Designation: E1169 − 17$^{\varepsilon 1}$

An American National Standard

Standard Practice for Conducting Ruggedness Tests[1]

This standard is issued under the fixed designation E1169; 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.

ε¹ NOTE—Editorial corrections were made throughout in April 2018.

1. Scope

1.1 This practice covers conducting ruggedness tests. The purpose of a ruggedness test is to identify those factors that strongly influence the measurements provided by a specific test method and to estimate how closely those factors need to be controlled.

1.2 This practice restricts itself to designs with two levels per factor. The designs require the simultaneous change of the levels of all of the factors, thus permitting the determination of the effects of each of the factors on the measured results.

1.3 The system of units for this practice is not specified. Dimensional quantities in the practice are presented only as illustrations of calculation methods. The examples are not binding on products or test methods treated.

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, health, and environmental practices and determine the applicability of regulatory limitations prior to use.*

1.5 *This international standard was developed in accordance with internationally recognized principles on standardization established in the Decision on Principles for the Development of International Standards, Guides and Recommendations issued by the World Trade Organization Technical Barriers to Trade (TBT) Committee.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

E456 Terminology Relating to Quality and Statistics
E1325 Terminology Relating to Design of Experiments
E1488 Guide for Statistical Procedures to Use in Developing and Applying Test Methods
E2282 Guide for Defining the Test Result of a Test Method
F2082 Test Method for Determination of Transformation Temperature of Nickel-Titanium Shape Memory Alloys by Bend and Free Recovery

3. Terminology

3.1 *Definitions*—The terminology defined in Terminology E456 applies to this practice unless modified herein.

3.1.1 *fractional factorial design, n*—a factorial experiment in which only an adequately chosen fraction of the treatments required for the complete factorial experiment is selected to be run. **E1325**

3.1.2 *level (of a factor), n*—a given value, a specification of procedure or a specific setting of a factor. **E1325**

3.1.3 *Plackett-Burman designs, n*—a set of screening designs using orthogonal arrays that permit evaluation of the linear effects of up to $n = t - 1$ factors in a study of t treatment combinations. **E1325**

3.1.4 *ruggedness, n*—insensitivity of a test method to departures from specified test or environmental conditions.

3.1.4.1 *Discussion*—An evaluation of the "ruggedness" of a test method or an empirical model derived from an experiment is useful in determining whether the results or decisions will be relatively invariant over some range of environmental variability under which the test method or the model is likely to be applied.

3.1.5 *ruggedness test, n*—a planned experiment in which environmental factors or test conditions are deliberately varied in order to evaluate the effects of such variation.

3.1.5.1 *Discussion*—Since there usually are many environmental factors that might be considered in a ruggedness test, it is customary to use a "screening" type of experiment design which concentrates on examining many first order effects and generally assumes that second order effects such as interactions and curvature are relatively negligible. Often in evaluating the ruggedness of a test method, if there is an indication that the results of a test method are highly dependent on the levels of the environmental factors, there is a sufficient indication that certain levels of environmental factors must be included in the specifications for the test method, or even that the test method itself will need further revision.

---

[1] This practice is under the jurisdiction of ASTM Committee E11 on Quality and Statistics and is the direct responsibility of Subcommittee E11.20 on Test Method Evaluation and Quality Control.

Current edition approved Oct. 1, 2017. Published January 2018. Originally approved in 1987. Last previous edition approved in 2014 as E1169 – 14. DOI: 10.1520/E1169-17E01.

[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.

3.1.6 *screening design, n*—a balanced design, requiring relatively minimal amount of experimentation, to evaluate the lower order effects of a relatively large number of factors in terms of contributions to variability or in terms of estimates of parameters for a model. **E1325**

3.1.7 *test result, n*—the value of a characteristic obtained by carrying out a specified test method. **E2282**

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *factor, n*—test variable that may affect either the result obtained from the use of a test method or the variability of that result.

3.2.1.1 *Discussion*—For experimental purposes, factors must be temporarily controllable.

3.2.2 *foldover, n*—test runs, added to a two-level fractional factorial experiment, generated by duplicating the original design by switching levels of one or more factors in all runs.

