Designation: E 1169 – 89 (Reapproved 1996)

An American National Standard

AMERICAN SOCIETY FOR TESTING AND MATERIALS
100 Barr Harbor Dr., West Conshohocken, PA 19428
Reprinted from the Annual Book of ASTM Standards. Copyright ASTM

Standard Guide for Conducting Ruggedness Tests[1]

This standard is issued under the fixed designation E 1169; 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 In studying a test method, it is necessary to consider the effect of environmental factors on the results obtained using the test method. If this effect is not considered, the results from the original developmental work on the test method may not be as accurate as expected.The purpose of a ruggedness test is to find the variables (experimental factors) that strongly influence the measurements provided by the test method, and to determine how closely these variables need to be controlled. Ruggedness tests do not determine the optimum conditions for the test method.

1.2 The experimental designs most often used in ruggedness testing are the so called "Plackett-Burman" designs **(1)**.[2] Other experimental designs also can be used. This guide, however, will restrict itself to Plackett-Burman designs with two levels per variable because these designs are particularly easy to use and are efficient in developing the information needed for improving test methods. The designs require the simultaneous change of the levels of all of the variables, and allow the determination of the separated effects of each of the variables on the measured results. In ruggedness tests the two levels for each variable are set so as not to be greatly different. For such situations, the calculated effect for any given variable is generally not greatly affected by changes in the level of any of the other variables. A detailed example involving glass electrode measurements of the pH of dilute acid solutions is used to illustrate ruggedness test procedures. A method is presented for evaluating the experimental uncertainties.

1.3 The information in this guide is arranged as follows:

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. Summary of Guide

2.1 A ruggedness test is conducted by making systematic changes in the variables associated with the test method and observing the size of the associated changes in the test method results. Generally, the designs (systematic plans of experimentation) associated with ruggedness tests are taken from the field of statistics.

3. Significance and Use

3.1 The ruggedness test of a test method should precede an interlaboratory study. The interlaboratory (round robin) study should be the final proof test for determining the precision of the test method. If a ruggedness test has not been run to determine, and subsequently to restrict the allowable ranges of the critical variables in the test method, then the precision from the round robin may be poor. It may not be known what went wrong, or how to fix the test method. The ruggedness test, by studying the influence of the test method variables and by indicating the need for selective tightening of test method specifications, helps avoid such situations. The use of ruggedness tests encourages the orderly development of a test method.

3.2 Ruggedness testing should be done within a single laboratory so the effects of the variables are easier to see. Only the effects of changes in the test method variables from high levels to low levels need to be determined. Numerous variables such as temperature, pressure, relative humidity, etc., may need to be studied. The influences of these changes are best studied under the short-term, high-precision conditions found within a single laboratory.

[1] This guide is under the jurisdiction of ASTM Committee E-11 on Statistical Methods and is the direct responsibility of Subcommittee E11.20 on Test Method Evaluation and Quality Control.

Current edition approved Nov. 20, 1989. Published January 1990. Originally published as E 1168 – 87. Last previous edition E 1168 – 87.

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

† Editorially corrected.

4. Plackett-Burman Designs Applied to Ruggedness Tests

4.1 A series of Plackett-Burman (P-B) designs are available for use with ruggedness tests for determining the effects of the test method variables (see 4.3 and Appendix X1). The effect for each variable is calculated on the basis that a given change in a variable from a high to a low level results in a fixed change in the test result. It is common in ruggedness testing to assume that the observed effect of the simultaneous change of a number of variables can be described as the simple addition of the fixed effects for each variable. It is also assumed that the effect for each variable is independent of the effects of other variables, that is, there are no coupled influences. The effects that are calculated on the basis of this assumption are called "main effects." If a considerable lack of independence among the effects of the variables is observed, the observer is than forced to recognize additional factors, which are called "interactions." The ruggedness test procedures for dealing with interactions are more complex, and are given in Refs (2) and (3). These more involved procedures, however, require additional measurements to develop information about the interactions. This guide is written only for evaluating main effects.

4.2 P-B designs require that N must be an integer multiple of four, for example, 4, 8, 12, 16, etc. P-B designs for N measurements per replicate can be used to estimate up to N-1 main effects. The calculated main effects, however, will be confounded (contaminated) with the interactions. If the interactions are relatively small, then the user may be satisfied in making only N ruggedness test measurements and obtaining slightly contaminated estimates for the N-1 main effects.

