



Designation: E 180 – 99

## Standard Practice for Determining the Precision of ASTM Methods for Analysis and Testing of Industrial and Specialty Chemicals<sup>1</sup>

This standard is issued under the fixed designation E 180; 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 ( $\epsilon$ ) indicates an editorial change since the last revision or reapproval.

### 1. Scope

1.1 This practice establishes uniform standards for expressing the precision and bias of test methods for industrial and specialty chemicals. It includes an abridged procedure for developing this information, based on the simplest elements of statistical analysis. There is no intent to restrict qualified groups in their use of other techniques.

1.2 *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:

- D 1013 Test Method for Total Nitrogen in Resins and Plastics<sup>2</sup>
- D 1727 Test Method for Urea Content of Nitrogen Resins<sup>3</sup>
- E 29 Practice for Using Significant Digits in Test Data to Determine Conformance with Specification<sup>4</sup>
- E 177 Practice for Use of the Terms Precision and Bias in ASTM Test Methods<sup>4</sup>
- E 178 Practice for Dealing with Outlying Observations<sup>4</sup>
- E 456 Terminology Relating to Quality and Statistics<sup>4</sup>
- E 691 Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method<sup>4</sup>
- E 1169 Guide for Conducting Ruggedness Tests<sup>4</sup>

### 3. Significance and Use

3.1 All test methods require statements of precision and bias. The information for these statements is generated by an interlaboratory study (ILS). This practice provides a specific design and analysis for the study, and specific formats for the precision and bias statements. It is offered primarily for the guidance of task groups having limited statistical experience.

3.2 It is recognized that the use of this simplified procedure will sacrifice considerable information that could be developed through other designs or methods of analyzing the data. For example, this practice does not afford any estimate of error to be expected between analysts within a single laboratory. Statements of precision are restricted to those variables specifically mentioned. Task groups capable of handling the more advanced procedures are referred to the literature (1, 2, 3, 5, 13)<sup>5</sup> and specifically to Practice E 691, the current Committee E-11 practice for interlaboratory studies. The latter includes graphical display and interpretation of ILS data.

3.3 The various parts appear in the following order:

- Part A—Glossary.
- Part B—Preliminary Studies.
- Part C—Planning the Interlaboratory Study.
- Part D—Testing for Outlying Observations.
- Part E—Statistical Analysis of Collaborative Data.
- Part F—Format of Precision Statements.
- Part G—Bias (Systematic Error).
- Part H—Presentation of Data.

### 4. Keywords

4.1 bias; industrial chemicals; interlaboratory study; precision

## PART A—GLOSSARY

### 5. Scope

5.1 The following statistical terms are defined in the sense in which they will be used in presenting precision and bias information. These definitions have been simplified and are not necessarily universally acceptable nor as defined in Terminology E 456 and Practice E 177. For definitions and explanations of other statistical terms used in this practice, refer to Terminology E 456 and Practice E 177.

### 6. Terminology

#### 6.1 Definitions and Descriptions of Terms:

6.1.1 *accuracy*—the agreement between an experimentally determined value and the accepted reference value. In chemical work, this term is frequently used to express freedom from

<sup>1</sup> This practice is under the jurisdiction of ASTM Committee E15 on Industrial and Specialty Chemicals and is the direct responsibility of Subcommittee E15.01 on General Standards.

Current edition approved Sept. 10, 1999. Published December 1999. Originally published as E 180 – 61 T. Last previous edition E 180 – 94.

<sup>2</sup> Annual Book of ASTM Standards, Vol 06.03.

<sup>3</sup> Discontinued. See 1983 Annual Book of ASTM Standards, Vol 06.01.

<sup>4</sup> Annual Book of ASTM Standards, Vol 14.02.

<sup>5</sup> The boldface numbers in parentheses refer to the list of references at the end of this practice.



bias, but in other fields it assumes a broader meaning as a joint index of precision and bias (see Practice E 177 and (4)). To avoid confusion, the term “bias” will be used in appraising the systematic error of test methods for industrial chemicals.

**6.1.2 bias**—a constant or systematic error as opposed to a random error. It manifests itself as a persistent positive or negative deviation of the method average from the accepted reference value.

**6.1.3 coefficient of variation**—a measure of relative precision calculated as the standard deviation of a series of values divided by their average. It is often multiplied by 100 and expressed as a percentage.

**6.1.4 duplicates**—two independent determinations performed by one analyst at essentially the same time.

**6.1.5 error**—in a statistical sense, any deviation of an observed value from the true, but generally unknown value. When expressed as a fraction or percentage of the value measured, it is called a relative error. All statements of precision or bias should indicate clearly whether they are expressed in absolute or relative sense.