3.2.2.1 *Discussion*—The most useful type of foldover is with signs of all factors switched. The foldover runs are combined with the initial test results. The combination allows main effects to be separated from interactions of other factors that are aliased in the original design.

4. Summary of Practice

4.1 Conducting a ruggedness test requires making systematic changes in the variables, called factors, and then observing the subsequent effect of those changes upon the test result of each run. Factors are features of the test method or of the laboratory environment that are known to vary across laboratories and are subject to control by the test method.

4.2 The factors chosen for ruggedness testing are those believed to have the potential to affect the results. However, since no limits may be provided in the standard for these factors, ruggedness testing is intended to evaluate this potential.

4.3 This practice recommends statistically designed experiments involving two levels of multiple factors. The steps to be conducted include:

4.3.1 Identification of relevant factors;

4.3.2 Selection of appropriate levels (two for each factor) to be used in experiment runs;

4.3.3 Display of treatment combinations in cyclic shifted order (see Annex A1 for templates), which assigns factors and levels to runs;

4.3.4 Execution of runs arranged in a random order;

4.3.5 Statistical analysis to determine the effect of factors on the test method results; and

4.3.6 Possible revision of the test method as needed.

5. Significance and Use

5.1 A ruggedness test is a special application of a statistically designed experiment. It is generally carried out when it is desirable to examine a large number of possible factors to determine which of these factors might have the greatest effect on the outcome of a test method. Statistical design enables more efficient and cost effective determination of the factor effects than would be achieved if separate experiments were carried out for each factor. The proposed designs are easy to use in developing the information needed for evaluating quantitative test methods.

5.2 In ruggedness testing, the two levels for each factor are chosen to use moderate separations between the high and low settings. In general, the size of effects, and the likelihood of interactions between the factors, will increase with increased separation between the high and low settings of the factors.

5.3 Ruggedness testing is usually done within a single laboratory on uniform material, so the effects of changing only the factors are measured. The results may then be used to assist in determining the degree of control required of factors described in the test method.

5.4 Ruggedness testing is part of the validation phase of developing a standard test method as described in Guide E1488. It is preferred that a ruggedness test precedes an interlaboratory (round robin) study.

6. Ruggedness Test Design

6.1 A series of fractional factorial designs are recommended for use with ruggedness tests for determining the effects of the test method variables (see Annex A1). All designs considered here have just two levels for each factor. They are known as Plackett-Burman designs (**1**).[3]

6.1.1 Choose the level settings so that the measured effects will be reasonably large relative to measurement error. It is suggested that the high and low levels be set at the extreme limits that could be expected to exist between different qualifying laboratories.

6.2 Table 1 shows the recommended design for up to seven factors, each factor set at two levels. The level setting is indicated by either (–1) or (1) for low or high levels, respectively. For factors with non-ordered scales (categorical), the designation "low" or "high" is arbitrary.

6.3 The design provides equal numbers of low and high level runs for every factor. In other words, the designs are balanced. Also, for any factor, while it is at its high level, all other factors will be run at equal numbers of high and low levels; similarly, while it is at its low level, all other factors will be run at equal numbers of high and low levels. In the terminology used by statisticians, the design is orthogonal.

6.4 The difference between the average response of runs at the high level and the average response of runs at the low level of a factor is the "main effect" of that factor. When the effect of a factor is the same regardless of levels of other factors, then the main effect is the best estimate of the factor's effect.

6.5 If the effect of one factor depends on the level of another factor, then these two factors interact. The interaction of two factors can be thought of as the effect of a third factor for which the column of signs is obtained by multiplying the columns of signs for the two initial factors. For example, the eight signs for Column C of Table 1, multiplied by the corresponding eight signs in Column D, gives a column of signs for the interaction CD. The complication of the fractional factorial designs

---

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

TABLE 1 Recommended Design for Up to Seven Factors

NOTE 1—For four factors, use Columns A, B, C, and E; for five factors, use Columns A, B, C, D, and F; for six factors, use Columns A, B, C, D, F, and G.

