Designation: E2263 – 04

StandardTest Method for Paired Preference Test[1]

This standard is issued under the fixed designation E2263; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This document covers a procedure for determining preference between two products using a two-alternative forced-choice, which may or may not include the option of choosing no preference.

1.2 A paired preference test determines whether there is a statistically significant preference between two products for a given population of respondents. The target population must be carefully considered.

1.3 This method establishes preference in a single evaluation context. Replicated tests will not be covered within the scope of this document.

1.4 Paired preference testing can address overall preference or preference for a specified sensory attribute.

1.5 The method does not directly determine the magnitude of preference.

1.6 This method does not address whether or not two samples are perceived as different. See Test Method E2164.

1.7 A paired preference test is a simple task for respondents, and can be used with populations that have minimal reading or comprehension skills, or both.

1.8 Preference is not an intrinsic attribute of the product, such as hue is, but is a subjective measure relating to respondents' affective or hedonic response. It differs from paired comparison testing which measures objective characteristics of the product. Preference results are always dependent on the population sampled.

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

2. Referenced Documents

2.1 *ASTM Standards:*[2]

E253 Terminology Relating to Sensory Evaluation of Materials and Products

E456 Terminology Relating to Quality and Statistics

E1858 Test Method for Determining Oxidation Induction Time of Hydrocarbons by Differential Scanning Calorimetry

E1871 Guide for Serving Protocol for Sensory Evaluation of Foods and Beverages

E2164 Test Method for Directional Difference Test

2.2 *ASTM Publication:*

Manual 26 Sensory Testing Methods, 2nd Edition[2]

2.3 *ISO Standard:*

ISO 5495 Sensory Analysis—Methodology—Paired Comparison[3]

3. Terminology

3.1 For definition of terms relating to sensory analysis, see Terminology E253, and for terms relating to statistics, see Terminology E456.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 α *(alpha) risk*—the probability of concluding that a preference exists when, in reality, one does not. (Also known as Type I Error or significance level.)

3.2.2 β *(beta) risk*—the probability of concluding that no preference exists when, in reality, one does. (Also known as Type II Error.)

3.2.3 *common responses*—for a one-sided test, the number of respondents selecting the product that is expected to be preferred. For a two-sided test, the largest number of respondents selecting either product.

3.2.4 *one-sided test*—a test in which the researcher has an *a priori* interest concerning the direction of the preference. In this case, the alternative hypothesis will express that a specific

[1] This test method is under the jurisdiction of ASTM Committee E18 on Sensory Evaluation and is the direct responsibility of Subcommittee E18.04 on Fundamentals of Sensory.

Current edition approved March 1, 2004. Published March 2004. DOI: 10.1520/E2263-04.

[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] Available from American National Standards Institute (ANSI), 25 W. 43rd St., 4th Floor, New York, NY 10036.

product is preferred over another product (that is, A > B or A < B), depending on the *a priori* belief.

3.2.5 *two-sided test*—a test in which the researcher does not have any *a priori* interests concerning direction of the preference. In this case, the alternative hypothesis is that the two products are not equally preferred (that is, A ≠ B).

3.2.6 P_{max}— a test sensitivity parameter established prior to testing and used along with the selected values of α and β to determine the number of respondents needed in a study. P_{max} is the proportion of common responses that the researcher wants the test to be able to detect with a probability of 1-β. For example, if a researcher wants to have a 90 % confidence level of detecting a 60:40 split in preference, then P_{max} = 60 % and β = 0.10.

3.2.7 *sensitivity*—a general term used to summarize the performance characteristics of the paired preference test. The sensitivity of the test is defined, in statistical terms, by the values selected for α, β, and P_{max}. Smaller values for α, β, and P_{max} indicate a more sensitive test.

3.2.8 p_c—the proportion of common responses which is calculated from the test data.

3.2.9 *product*—the material from which samples are selected.

3.2.10 *sample*—the unit of product prepared, presented, and evaluated in the test.

3.2.11 *respondent*—also known as assessor; a general term for any individual responding to stimuli in a sensory test.

