samples: a receiver sample, a supplier sample,

and a retain sample. The retain sample should itself be large enough to permit further subdivision into three portions in case additional test work is desirable.

6. Applying Precision Data to Test Methods

6.1 This section describes procedures in which the precision limits of test methods may be used to indicate when results obtained by two laboratories differ significantly. This section may also be used for rejection of results of replicate tests by an operator.

6.2 *Significance of Repeatability (r)*:

6.2.1 *Acceptance of Results*—When only two results are obtained under conditions of repeatability and the difference is equal to or less than the repeatability of the method, the operator may report the average of the two results as being applicable to the sample tested.

6.2.2 *Rejection of Results*—When two results are obtained that differ by more than the repeatability of the method, both should be rejected. Obtain two additional results immediately under conditions of repeatability. If the difference between these two results is equal to or less than the repeatability of the method, the operator should report the average of the two as being applicable to the sample tested. If, however, the difference so obtained again exceeds the repeatability, reject the results and investigate the application of the method.

6.3 *Significance of Reproducibility (R)*:

6.3.1 *Acceptance of Results*—When two results are obtained in different laboratories (Note 1) and the difference is equal to or less than the reproducibility of the method, both results should be considered acceptable. The value assigned to the sample should be the average of the two results.

NOTE 1—When a comparison for reproducibility is made between results from two laboratories, it is assumed that single results from each will be compared. If each of the laboratories has produced more than a single result, see 6.4.

6.3.2 *Rejection of Results*—When the results from two laboratories differ by more than the reproducibility of the method, reject both results and each laboratory should repeat the test on the retained sample. If the difference is now equal to or less than the reproducibility, both results should be considered acceptable and their average reported. If, however, the difference between these results is still greater than the reproducibility, reject the results and investigate the application of the method at each laboratory.

6.4 *Multiple Testing*— If the number of results obtained in either one or both laboratories is more than one, then the allowable difference between the averages from the two laboratories is given as follows:

$$\text{Difference, } R' = \sqrt{R^2 - r^2\left(1 - \frac{1}{2n_1} - \frac{1}{2n_2}\right)} \tag{1}$$

where:

R = reproducibility of the method,
r = repeatability of the method,
n_1 = number of results of the first laboratory, and
n_2 = number of results of the second laboratory.

6.5 *Referee Laboratory*—In the event a third or referee laboratory is invited to make the test using a portion of one of the samples described in 6.3.2, multiply the reproducibility, R, by 1.2 (to convert a range for two to a range for three) and compare this value with the difference between the two extreme results for acceptance. If acceptance is indicated, the assigned test value (ATV) for the sample should be the average of the three results.

7. Applying Precision Data to Specifications

7.1 *Specifications*— A specification fixes a limit to the *true value* of a given property. In practice, however, this *true value* can never be established exactly. The property is measured in the laboratory by applying a standard test method, the results of which may show some scattering as defined by the repeatability and reproducibility limits. There is always, therefore, some uncertainty as to the *true value* of the tested property.

7.2 Although the *true value* is never known exactly, the probability of obtaining any specific test result, relative to the true value, can be calculated if the probability distribution function for the test method is known (for example, the normal distribution curve with its associated reproducibility).

7.2.1 Some specifications, because of the product characteristic or the end use of the product, or both, require that the receiver have a high degree of assurance that the product actually meets or exceeds the quality level indicated by the specification value. For the purpose of this practice, such specifications are called *critical* specifications.

7.2.2 Specifications that require assurance only that the product quality is not substantially poorer than is indicated by the specification level are called *noncritical* specifications for the purposes of this practice.

7.3 *Specification Conformance Guidelines*:

7.3.1 Whenever a product is tested for conformity to a specification, a decision must ultimately be made as to the acceptance or rejection of the product.

7.3.2 The numerical value that divides the regions of acceptable and unacceptable product test values is the acceptance limit (*AL*). The *AL* may or may not coincide with the specification value (*S*) used to define a product quality or grade.