**6.1.6 95 % limit (difference between two results)**—the maximum absolute difference expected for approximately 95 % of all pairs of results from laboratories similar to those in the interlaboratory study.

**6.1.7 precision**—the degree of agreement of repeated measurements of the same property. Precision statements in ASTM methods for industrial and specialty chemicals will be derived from the estimated standard deviation or coefficient of variation of a series of measurements and will be expressed in terms of the repeatability; the within-laboratory, between days variability; and the reproducibility of a method (see 6.1.14, 6.1.3, 6.1.10, 6.1.16, 6.1.12).

**6.1.8 random error**—the chance variation encountered in all experimental work despite the closest possible control of variables. It is characterized by the random occurrence of both positive and negative deviations from the mean value for the method, the algebraic average of which will approach zero in a long series of measurements.

**6.1.9 range**—the absolute value of the algebraic difference between the highest and the lowest values in a set of data.

**6.1.10 repeatability**—the precision of a method expressed as the agreement attainable between two independent determinations performed at essentially the same time (duplicates) by one analyst using the same apparatus and techniques. (see also 6.1.17.)

**6.1.11 replicates**—two or more repetitions of a test determination.

**6.1.12 reproducibility**—the precision of a method expressed as the agreement attainable between determinations performed in different laboratories (12).

**6.1.13 result**—a value obtained by carrying out the test method. The value can be a single determination, an average of duplicates, or other specified grouping of replicates.

**6.1.14 significance level**—the decimal probability that a result will exceed the critical value. (see 21.3 and 21.4.)

**6.1.15 standard deviation**—a measure of the dispersion of a series of results around their average, expressed as the positive square root of the quantity obtained by summing the squares of

the deviations from the average of the results and dividing by the number of observations minus one. It is also the square root of the variance and can be calculated as follows:

$$s = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n - 1}} \quad (1)$$

where:

$s$  = estimated standard deviation of the series of results,

$X_i$  = each individual value,

$\bar{X}$  = average (arithmetic mean) of all values, and

$n$  = number of values.

The following forms of this equation are more convenient for computation, especially when using a calculator:

$$s = \sqrt{\frac{\sum X^2 - (\sum X)^2/n}{n - 1}} \quad (2)$$

or

$$s = \sqrt{\frac{n\sum X^2 - (\sum X)^2}{n(n - 1)}} \quad (3)$$

where:

$s$  = estimated standard deviation,

$\sum X^2$  = sum of the squares of all of the individual values,

$(\sum X)^2$  = square of the total of the individual values, and

$n$  = number of values.

**NOTE 1**—Care must be taken in using either of these equations that a sufficient number of decimal places is carried in the sum of the values and in the sum of their squares so that serious rounding errors do not occur. For best results, all rounding should be postponed until after a value has been obtained for  $s$ .

In this practice, the standard deviation is obtained from the difference between duplicate determinations and from an analysis of variance of an interlaboratory test program (see Part E).

**6.1.16 variance**—a measure of the dispersion of a series of results around their average. It is the sum of the squares of the individual deviations from the average of the results, divided by the number of results minus one.

**6.1.17 within-laboratory, between-days variability (formerly called repeatability)**—the precision of a method expressed as the agreement attainable between independent determinations (each the average of duplicates) performed by one analyst using the same apparatus and techniques on each of two days. (This term is further defined and limited in 10.1.6, 25.1, and 25.2.9.2) (12).

## PART B—PRELIMINARY STUDIES

### 7. Scope

7.1 This part covers the preliminary work that should be carried out in a few laboratories before undertaking a full interlaboratory evaluation of a method.

### 8. Discussion

8.1 When a task group is asked to provide a specific test procedure, there may be available one or more methods from the literature or from laboratories already performing such analyses. In such cases, these methods have usually been the subject of considerable research and any additional study

variables, at this stage, would be wasteful of available task group time. It is recommended that such methods be rewritten in ASTM format, with full descriptions of the equipment and procedure, and be evaluated in a pilot run by a few laboratories on selected materials. Three laboratories and at least three such materials, using one or two analysts performing duplicate determinations on each of two days, by each method, constitutes a practical plan which can be analyzed by the procedures described in Part E—Statistical Analysis of Collaborative Data. Such a pilot study will confirm the adequacy of the methods and supply qualitative indications of relative precision and bias.