4. Summary of Test Method

4.1 Clearly define the test objective in writing, specifying the type of audience or population you wish to recruit as respondents. (If objective involves substantiating an advertising claim, refer to Guide E1858.)

4.2 Choose the number of respondents (*N*) to be recruited based on the sensitivity level desired for the test (P_{max}, α, and β). The sensitivity of the test is, in part, a function of two competing risks—the risk of declaring a preference when there is none (that is, α-risk) and the risk of not declaring that a preference exists when there is a preference (that is, β-risk). Acceptable values of α and β vary depending on the test objective. The values should be agreed upon by all parties affected by the results of the test before the test is conducted.

4.3 In paired preference testing, an assessor receives a pair of coded samples that are identified with appropriate non-biasing codes. The assessor is asked to choose the sample that is preferred.

4.3.1 When using a forced choice procedure, a sample must be chosen even if the selection is based only on a random selection by the assessor.

4.3.2 If a choice is not forced, a "no preference" option should be included, and the data must be handled in a different way.

4.4 Results are tallied and significance determined by reference to a statistical table (or calculation).

4.5 Testing is generally conducted for one pair of samples to avoid bias from one set of samples to another.

5. Significance and Use

5.1 The paired preference test determines whether or not there is a preference for one product over another product among a specific target population. Knowledge of consumer segments, brand loyalties, the range of product offerings in the marketplace, and the decision risk must be understood when planning a paired preference test.

5.2 The paired preference method is commonly used in tests with one or more of the following objectives: (*1*) to establish superiority in preference versus the competition for advertising claims support; (*2*) to establish the preference of a new product for launch versus a competitor's product; (*3*) to establish the preference of a reformulated product in a product improvement or product modification project (for example, process change or ingredient change); and (*4*) to establish the preference of a cost improved product versus the current formulation in a cost savings project. Selected values of P_{max}, α, and β will change with all four types of test objectives. These should be selected prior to determination of *N*.

5.2.1 *Preference versus Competition for Advertising Claims, Launching a New Product versus Competition, or Product Improvement versus Current Product*—Select a P_{max} to represent what you expect a reasonable preference split to be. The main risk to avoid is to wrongly claim your product is preferred over the competitors. Thus, low values of α are selected, for example, 0.05, 0.01, or 0.001. The desired outcome of this test is to reject the null hypothesis. The alternative hypothesis is one sided: a new or improved product (A) is preferred over the competitors or the current formulation product (B). The test is one-sided. The value of β will be determined by the sample size chosen and the size of the preference in the consumer segment selected for the test. Selection of the appropriate number of respondents is determined by P_{max}, α, and β, as well as the market segment that must be included in the test (see Guide E1858).

5.2.2 *Cost Reduction or Reformulation of an Existing Product*—When parity preference is the desired test outcome, values of α are increased and values of β are decreased. For example, if a product is developed which represents a significant cost savings over the current formulation and there is concern over alienation of current users, α might be selected at 0.20 and β might be selected at 0.01. Parity testing can be either one-or two-sided depending on the action standards of the test. The test is one-sided if the action standard is that the product must be parity or better. The test is two-sided if the action standard is parity only. The number of respondents chosen must reflect the risk of replacing the current product with the cost-reduced product.

5.3 A test result of superiority or parity does not ensure that the test conclusion is correct. An incorrect test result can be obtained when the sample of respondents is selected in a way that does not reflect the true preference in the population of interest, or when the number of respondents is too small to correctly reflect the preference status of the two products among the target consumer group. Careful selection of P_{max}, α,

and β and an appropriate selection of respondents is needed to minimize the risk of drawing an incorrect conclusion in forced-choice paired preference testing.

6. Apparatus

6.1 Carry out the test under conditions that prevent contact between respondents until the evaluations have been completed.