PB Order	Run #	A	B	C	D	E	F	G	Test Result
1		1	1	1	–1	1	–1	–1	
2		–1	1	1	1	–1	1	–1	
3		–1	–1	1	1	1	–1	1	
4		1	–1	–1	1	1	1	–1	
5		–1	1	–1	–1	1	1	1	
6		1	–1	1	–1	–1	1	1	
7		1	1	–1	1	–1	–1	1	
8		–1	–1	–1	–1	–1	–1	–1	
Ave+									
Ave–									
Effect									

presented here is that main effects are confounded (aliased) with the two-factor interactions. Factors are aliased when their columns of signs are the negatives or positives of each other. For example, the column of signs for the interaction CD is identical to minus the column of signs for Column A.

6.6 To separate factor main effects from interactions, the design shall be increased with additional runs. A "foldover," as shown in Table 2, is recommended to separate the main effects from the aliased interactions. When the runs in Tables 1 and 2 are combined, all main factors will no longer be aliased with two-factor interactions.

6.7 Sensitivity of the experiment can be increased by the addition of a second block of runs that replicates the first (that is, runs with the same factor settings as the first block). Increasing the size of the experiment improves the precision of factor effects and facilitates the evaluation of statistical significance of the effects. However, the preference of this practice is to use a foldover rather than a repeat of the original design.

6.8 The sequence of runs in Tables 1 and 2 is not intended to be the actual sequence for carrying out the experiments. The order in which the runs of a ruggedness experiment are carried out should be randomized to reduce the probability of encountering any potential effects of unknown, time-related factors. Alternatively, optimum run orders to control the number of required factor changes and the effect of linear time trends have been derived (**2**). In some cases, it is not possible to change all factors in a completely random order. It is best if this limitation is understood before the start of the experiment. A statistician may be contacted for methods to deal with such situations.

7. Ruggedness Test Calculations

7.1 Estimate factor effects by calculating the difference between average responses at the high and the low levels. When the design is folded over, obtain the main effect of a factor by averaging effects from the design and its foldover. Estimate the corresponding confounded interactions by taking half the difference of the main effects.

7.2 A half-normal plot is used to identify potentially statistically significant effects.

7.2.1 Construct a half-normal plot by plotting the absolute values of effects on the *X*-axis, in order from smallest to largest, against the half-normal plotting values given in Annex A2 on the *Y*-axis. Effects for all columns in the design, including columns not used to assign levels to any real experiment factor, are plotted. The half-normal plotting values do not depend on data. They depend only on the half-normal distribution and the number of effects plotted.

7.2.2 A reference line in the half normal plot is provided with slope $1/s_{effect}$, if an estimate of precision is available. Potentially significant effects are those that fall farthest to the right of the line.

7.3 If an estimate of precision is available or can be derived from the experiment, statistical tests of factor effects can be determined using the Student's *t*-test. The *t*-test statistic for a factor is the effect divided by the standard error s_{effect}, which is the same for all factors with a balanced and orthogonal design. If the *t*-value is greater than the *t*-value corresponding to the 0.05 significance level, the factor is statistically significant at level 0.05.

TABLE 2 Foldover of Design Shown in Table 1

PB Order	Run #	A	B	C	D	E	F	G	Test Result
1		–1	–1	–1	1	–1	1	1	
2		1	–1	–1	–1	1	–1	1	
3		1	1	–1	–1	–1	1	–1	
4		–1	1	1	–1	–1	–1	1	
5		1	–1	1	1	–1	–1	–1	
6		–1	1	–1	1	1	–1	–1	
7		–1	–1	1	–1	1	1	–1	
8		1	1	1	1	1	1	1	
Ave+									
Ave–									
Effect									

7.3.1 If fewer factors are used with the design than the maximum number, then "effects" estimated for the unused columns differ from zero only as a result of experimental error (or interactions of other factors). The root mean square of unused effects is an estimate of the standard error of an effect having degrees of freedom equal to the number of unused effects averaged (**3**).

7.3.2 The design may be replicated; that is, a second block of runs using the same factor settings as the original design is run. Then an estimate of the standard error of an effect is:

$$s_{effect} = \sqrt{\frac{4s_{rep}^2}{N \times reps}} \quad (1)$$

with degrees of freedom of $(N - 1) \times (reps - 1)$,

where:

N = number of runs in the design,
$reps$ = number of replicates of the design, and
s_{rep} = the estimated standard deviation of the test results.