8.2 When the method to be evaluated is new, or represents an extensive modification of an available method, it is recommended that a study on variables be carried out by at least one laboratory to establish the parameters and conditions to be used in the description of the method. This should be followed by a three-laboratory pilot study before undertaking a full interlaboratory evaluation.

8.3 Detailed procedures for executing such preliminary studies are not described in this practice but are available in the general statistical literature.<sup>6</sup> Practice E 691 and Guide E 1169 also provide information on this subject.

## PART C—PLANNING THE INTERLABORATORY STUDY

### 9. Scope

9.1 This part covers some commonsense recommendations for the planning of interlaboratory studies.

### 10. Variables

10.1 The major variables to be considered are the following: methods, materials or levels, laboratories, apparatus, analysts, days, and runs. These are discussed as follows:

10.1.1 *Methods*—The preliminary studies of Part B should lead to agreement on a single method, which can then be evaluated in a full interlaboratory study. If it is necessary to evaluate two or more methods, the complete program must be carried out on each such method. In either case, it will be assumed that the method variables have been explored and that a well-standardized, fully detailed procedure has been prepared. Nothing short of this will justify the time and expense required for an extensive precision study.

10.1.2 *Materials or Levels*—The number of samples distributed should be held to the minimum needed to evaluate the method adequately. (Increasing the number of samples will not increase significantly the degrees of freedom (see 25.2.8) available for predicting the reproducibility of the method. This can be achieved only by increasing the number of laboratories.) Some interlaboratory studies can be limited to a single sample, as in the case of preparing a specific standard solution. Methods applicable to a single product of high purity can usually be evaluated with one or two samples. When different

concentrations of a constituent or values of a physical property are involved, the samples should represent the approximate lower, middle, and top levels of the expected range. If these vary over a wide range, the number of levels should be increased and spaced to cover the range. If technical grade products are used in a precision study, the bias of the method may be undeterminable unless the accepted reference value and its limits of error are known from other sources. For this reason, it is well to include one or more samples of known purity in the interlaboratory study.

10.1.3 *Laboratories*—To obtain a reliable precision estimate, it is recommended that the interlaboratory study include approximately ten qualified laboratories.<sup>7</sup> When this number of independent laboratories cannot be recruited, advantage can be taken of a liberalized definition of collaborating laboratories, quoted as follows from the *ASTM Manual for Conducting an Interlaboratory Study of a Test Method (STP 335)*, p. 9 (5):

Here the term “collaborating laboratory” has a more specific meaning than in common usage. For example, a testing process often consists of an integrated sequence of operations using apparatus, reagents, and measuring instruments; and several more or less independent installations may be set up in the same area or “laboratory.” Each such participating installation should be considered as a collaborating laboratory so far as this procedure is concerned. Similarly, sets of test results obtained with different participants or under different conditions of calibration would in general constitute results from different collaborating laboratories even though they were obtained on the same sets of equipment.

This concept makes it possible to increase the available “laboratories” by using two analysts (but not more than two) in as many laboratories as needed to bring the total to the recommended minimum of ten. In such cases the two analysts must evaluate the method independently in the fullest sense of the word, interpreted as using different samples, different reagents, different apparatus where possible, and performing the work on different calendar days. (In the design in Section 16, laboratories using two analysts are designated as A-1, A-2, B-1, B-2, etc.) The most desirable laboratories and analysts are those having previous experience with the proposed method or with similar methods. It is essential that enough experience be acquired to establish confidence in the performance of a laboratory before starting the interlaboratory test series. Such preliminary work must be done with samples other than those to be used in the formal interlaboratory test program.

10.1.4 *Apparatus*—The effect of duplicate setups is not often a critical variable in chemical analysis. In instrumental methods, however, apparatus can become an important factor because the various laboratories may be using different makes or types of equipment, for example, the various colorimeters and spectrophotometers used in photometric methods. In such cases, the effect of apparatus becomes confounded with between-laboratory variability, and special care must be used to avoid misinterpreting the results. Of course, if enough laboratories have instruments of each type, “apparatus” can be made a planned variable in the study.

<sup>6</sup> Task group chairmen are referred specifically to Youden, W. J. “Experimental Design and ASTM Committees,” *Materials Research & Standards*, MTRSA Vol 1, No. 11, November 1961, p. 862.

<sup>7</sup> Practice E 691 insists on a minimum of six laboratories, but would prefer more than ten.



**10.1.5 Analysts**—The use of a single analyst in each “laboratory” (as defined in 10.1.3) is adequate to provide the information needed for calculating the within-laboratory, between-days variability and reproducibility of the method as defined in this practice. It is essential that all analysts complete the entire interlaboratory test program. With regard to analyst qualifications, an analyst who is proficient in the method should be selected.