4.3 A P-B design for seven factors (A through G) and eight measurements is given in Fig. 1. This design is suitable for use whenever an independent estimate of measurement variability is available. Note that each column of the design contains an equal number of plus (+) and minus (−) factor settings. A (+) for a given factor indicates that the measurement is made with that factor set at the high level, and a (−) indicates the factor is to be at the low level. All seven factors are set for each measurement (test result). The eight measurements should be made in a random order. Typical test results are shown at the far right of the design in Fig. 1. If slightly less than seven factors are being investigated, simply drop the "excess" columns from the design. For such situations, the experimenter should consult a statistician to evaluate the measurement variability. In this regard, Ref (1) (pp. 310 to 320) may be of interest. The experimenter can, however, still use the techniques described in Sections 6, 7, and 8 of this guide.

	Factor							
	A	*B*	*C*	*D*	*E*	*F*	*G*	Test Result
1	+	+	+	−	+	−	−	1.1
2	−	+	+	+	−	+	−	6.3
3	−	−	+	+	+	−	+	1.2
4	+	−	−	+	+	+	−	0.8
5	−	+	−	−	+	+	+	6.0
6	+	−	+	−	−	+	+	0.9
7	+	+	−	+	−	−	+	1.1
8	−	−	−	−	−	−	−	1.4

FIG. 1 A Plackett-Burman Design for $N = 8$

4.4 A P-B design is constructed such that the four A(+) and the four A(−) terms will *each* be associated with an equal number of B(+) and B(−) terms. The A effect is orthogonal to the B effect, that is, it is not affected by the B effect. In the P-B design, all main effects (columns) are orthogonal to all other main effects (columns). This orthogonality of the main effects, and the acceptance of possible contamination of estimates for the main effects (by the interactions) are the major characteristics of most ruggedness tests. For many practical problems these characteristics are acceptable.

5. P-B Design Calculations

5.1 The effect of any factor, such as A, is calculated as the average of the measurements made at the high level minus the average of the measurements made at the low level, for example:

$$\text{Effect } A = \frac{\Sigma A(+)}{N/2} - \frac{\Sigma A(-)}{N/2} = (2/N)[\Sigma A(+) - \Sigma A(-)] \quad (1)$$

$$\begin{aligned}\text{Effect } A &= (2/8)\,[(1.1 + 0.8 + 0.9 + 1.1) \\ &\quad - (6.3 + 1.2 + 6.0 + 1.4)] \\ &= -2.75.\end{aligned}$$

5.2 For the P-B design, the standard deviation for an effect, such as A, is easily derived by using Eq 1 along with the standard deviation of a single measurement, σ.

$$\begin{aligned}\sigma_{\text{effect}}\, A &= \sqrt{(2/N)^2 \text{ variance } [\Sigma A(+) - \Sigma A(-)]} \\ &= \sqrt{(2/N)^2\, N\sigma^2} \quad (2) \\ \sigma_{\text{effect}}\, A &= 2\sigma/\sqrt{N}\end{aligned}$$

The same equations for the P-B design apply when the standard deviation σ is replaced by its sample estimate, s, as follows:

$$s_{\text{effect } A} = 2s/\sqrt{N} \quad (3)$$

Sections 7 and 9 present two methods for determining a sample estimate of the standard deviation of a single measurement, s.

6. P-B Design Considerations

6.1 Eq 3 shows that the standard deviation of an effect is inversely proportional to $\sqrt{N}$, the number of measurements made. The user may therefore be tempted to use large P-B designs. Practical experience, however, favors moderate size designs. Overly large designs require the correct setting of too many factors, and this increases the chance for blunders. In addition, large designs require more time to complete and other factors not being considered in the design can change and distort the results. The effects of incorrect factor settings and of shifting experimental conditions are propagated into all of the calculated results (see Eq 1). The ($N = 8$) P-B design in Fig. 1 is a suitable size for many experiments. If more factors need to be studied, a second ($N = 8$) P-B design may be used. This latter procedure may involve the repeated testing of some of the more important factors from the first design.

6.2 Ruggedness tests that have small or only moderate changes in the levels of the factors tend to have interactions that are relatively small, that is, the interactions tend to be unimportant relative to the main effects. For such situations, useful information may be obtained by investigating additional main effects rather than by investigating the numerous interactions.

6.3 In general, the size of all effects in a P-B design will increase with increased separation of the high and low settings of the factors. It seems prudent to use only moderate separations of the high and low settings so that the measured effects will be approximately additive and, at the same time, reasonably large relative to the measurement error. For the high and low settings of the factors, it is suggested that the extreme limits that may be expected to be observed between different qualified laboratories be used.

7. Interpretation of Results

7.1 Since the main effects are expressed in the units of the measurement, direct judgment can be made as to whether or not the change associated with the shift of the factor from a high level to a low level is too large. Other, more quantitative methods of judgment that analyze the variance of the measurements are given in 7.2, 7.3, and 7.4. These quantitative methods still only give tentative answers and follow-up or confirmatory experiments are frequently needed.