7.3.2.1 An example showing calculation of s_{rep} and s_{effect} is given in 8.2.

8. Example of a Replicated Ruggedness Experiment

8.1 An example of a seven-factor ruggedness experiment comes from a study done for Test Method F2082. This test method determines a transformation temperature for nickel-titanium shape memory alloys. The factors of interest are quench method, bath temperature at deformation, equilibrium time, bending strain, pin spacing, linear variable differential transducer (LVDT) probe weight, and heating rate. Table 3 provides the levels of factors chosen in this example.

8.2 After all tests are completed, the transformation temperature results are entered in Table 4 in the Rep 1 and Rep 2 Test Result columns.

8.2.1 Factor main effects are then calculated using the average values (Rep Ave) of each design point for the two replicates. At the bottom of each column are the averages of the replicate averages corresponding to the (1) and the averages of the replicate averages corresponding to the (–1) signs in that column. For instance, in Table 4, for factor A, the (Ave+) value is the average of measurements values corresponding to the (1 = water) signs in Column A: –27.29, –17.28, –31.70, and –15.45, which yield an average of –22.93. The (Ave–) value is the average of the measurement values corresponding to the (–1 = air) signs in Column A: –17.40, –27.76, –35.10, and –43.10, which average –30.84.

8.2.2 The effect row contains the difference [(Ave+) – (Ave–)] for that column. It may be interpreted as the result of changing the factor shown in that column from low to high level. For factor A, since the Ave+ is 7.91 more than the Ave–, the effect is 7.91.

8.2.3 Estimate the standard deviation of the test and the standard error of effects from the dispersion of differences between replicates. The first pair of replicate readings is –26.95 and –27.63 and the difference (Rep2 – Rep1) is –0.68. The remaining differences are: 0.74, 2.85, 1.15, –2.68, –2.55, 3.23, and –0.69. The standard deviation of the differences is 2.23.

8.2.4 The estimate of the standard deviation of the test results, s_r (see 7.3.2), is:

$$s_r = s_d / \sqrt{2} = 2.23/1.414 = 1.58 \quad (2)$$

for the example data. For this example $N = 8$, and Rep = 2 and

$$s_{effect} = \sqrt{\frac{4s_r^2}{8 \times 2}} = 1.58/2 = 0.79 \quad (3)$$

8.3 Statistical significance of the factor effects and half-normal values for the half-normal plot are shown in Table 5.

8.3.1 Dividing the effect by s_{effect} provides a Student's *t*-value, which has $(N - 1)(\text{reps} - 1)$ degrees of freedom, seven degrees of freedom for this experiment. For example, for Effect A, the *t*-value is 7.91 / 0.79 = 10.04. Based on the assumption that there is no effect, the probability of a *t* score as large as 10.04 is approximately 0 (*p*-value < 0.001).

8.3.2 The half-normal plot is shown in Fig. 1. A line for comparison to factor effects is plotted with slope determined by $1 / s_{effect}$. Potentially significant effects are those which fall farthest to the right of the line. The conclusion of this test is that four of the design factors (D, A, B, and F) have significant effects on the response, the largest being bending strain factor D. The *p*-values for these four factors are all smaller than 0.05.

8.3.3 For the method evaluated in this example, the experimenters performed more testing on the effect of bending strain and bath temperature. Test Method F2082 was then revised by

TABLE 3 Test Method F2082 Ruggedness Test Factors, Levels, and Description

Factor No.	Variable	Discussion	Units	F2082 Limits	Level 1 (–)	Level 2 (+)
A	quench method	method of cooling after heat treatment of test specimen			air cool	water
B	bath temperature at deformation	temperature at which strain is applied to the test specimen	°C	–40 maximum	–60	–40
C	equilibration time	time at which the test specimen and fixture rest in the liquid bath before application of strain	minutes	2 minimum	2	4
D	bending strain	strain applied to test specimen at the deformation temperature	%	2–4	2	4
E	pin spacing	distance between test specimen supports	% of mandrel diameter	80–95	80	95
F	LVDT probe weight	load that the displacement transducer places on the test specimen	grams	3 maximum	1	3
G	heating rate		°C/min	4 maximum	2	4