**10.1.6 Days**—As defined in 6.1.17, the within-laboratory, between-days variability of the method shall be evaluated in terms of independent determinations by the same analyst. To achieve this, all scheduled determinations must be performed on each of two days (see Sections 16 and 25).

**NOTE 2**—As used in this practice, the term “days” represents replication of a set of determinations performed on any day other than that on which the first set was run. It may become a systematic variable to the extent that it is desirable that a given laboratory run the entire set of samples on one day and repeat the entire set on another. Although this may introduce a bias for that laboratory, there appears to be little chance that such a bias would be common to all laboratories. Where preliminary studies suggest that instability may result in an over-all systematic “days” effect, special planning will be required to take care of this problem.

**10.1.7 Runs**—The multiple determinations performed at the same time or within a very short time interval, on each day. In this practice, two runs (that is, duplicate determinations) are performed on each of two days.

## **11. Number of Determinations**

**11.1** Each analyst is required to perform duplicate determinations on each sample on each of two days. If one determination of a paired set is accidentally ruined, another pair must be run. An odd or unusual value does not constitute a “ruined” determination. In such cases, an additional set of duplicate determinations should be run and all values reported, with an assignable cause if at all possible.

## **12. Samples**

**12.1** One person should be made responsible for accumulating, subdividing, and distributing the materials to be used in the test program. Extra samples should be held in reserve to permit necessary replacement of any that may be lost or damaged in transit. Proper techniques in packaging and sampling should be followed, particularly with corrosive or otherwise hazardous materials. It is recommended that: all liquid samples be tested for closure leakage by laying the bottles on their side for 24 h prior to packaging, sample bottles be packed in boxes with strict attention to right side up labels, sample bottles be enclosed in plastic bags with plastic ties, packing of severely corrosive liquids be supervised by a technically trained person, and that strict attention be paid to DoT regulations. If a collaborating laboratory should receive a sample which shows evidence of leakage, or which is suspect for any other reason, the recipient should not use it but should immediately request a replacement.

**12.2** The most important requirement is that the sampling units to be distributed to the participating laboratories be random selections from a reasonably homogeneous quantity (sample) of material. Single-phase liquids usually present no problem unless they are hygroscopic or unstable. Solid mix-

tures, in which the components vary in particle size, should be ground, sieved, and recombined to give a homogeneous product, and then checked (microscopically, or by any other available means) to confirm its homogeneity.

**12.3** In the case of stable, homogeneous materials, one sampling unit can be distributed to each collaborating laboratory. If the material is hygroscopic, or otherwise unstable, multiple sampling units should be provided for each day’s run by each analyst.

**12.4** Instability of any type may impose other restrictions on the execution of a planned program. It is the responsibility of the task group chairman to include in the plans for the interlaboratory study specific instructions on selecting, preparing, storing, and handling of the standard samples.

**12.5** The sampling units distributed for the formal interlaboratory test program should not be used for practice runs. Where “dry-runs” are performed to develop proficiency in an inexperienced analyst or laboratory, this must be done on samples other than these.

## **13. Scheduling and Timing**

**13.1** Interlaboratory studies fail occasionally because no timetable had been established to cover the program, particularly in cases where the materials have changed in storage, after opening the container, etc. The instructions to the collaborators should cover such points as the time between receipt of samples and their testing, time elapsing between start and finish of the program, the order of performing the tests, etc., with particular attention to randomizing as a means of avoiding systematic errors.

**NOTE 3**—A discussion of randomizing is beyond the scope of this practice. Refer to standard textbooks on statistics and specifically to the indicated references (9, 10).

## **14. Instructions and Preliminary Questionnaire**

**14.1** Having decided on the variables and levels for each, the task group chairman should distribute to all participants a complete description of the planned collaborative study, emphasizing any special conditions or precautions to be observed. A detailed procedure and description of equipment, prepared in ASTM format, must be included. A questionnaire similar to the one in Table 1 will aid materially in the successful execution of the interlaboratory study.

## **15. Report Form**

**15.1** A form for reporting the essential data should be prepared and distributed (in duplicate) to all collaborators, who should be instructed on the number of decimal places to be used. It is recommended that interlaboratory studies be reported to one decimal place beyond that called for in the “Report” instructions of the method under study. Any subsequent rounding should be done by the task group chairman or the data analyst.

## **16. Design for an Interlaboratory Test Program**

**16.1** The plan given in Table 2 should cover most cases where laboratories and levels (or materials) are the principal variables. It calls for each analyst to perform two determinations in parallel on each of two days, at each level.