6.2 Sample preparation and serving sizes should comply with Practice E1871, or see Refs (**4**) or (**5**).

7. Respondents

7.1 Choose the appropriate set of respondents on the basis of the test objective. Selecting the appropriate set of assessors for a preference test is critical since preference responses vary depending on the consumer group targeted. The most appropriate respondents to determine product preference are the current or potential consumers of the product category.

7.2 Respondents must be selected based upon the objective of the study and are dependent on the business implication. For a new product, the respondents should represent target consumers. For an existing product, respondents may include users of the product. If the business objective is to ensure that market share is not lost when making formula changes, respondents should include heavy category or product users.

8. Number of Respondents

8.1 Once the target population has been clearly defined, choose the number of respondents required for the test as follows: (*1*) first determine if the test is one-sided or two-sided, and (*2*) establish the sensitivity required by the test objectives by selecting values for the three test-sensitivity parameters, the α-risk, the β-risk, and the proportion of common responses, P_{max}, that would represent a meaningful departure from parity (50:50) preference as decided by the research team.

8.1.1 The test is one-sided if the researcher has an *a priori* interest in only one of the samples being preferred. For example, the test is one-sided if the researcher wants to determine if the product is preferred to the major competitor's product. The test is two-sided if the researcher has no *a priori* interest in a particular sample being preferred. For example, the test is two-sided if two prototype samples are being compared and the researcher wants to establish if one sample is preferred over the other sample. More respondents are needed for a two-sided test than for a one-sided test (see 5.2.1 and 5.2.2).

8.1.2 When the researcher wants to take only a small chance of concluding that a preference exists when it does not (for example, when testing to support a claim of superiority), the most commonly used values for α-risk and β-risk are $\alpha = 0.05$ and $\beta = 0.20$. These values can be adjusted on a case-by-case basis to reflect the sensitivity desired versus the number of respondents available. When testing for a preference with a limited number of respondents, hold the α-risk at a relatively small value and allow the β-risk to increase in order to control the risk of falsely concluding that a preference is present.

8.1.3 When the researcher wants to take only a small chance of missing a preference that exists (for example, when testing to support a claim of parity preference), the most commonly used values for α-risk and β-risk are $\alpha = 0.20$ and $\beta = 0.05$. These values can be adjusted on a case-by-case basis to reflect the sensitivity desired versus the number of respondents available. When testing for parity with a limited number of respondents, hold the β-risk at a relatively small value and allow the α-risk to increase in order to control the risk of missing a preference that truly exists.

8.1.4 For P_{max}, the proportion of common responses falls into three ranges: (*1*) $P_{max} < 55\ \%$ represents "small" values; (*2*) $55\ \% \leq P_{max} \leq 65\ \%$ represents "medium sized" values; and (*3*) $P_{max} > 65\ \%$ represents "large" values.

8.1.5 For example, if a researcher is planning a test to support a superior preference claim for a product over the major competitor's product, the researcher might choose the following values for the test-sensitivity parameters: $\alpha = 0.05$, $\beta = 0.20$, and $P_{max} = 60\ \%$. The test is one-sided because the researcher is only interested in the situation where their product is preferred.

8.2 Having defined the required sensitivity for the test using 8.1, use Table X1.1 to determine the number of respondents necessary for a one-sided test, or Table X1.2 to determine the number of respondents necessary for two-sided test. Select the section of the table corresponding to the selected P_{max} value and the column corresponding to the selected β value. The minimum required number of respondents is found in the row corresponding to the selected value of α. Alternatively, Table X1.1 can be used to develop a set of values for P_{max}, α, and β that provide acceptable sensitivity while maintaining the number of respondents within practical limits.

8.2.1 Using the values from the example in 8.1.5, the researcher would use the section of Table X1.1 corresponding to $P_{max} = 60\ \%$ and the column corresponding to $\beta = 0.20$. In the row corresponding to $\alpha = 0.05$, it is found that 158 respondents will be needed for the test.