7.2 If *m* auxiliary measurements, all made under the same conditions as each other are available from other experimentation, the within-laboratory measurement variability, *s*, can be calculated. A *t*-test (with *m*-1 dF) can be used to judge if a main effect, such as *A*, is statistically significant relative to the measurement variability, for example:

$$t_{m-1} = \frac{\text{effect } A}{s_{\text{effect}A}} \quad (4)$$

Note that the *m* from the auxiliary measurements will not generally be the same as the *N* of the ruggedness test. Using Eq 3, the *t*-test can be calculated as follows:

$$t_{m-1} = \frac{\text{effect } A}{2s/\sqrt{N}} \quad (5)$$

A proper estimate for the *s*-value in Eq 5 should include all of the uncertainties of a single ruggedness test measurement. It is therefore desirable that the auxiliary measurements be made as independently as possible with experimental conditions being reset for each measurement.

7.3 Tighten the test method specification if the calculated *t*-value from Eq 5 is statistically significant, and if the size of the effect is of practical importance. This change should help reduce the interlaboratory variability.

7.4 The complete P-B-experiment can be replicated to obtain better estimates of the effects of the factors and to get a current estimate of the measurement variability, *s*. In estimating the measurement variability, it is necessary to guard against the occurrence of a possible measurement shift between the running of the two designs. This can be handled mathematically (see Section 8).

8. Example

8.1 This ruggedness testing example deals with factors that may influence the determination of the pH in dilute acid solutions when measurements are made by use of a glass electrode. The measurement procedures used with the glass electrode have been described in Ref 4. The seven factor ($N = 8$) P-B design that was used is given in Fig. 2. This convenient design was first suggested by F. Yates **(5)** and was frequently used by W. J. Youden **(6)** who did much of the pioneering work in ruggedness testing. For those experienced with the use of fractional factorial designs, it is a 2^{7-4} design. It has been shown **(2)**, by a rearrangement of the rows and columns, that this design is equivalent to the previously listed P-B design.

8.2 The seven factors that were studied are listed in 8.2.1-8.2.7. The first listed level for each factor has been arbitrarily assigned the positive sign:

8.2.1 *Factor A*—Temperature, 25 or 30°C,

8.2.2 *Factor B*—Stirring during the pH measurement: yes or no (denoted as Y or N in Table 1),

8.2.3 *Factor C*—Dilution (0.5 mL distilled H_2O/20 mL of solution), yes or no,

8.2.4 *Factor D*—Depth of electrode immersion, 1 or 3 cm below liquid surface,

8.2.5 *Factor E*—Addition of sodium nitrate ($NaNO_3$ = 0.67 meq/20 mL solution), yes or no,

8.2.6 *Factor F*—Addition of potassium chloride (KCl = 1.34 meq/20 mL of solution), yes or no, and

8.2.7 *Factor G*—Electrode equilibration time before measuring the pH, 10 or 5 min.

8.3 The seven factors are only a partial list of factors that may change the observed value of the pH. Obviously, all other factors that are not listed above need to be kept constant. The particular, constant levels of these other factors will result in some specific offset in the pH measurements. In the ruggedness test, however, this fixed offset need not be of concern since measurement changes (the effects) that occur when the seven factors (8.2.1-8.2.7) are changed is the primary interest.

8.4 Results from a ruggedness test with a hydrochloric acid (HCl) solution are given in Table 1. The complete experiment was replicated on a second day. A different random order of measurement was used for each day. The two sets of measurement results are given at the far right of Table 1.

8.4.1 For the first set of measurements in Table 1, the effect of factor *A* is calculated from Eq 1 as the difference of the average value when 25°C is used and the average value when 30°C is used, for example (2999 + 3055 + 3049 + 2949)/4 − (2904 + 3015 + 3006 + 2964)/4 = 3013 − 2972 = + 41. The averages and differences of the averages (the effects) are given for Factors A-G (8.2.1-8.2.7) in the third and fourth columns of Table 2. Similar calculations for the second set of measurements are given in the fifth and sixth columns of the table. A short-cut method for doing these calculations is given in X1.3.

	Factor						
	A	*B*	*C*	*D*	*E*	*F*	*G*
1	−	−	−	−	−	−	−
2	−	−	+	−	+	+	+
3	−	+	−	+	−	+	+
4	−	+	+	+	+	−	−
5	+	−	−	+	+	−	+
6	+	−	+	+	−	+	−
7	+	+	−	−	+	+	−
8	+	+	+	−	−	−	+

FIG. 2 Alternate Form of Plackett-Burman Design for *N* = 8