**TABLE 1 Questionnaire on Interlaboratory Study**

Title of Method (attached):

1. Our laboratory wishes to participate in the cooperative testing of this method for precision data.

YES... NO...

2. As a participant, we understand that:

- (a) All essential apparatus, chemicals, and other requirements specified in the method must be available in our laboratory when the program begins,
- (b) Specified "timing" requirements (such as starting date, order of testing specimens, and finishing date) of the program must be rigidly met,
- (c) The method must be strictly adhered to,
- (d) Samples must be handled in accordance with instruction, and
- (e) A qualified analyst must perform the tests.

Having studied the method and having made a fair appraisal of our capabilities and facilities, we feel that we will be adequately prepared for cooperative testing of this method.

3. We can supply \_\_\_ qualified analysts.

YES... NO...

4. Comments:

\_\_\_\_\_  
Signature  
\_\_\_\_\_  
Company

**TABLE 2 Single Method, Single Analyst, Ten Laboratories, *N* Levels or Materials  
Level or Material I**

| Laboratory                                                       |       | A   | B   | C   | D   | E   | F   | G   | H   | I   | J   |
|------------------------------------------------------------------|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| or Laboratory                                                    |       | A-1 | A-2 | B-1 | B-2 | C   | D   | E   | F   | G   | H   |
| Day 1                                                            | Run a | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
|                                                                  | Run b | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| Day 2                                                            | Run a | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
|                                                                  | Run b | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| Level or Material II                                             |       |     |     |     |     |     |     |     |     |     |     |
| Laboratory                                                       |       | A   | B   | C   | D   | E   | F   | G   | H   | I   | J   |
| or Laboratory                                                    |       | A-1 | A-2 | B-1 | B-2 | C   | D   | E   | F   | G   | H   |
| Day 1                                                            | Run a | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
|                                                                  | Run b | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| Day 2                                                            | Run a | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
|                                                                  | Run b | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| etc. to<br>Level or Material <i>N</i> ( <i>N</i> = 3 or Greater) |       |     |     |     |     |     |     |     |     |     |     |
| Laboratory                                                       |       | A   | B   | C   | D   | E   | F   | G   | H   | I   | J   |
| or Laboratory                                                    |       | A-1 | A-2 | B-1 | B-2 | C   | D   | E   | F   | G   | H   |
| Day 1                                                            | Run a | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
|                                                                  | Run b | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| Day 2                                                            | Run a | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
|                                                                  | Run b | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

additional variables must be included, the proposed program should be referred to a statistician, the Subcommittee on Precision and Bias, or to Committee E-11 on Quality and Statistics for a specific recommendation.

## PART D—TESTING FOR OUTLYING OBSERVATIONS

### 17. Scope

17.1 This part covers some elementary recommendations for dealing with outlying observations and rejection of data. Lacking a universally accepted practice for the rigid application of available statistical tests, considerable technical and common sense judgment must be exercised in using them. Accordingly, the following procedures are offered only as guides for the data analyst and all decisions to exclude or to include any suspect data shall be subject to the approval of the task group concerned. Rejection of data as outliers should be done only after attempts have been made to ascertain why the suspect values differ from other values; for example, a calculation error, transposition of digits, misunderstanding of, or failure to follow the test method provisions, etc.

NOTE 4—The test for outlying observations should be applied only once to a set of interlaboratory test data. Although two or more values can

be rejected simultaneously, in no case should the remaining data again be tested for outliers.

### 18. Principle of Method

18.1 The tests for outliers among the "multiple runs" and "different days" data are based on control chart limits for the range, as described in the *ASTM Manual on Presentation of Data and Control Chart Analysis*, MNL 7, (14).

18.2 The test for outlying observations among laboratory averages is that described in Practice E 178.

18.3 The choice of significance levels for each of the three tests is based on practical experience gained from a number of interlaboratory studies involving chemical or physical properties.

NOTE 5—In choosing significance levels, there are two alternatives: (1) use of a low-significance level, accepting the divergent data, inflating variances, and perhaps failing to find significant differences, or (2) use of a higher significance level, rejecting the divergent data, deflating variances, and perhaps finding significance where none exists. In the case of multiple runs in an interlaboratory test program, the choice of the 0.001 level is based on the premise that only a high degree of divergence should justify rejection of data from a laboratory for this reason. The 0.01 level for days also reflects this premise. The 0.05 level for laboratories is frequently used and is chosen here because an outlying laboratory average, even at this significance level, may have a pronounced effect on