8.3 Often in practice, the number of respondents is determined by project constraints (for example, duration of the experiment, number of available respondents, quantity of sample, budgetary restraints). The power of the test should then be computed. For this purpose, the following parameters need to be defined: α, observed P_{max}, and the number of respondents, n. The observed P_{max} corresponds to the observed proportion of common responses, n is determined by the test realization, and α should be fixed by the experimenter prior to the test conduct. With this information, an exact power computation can be achieved using appropriate software. However, an approximate value can be inferred by reverse lookup using Table X1.1 or Table X1.2, depending on whether the alternative is one- or two-sided. First, use the value of P_{max} closest to the observed one to select a group of rows, then select among these rows the one corresponding to the selected value of α. Finally, select the cell having the number of assessors closest to the actual number of assessors. The corresponding column heading will give a close estimate of the actual power of the test (1-β). Lower sample sizes will reduce the power of the test.

9. Procedure

9.1 Paired preference can be used in either CLT (Central Location Test) or IHUT (Inhome Use Test) designs. The following discussion focuses on CLT testing procedures, however, randomizations and data analyses would be similar for IHUT's.

9.2 Prepare serving order worksheet and ballot in advance of the test to ensure a balanced order of presentation of the two samples. Balance the serving sequences of the samples (AB and BA) across all respondents. Serving order worksheets should also include complete sample identification information either by product name or coded reference for double blind studies. See Appendix X1.

9.3 It is critical to the validity of the test that respondents cannot differentiate the samples based on the way they are presented. For example, in a test evaluating flavor differences, one should avoid any subtle differences in temperature or appearance caused by factors such as the time sequence of preparation. Code the vessels containing the samples in a uniform manner, using three digit numbers chosen at random for each test. Prepare samples out of sight and in an identical manner, that is, same apparatus, same vessels, same quantities of sample (see Practice E1871).

9.4 Present the pair of samples simultaneously if possible, following the same spatial arrangement for each assessor (on a line to be sampled always from left to right, or from front to back, and so forth). Respondents are typically allowed to try each sample more than once. If the conditions of the samples restrict retrying the samples (for example, if samples are bulky, leave an aftertaste, or show slight differences in appearance that cannot be masked), present the samples sequentially and do not allow repeated evaluations.

9.5 It is not recommended that more than one question be asked about the samples, because the selection the assessor has made on the initial question may bias the reply to subsequent questions. Responses to additional questions may be obtained through separate tests for acceptance, degree of difference, and so forth (see Manual 26). A section soliciting comments may be included following the initial preference question.

9.6 The paired preference test can be either forced-choice or have the option of no preference.

9.6.1 When using the paired preference test as a forced-choice procedure, respondents are not allowed the option of reporting "no preference." An assessor who has no preference for either of the samples should be instructed to randomly select one of the samples, and can indicate in the comments section that they had no preference.

10. Analysis and Interpretation of Results

10.1 The procedure used to analyze the results of a paired preference test depends on whether or not a "no preference" option is allowed.

10.1.1 If a forced choice procedure is used, analyze as detailed in 10.2.

10.1.2 If a "no preference" option is allowed, then there are various ways to handle the data depending on the test objectives. Typically the no preference data is split in some manner between "A" and "B." Regardless of how the no preference data is handled, it is always important to report the percentage of no preference responses and take those into account for your final action steps.

10.1.2.1 *For Ad Claim testing for superiority*, "no preference" responses go against your company's product superiority. Therefore, those responses are given to the competitive product.

10.1.2.2 *For Ad Claim testing for parity*, "no preference" responses are arguments against the competitive product superiority. Therefore, those responses are given to your company's product.

10.1.2.3 *For cost reduction or ingredient/supplier changes*, "no preference" responses are split between current and test product.

10.1.2.4 *For product improvement*, "no preference" responses are handled similarly to an ad claim superiority claim and given to the current (not "improved") product.

10.1.2.5 *For comparison of formulation options*, where there is no control or current product, no preference responses are split equally between the two products. It is important to also report the percentage of no preference responses and take those into account for your final action step.

10.2 *Analysis for Preference*—Different analyses are used depending on if the number of respondents is equal to or greater than planned or fewer than planned.