TABLE 4 Test Method F2082 Ruggedness Test Calculations

PB Specified Order Number	A	B	C	D	E	F	G	Rep 1 Test Result	Rep 2 Test Result	Rep Ave	Rep Difference
1	1	1	1	–1	1	–1	–1	–26.95	–27.63	–27.29	–0.68
2	–1	1	1	1	–1	1	–1	–17.77	–17.03	–17.40	0.74
3	–1	–1	1	1	1	–1	1	–29.18	–26.33	–27.76	2.85
4	1	–1	–1	1	1	1	–1	–17.85	–16.70	–17.28	1.15
5	–1	1	–1	–1	1	1	1	–33.76	–36.44	–35.10	–2.68
6	1	–1	1	–1	–1	1	1	–30.42	–32.97	–31.70	–2.55
7	1	1	–1	1	–1	–1	1	–17.06	–13.83	–15.44	3.23
8	–1	–1	–1	–1	–1	–1	–1	–42.75	–43.44	–43.10	–0.69
Ave+	–22.93	–23.81	–26.04	–19.47	–26.86	–25.37	–27.5				
Ave–	–30.84	–29.96	–27.73	–34.3	–26.91	–28.4	–26.27				
Main Effect	7.91	6.15	1.69	14.83	0.054	3.03	–1.23			Std error effect	0.79

TABLE 5 Statistical Significance of Effects for Test Method F2082 Ruggedness Test

Effect Order, *e*	Effect	Estimated Effect	Student's *t*	*p*-value^A	Half-Normal Plotting Values
7	D	14.83	18.82	<0.001^B	1.80
6	A	7.91	10.04	<0.001^B	1.24
5	B	6.15	7.80	<0.001^B	0.92
4	F	3.03	3.85	0.006^B	0.67
3	C	1.69	2.15	0.069	0.46
2	G	–1.23	–1.57	0.16	0.27
1	E	0.054	0.072	0.95	0.09

^{A} *p*-value is the two-sided tail probability of Student's *t* with seven degrees of freedom, which can be calculated in Microsoft Excel by function *t*dist(*t*,df,2).
^{B} The marked values are statistically significant at the 5 % level.

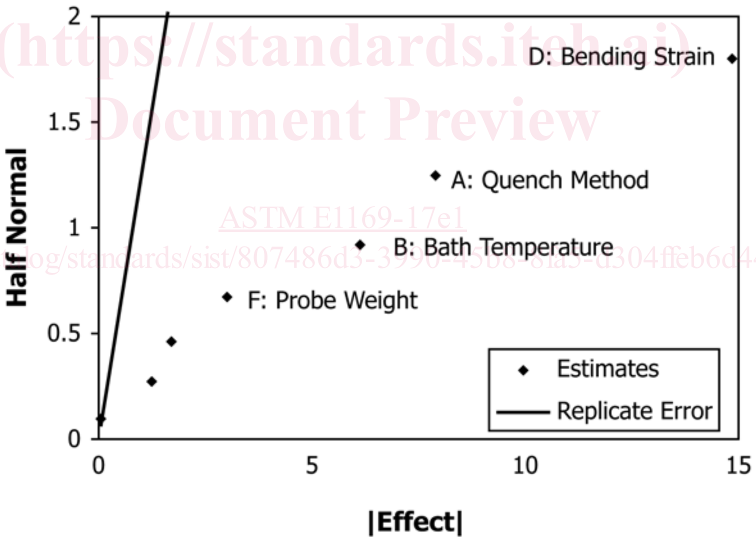

FIG. 1 Half-Normal Plot, Test Method F2082 Example

reducing the tolerance on these two parameters (bending strain was changed from 2–4 % to 2–2.5 % and bath temp changed to –55 max from –40 max). It was not practical to change the probe weight tolerance (a possibly significant factor), and quench method was related to sample preparation, not to the standard test method.