7.3.3 The *AL* value, which must be agreed upon by the supplier and receiver, is that level of quality such that, if the *true value* is exactly *AL,* they are willing to take a 50 % chance of either accepting or rejecting the product as tested.

7.3.4 In the absence of an agreement to the contrary, the specification will be considered a noncritical specification for which there is 95 % assurance that the product will be accepted if the true quality is the specification value. Thus, the *AL* will be set by using a confidence level $P = 0.95$ as shown in 7.3.6.

7.3.5 The probability of accepting a product (deciding that product quality is acceptable) when the true value equals the specification value is shown in Table 1 and Fig. 1 as a function of $D = (AL - S)/0.255R$, a direct measure of the difference between *AL* and *S*. This relationship is based (*1*) on the assumption of normally (Gaussian) distributed testing errors, which is adequate for most test procedures, and (*2*) on using an assigned test value (ATV) for making the specification conformance decision that is the average of precision-acceptable results from two laboratories.

7.3.6 Instead of deciding directly on an *AL,* the receiver may select a given probability *P* of accepting the product when the

TABLE 1 Deviation of AL from Specification for Product Acceptance at a Given Probability

NOTE 1—Based on $N = 2$ = number of different laboratories' results used to obtain ATV. See text for use of this table.

		$D = (AL - S)/0.255\ R$	
	Probability (P) of Acceptance	Maximum Specification Limit	Minimum Specification Limit
Critical	0.001	−3.090	3.090
	0.005	−2.576	2.576
	0.010	−2.326	2.326
	0.025	−1.960	1.960
	0.050	−1.645	1.645
	0.100	−1.282	1.282
	0.150	−1.036	1.036
	0.200	−0.842	0.842
	0.300	−0.524	0.524
Noncritical	0.500	0.000	0.000
	0.700	0.524	−0.524
	0.800	0.842	−0.842
	0.850	1.036	−1.036
	0.900	1.282	−1.282
	0.950	1.645	−1.645
	0.975	1.960	−1.960
	0.990	2.326	−2.326
	0.995	2.576	−2.576
	0.999	3.090	−3.090

true value equals the specification value S. From the relationship given, read a value D corresponding to the ordinate value P. The proper AL is then given by

$$AL = S + 0.255 \times R \times D \quad (2)$$

For N, other than two different laboratory results, the 0.255 factor should be multiplied by $\sqrt{2/N}$.

7.3.6.1 For specifications having both minimum and maximum limits, the procedure in 7.3.6 must be applied twice to give both upper and lower ALs. There must be some allowable region remaining between the lower and upper ALs.

7.3.7 When only a single test result is or will be available, the relationships given should be used with $N = 1$ (7.3.6). Obviously, no check on reproducibility precision can be made with a single test result, and the single value becomes the ATV for the sample.

7.3.8 The AL for critical specifications is thus set so that if the true value is less than or equal to S, there is only a small probability (defined by selection of P) of accepting the product.

7.3.9 For noncritical specifications, the AL is set so that if the true value is equal to or greater than S, there is a high probability (defined by selection of P) of accepting the product.

7.3.10 The relationships between the ALs for critical and noncritical specifications are shown in Fig. 2 for a minimum specification.

8. Obtaining the Assigned Test Value (ATV)

8.1 The following procedure will produce an ATV with precision control based on the reproducibility of the test method.

8.2 The receiver and supplier should obtain independent test results, X_R and X_S, respectively.

NOTE 2—The supplier's result must be on the *test sample* (see Section 5) and not a reported value by the supplier. In many cases, a reported value by the supplier is obtained on a different sample, for example, at point of manufacture, and may be the average of several determinations.

8.3 *ATV Procedure*:

8.3.1 If the absolute value of $\Delta = X_R - X_S$ is less than or equal to R, the reproducibility of the test method, average the two results to obtain the following in accordance with 6.3.1:

$$ATV = (X_R + X_S)/2 \quad (3)$$

8.3.2 If the absolute value of Δ exceeds R, reject both results and retest on portions of the retain sample to obtain X_R', X_S'.