10.2.1 When the actual number of respondents is equal to or greater than planned, refer to Table X1.3 (one-tailed) or Table X1.4 (two-tailed) to analyze the data. If the number of common responses is equal to or greater than the number given in the table, conclude that there is a preference between the products. If the number of common responses is fewer than the number given in the table, conclude that there is no preference. The conclusions, "preference" or "no preference," are based on the predetermined α, β, and P_{max} levels.

10.2.2 When the actual number of respondents is fewer than planned, then the data analysis is the same as 10.2.1 above. Understand that the β-risk is now larger than the value chosen because a smaller number of respondents participated in the test.

10.3 *Analysis for Parity*—Different analyses are used depending on if the number of respondents is equal to or greater than planned or fewer than planned.

10.3.1 When the actual number of respondents is equal to or greater than planned, then the analysis is conducted as outlined in 10.2.1.

10.3.2 When the number of respondents is fewer than planned, then data analysis consists of calculating a confidence interval. A confidence interval is calculated because the α, β, and P_{max} levels are different in parity preference testing. The calculations are as follows, where c = the number of common responses, and n = the total number of respondents:

Proportion of common responses

$$\frac{P_c}{1} = c/n$$

$$S_c \left(\text{standard deviation of } P_c\right) = \sqrt{P_c(1 - P_c)/n}$$

$$\text{Confidence Limit} = P_c + z_\beta S_c$$

10.3.3 z_β is the critical value of the standard normal distribution. Values of z_β for some commonly used values of β-risk are:

β-risk	z_β
0.50	0.000
0.40	0.253
0.20	0.842
0.10	1.282
0.05	1.645
0.01	2.326
0.001	3.090

Given the values chosen for β and P_{max}, if the confidence limit is less than P_{max}, then conclude that there is parity (that is, no more than P_{max} of the population would have a preference at the β-level of significance). If the confidence limit is greater than P_{max}, then conclude that the products are not at parity. Understand that the α-risk is larger than the value chosen when a smaller number of respondents than planned participate in the test.

10.4 If desired, calculate a two-sided confidence interval on the proportion of common responses.

11. Report

11.1 Report the test objective, the results, and the conclusions. The following additional information is recommended:

11.1.1 The purpose of the test and the nature of the treatment studied;

11.1.2 Full identification of the samples: origin, method of preparation, quantity, shape, storage prior to testing, serving size, and temperature. (Sample information should communicate that all storage, handling, and preparation was done in such a way as to yield samples that differed only in the variable of interest, if at all.);

11.1.3 The number of respondents, recruitment criteria, the number of selections of each sample, and the result of the statistical analysis;

11.1.4 Respondents: age, gender, frequency of product usage: typical/usual product consumption in the category (for example, brand loyal or rotators);

11.1.5 Any information or instructions given to the assessor in connection with the test; including how the product was identified when presented;

11.1.6 The test environment: use of booths, simultaneous or sequential presentation, light conditions, whether the identity of samples was disclosed after the test and the manner in which this was done; and

The location and date of the test and name of the test administrator.

12. Precision and Bias

12.1 Because results of paired preference tests are a function of individual preferences, a general statement regarding the precision of results that is applicable to all populations of respondents cannot be made. Unless the demographics of the test population are matched to U.S. census, results are not projected to the total U.S. population. However, adherence to the recommendations stated in this standard should increase the reproducibility of results and minimize bias.

13. Keywords

13.1 paired preference; preference; sensory; test method

APPENDIXES

(Nonmandatory Information)

X1. EXAMPLE 1—PRODUCT IMPROVEMENT: FORCED CHOICE PROCEDURE

X1.1 Background

X1.1.1 A beverage manufacturer wants to determine if a new chocolate flavoring "A" is preferred over the current chocolate flavor "B" in a milk alternative beverage prior to fielding a more expensive in-home consumer test. It was decided to force a choice between the two flavors.

X1.2 Test Objective

X1.2.1 To determine if chocolate flavoring "A" is preferred over "B" in a milk alternative beverage. This is a one-tailed test.