9. Example of a Ruggedness Experiment with Foldover

9.1 This example is part of a series of experiments that studied the effects of factors that influence determination of pH in dilute acid solutions (**4**, **5**). The factors and their levels are shown in Table 6. The data and calculated main effects, for the initial design and the foldover experiment subsequently performed, are shown in Tables 7 and 8. The results are recorded as 1000 × pH.

9.2 Based solely on the estimated effects for the first (Table 7) half of the experiment, factors B, D, E, and G appear to be significant. In Annex A3 and Table 9, it is shown that Interactions AF, CG, and DE are confounded with factor B. As a general rule, factors interact only when they have large main effects in their own right. Hence, AF and CG are unlikely to be

TABLE 6 Example: Factors That Influence Determination of pH in Dilute Acid Solutions

Factor No.	Variable	Units	Level 1 (–)	Level 2 (+)
A	Dilution with water	yes or no	No	yes
B	Addition of potassium chloride	yes or no	No	yes
C	Equilibration time	minutes	5	10
D	Depth of electrode immersion	cm	1	3
E	Addition of sodium nitrate	yes or no	No	yes
F	Stirring	yes or no	No	yes
G	Temperature	°C	2	4

TABLE 7 Results and Effects for Initial Design

	A	B	C	D	E	F	G	Test Result
1	1	1	1	–1	1	–1	–1	3015
2	–1	1	1	1	–1	1	–1	3006
3	–1	–1	1	1	1	–1	1	2999
4	1	–1	–1	1	1	1	–1	2964
5	–1	1	–1	–1	1	1	1	3049
6	1	–1	1	–1	–1	1	1	2949
7	1	1	–1	1	–1	–1	1	3055
8	–1	–1	–1	–1	–1	–1	–1	2904
Ave+	2995.8	3031.3	2992.3	3006.0	3006.8	2992.0	3013.0	
Ave–	2989.5	2954.0	2993.0	2979.3	2978.5	2993.3	2972.3	
Main Effect	6.3	77.3	–0.8	26.8	28.3	–1.3	40.8	

TABLE 8 Results and Effects for Foldover Factor—Settings Are at the Opposite Level to the First Set (Table 7)

	A	B	C	D	E	F	G	Test Result
1	–1	–1	–1	1	–1	1	1	2931
2	1	–1	–1	–1	1	–1	1	2978
3	1	1	–1	–1	–1	1	–1	2967
4	–1	1	1	–1	–1	–1	1	3030
5	1	–1	1	1	–1	–1	–1	2874
6	–1	1	–1	1	1	–1	–1	2979
7	–1	–1	1	–1	1	1	–1	2911
8	1	1	1	1	1	1	1	3040
Ave+	2962.8	2923.5	2963.8	2971.5	2950.5	2965.3	2932.8	
Ave–	2964.8	3004.0	2963.8	2956.0	2977.0	2962.3	2994.8	
Main Effect	2.0	80.5	0.0	–15.5	26.5	–3.0	62.0	

TABLE 9 Calculation of Estimated Effects Using Data from Tables 7 and 8 (Foldover of Table 7)

Factor	Table 4	Foldover	Average
A	6.3	2.0	4.1
B	77.3	80.5	78.9
C	–0.8	0.0	–0.4
D	26.8	–15.5	5.6
E	28.3	26.5	27.4
F	–1.3	–3.0	–2.1
G	40.8	62.0	51.4
			½ difference
A–I = –BF – CD – EG	6.3	2.0	–2.1
B–I = –AF – CG – DE	77.3	80.5	1.6
C–I = –AD – BG – EF	–0.8	0.0	0.38
D–I = –AC – BE – FG	26.8	–15.5	–21.1
E–I = –AG – BD – CF	28.3	26.5	–0.88
F–I = –AB – CE – DG	–1.3	–3.0	–0.88
G–I = –AE – BC – DF	40.8	62.0	10.6

important, but a DE interaction could be contributing to the estimated B effect. Similarly, AC, BE, and FG are confounded with D; a BE interaction could be contributing to the apparent D effect. Likewise, a BD interaction could be contributing to E. The foldover experiment is conducted to remove the ambiguity caused by this confounding. Results from the foldover experiment are shown in Table 8. (If fewer than seven factors are