8.3.3 If the absolute value of $\Delta' = X_R' - X_S'$ is less than or equal to R, average the two results to obtain the following in accordance with 6.3.2:

$$ATV = (X_R' + X_S')/2 \quad (4)$$

8.3.4 If the absolute value of Δ' exceeds R, obtain a new test value X_{RL} from a referee laboratory (6.5).

8.3.5 If $\Delta_3 = X_{max} - X_{min}$ is less than or equal to 1.2 R, obtain the following:

$$ATV = (X_R' + X_S' + X_{RL})/3 \quad (5)$$

8.3.6 If Δ_3 exceeds 1.2 R, obtain ATV as the average of the closer pair.

NOTE 3—This last step for obtaining an ATV does not comply rigidly to statistical concepts. It is done in this manner because in most cases the test sample (see Section 5) is depleted.

8.4 The above procedure will always yield an ATV. If the supplier's and receiver's laboratories have little or no bias relative to each other, then the procedure will end at 8.3.1 about 95 % of the time, and some 95 % of the remaining 5 %, at 8.3.3.

8.5 If any particular supplier and receiver pair find they frequently must go as far as calling for a reference laboratory test, they should carefully check their running of the test and, if possible, calibrate their results with other laboratories.

8.6 This procedure for obtaining an ATV is designed for the test of samples obtained according to Section 5.

8.6.1 If more extensive testing is needed for special situations, comparable procedures can be developed. A statistician or quality control expert should be consulted to do this.

9. Product Quality Conformance

9.1 A product should be considered as conforming to the specifications if the ATV of each property meets the AL value.

9.2 The supplier should ship product only if there is confidence that each property meets specification values.

9.3 When the receiver has obtained a single result, the product quality should be considered suspect if the test result fails the AL value (see A3.1.5).

9.4 A dispute between supplier and receiver may arise whenever a receiver's result fails the AL value.

9.5 The dispute should be resolved by obtaining an assigned test value (ATV) for the product as an estimate of the "true value" and comparing this to the acceptance limit (AL) as determined in 7.3.

10. Acceptance or Rejection of Product

10.1 If the ATV is equal to or better than the AL limit, the product is to be accepted as having met specification.

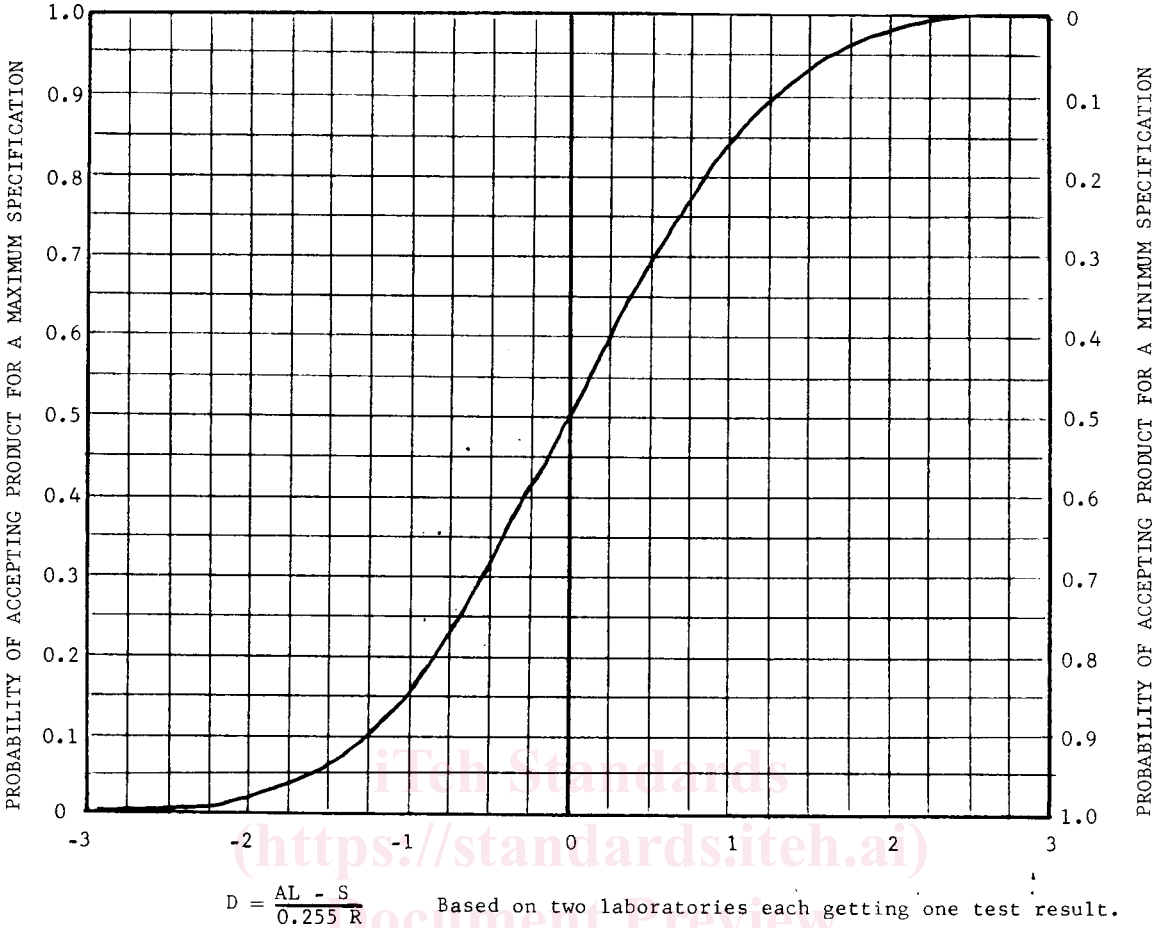
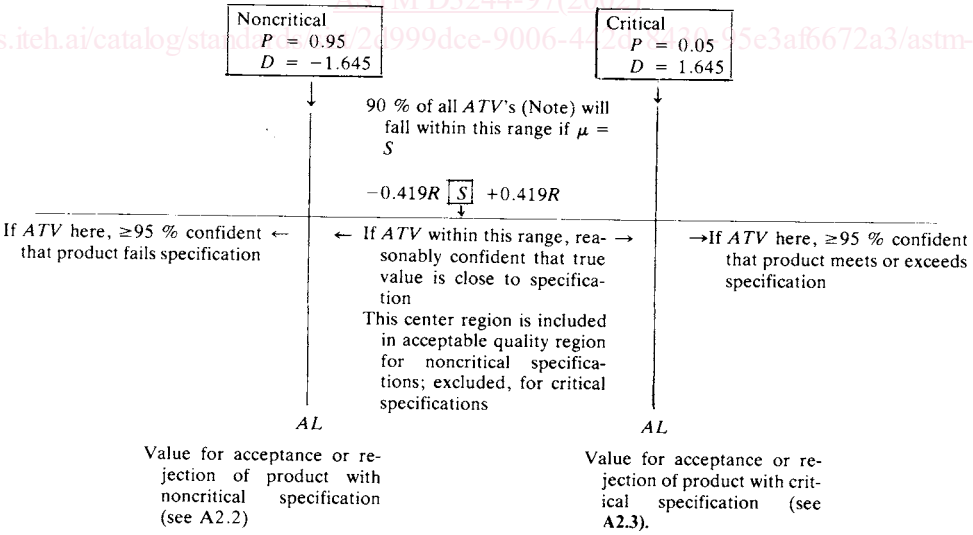


FIG. 1 Probability of Acceptance vs Deviation of *AL* from True Value = S



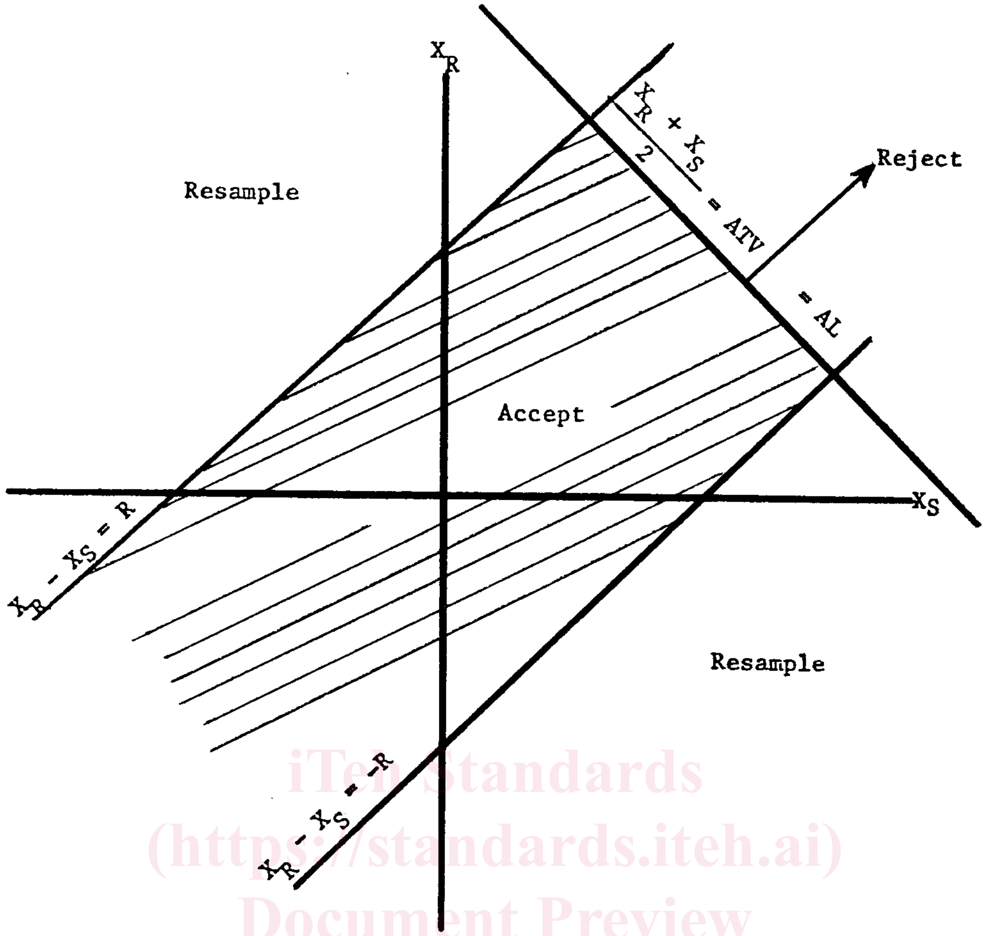
NOTE 1—This applies when ATV is established by the average of two results, one each from two different laboratories.

FIG. 2 Relationships Between *AL* s for Critical and Noncritical Specifications

10.2 If the ATV fails the AL value, the product is to be rejected as failing specification.

10.3 These concepts are presented graphically in Fig. 3.

10.3.1 The plotted lines are boundary conditions separating acceptable results from those indicating other alternative actions.



X_R = receiver's result
X_S = shipper's result
R = test reproducibility
AL = acceptance limit

FIG. 3 Diagram Showing Regions of Acceptance, Rejection, and Resampling

10.3.1.1 The sample is considered acceptable if the two results fall to the left of the line, $(X_R + X_S)/2 = ATV \ = AL$ if they are also within the lines. $X_R - X_S = \pm\ R$.

10.3.2 The sample is unacceptable if the results lie to the right of the line $(X_R + X_S)/2 = ATV \ = AL$.

10.3.3 Initial results falling in the region labeled *resample* call for a retest.

10.3.3.1 If results for a second sample also fall in the resample region, a reference laboratory should be included in the new testing program.

10.4 The actual consequences of rejecting a product for failure to meet specification are subject to prior agreement or negotiation between the parties concerned.

11. Keywords

11.1 acceptance; acceptance limits; agreement; conformance; dispute; precision; rejection; specifications