X1.3 Number of Respondents

X1.3.1 To protect the product developer from falsely concluding that a preference exists, the sensory analyst proposes α = 0.05, and a P_{max} of 70 % with β = 0.01. The analyst enters Table X1.1 in the section corresponding to P_{max} = 0.70 and the column corresponding to β = 0.01. Then, reading from the row corresponding to α = 0.05, it is determined that a minimum of 94 respondents will be needed for the test. The sensory analyst recruits more than 94 respondents that have been identified as users of the product category to ensure that the minimum number of respondents are tested.

X1.4 Conducting the Test

X1.4.1 One hundred cups of "A" and 100 cups of "B" are coded with unique random three digit numbers. Each sequence, AB and BA, is presented 47 times so as to cover at least 94 respondents in a balanced random order, with extra servings available in case of accidental spills, etc. An example of the worksheet and scoresheet is shown in Figs. X1.1 and X1.2. Ninety-six respondents participated in the test.

X1.5 Analysis and Interpretation of Results

X1.5.1 Sixty-seven respondents selected the sample with chocolate flavor "A" as preferred, and 38 selected sample with flavor "B." In Table X1.3, the row corresponding to 96 respondents and the column corresponding to α = 0.05, the

TABLE X1.1 Number of Respondents Needed for a Paired Preference Test One-Sided Alternative[A]

		β							
α		0.50	0.40	0.30	0.20	0.10	0.05	0.01	0.001
0.50	p_{max} = 75 %	2	4	4	4	8	12	20	34
0.40		2	4	4	6	10	14	28	42
0.30		2	6	8	10	14	20	30	48
0.20		6	6	10	12	20	26	40	58
0.10		10	10	14	20	26	34	48	70
0.05		14	16	18	24	34	42	58	82
0.01		22	28	34	40	50	60	80	108
0.001		38	44	52	62	72	84	108	140
0.50	p_{max} = 70 %	4	4	4	8	12	18	32	60
0.40		4	4	6	8	14	26	42	70
0.30		6	8	10	14	22	28	50	78
0.20		6	10	12	20	30	40	60	94
0.10		14	20	22	28	40	54	80	114
0.05		18	24	30	38	54	68	94	132
0.01		36	42	52	64	80	96	130	174
0.001		62	72	82	96	118	136	176	228
0.50	p_{max} = 65 %	4	4	4	8	18	32	62	102
0.40		4	6	8	14	30	42	76	120
0.30		8	10	14	24	40	54	88	144
0.20		10	18	22	32	50	68	110	166
0.10		22	28	38	54	72	96	146	208
0.05		30	42	54	70	94	120	174	244
0.01		64	78	90	112	144	174	236	320
0.001		108	126	144	172	210	246	318	412
0.50	p_{max} = 60 %	4	4	8	18	42	68	134	238
0.40		6	10	24	36	60	94	172	282
0.30		12	22	30	50	84	120	206	328
0.20		22	32	50	78	112	158	254	384
0.10		46	66	86	116	168	214	322	472
0.05		72	94	120	158	214	268	392	554
0.01		142	168	208	252	326	392	536	726
0.001		242	282	328	386	480	556	732	944
0.50	p_{max} = 55 %	4	8	28	74	164	272	542	952
0.40		10	36	62	124	238	362	672	1124
0.30		30	72	118	200	334	480	810	1302
0.20		82	130	194	294	452	618	1006	1556
0.10		170	240	338	462	658	862	1310	1906
0.05		282	370	476	620	866	1092	1584	2238
0.01		550	666	820	1008	1302	1582	2170	2928
0.001		962	1126	1310	1552	1908	2248	2938	3812

[A] The values recorded in this table have been rounded to the nearest whole number evenly divisible by two to allow for equal presentation of both pair combinations (AB and BA).

sensory analyst finds that 57 common responses were needed in order to conclude that there is a preference.

X1.6 Report and Conclusions

X1.6.1 The sensory analyst reports that there was a significant preference for product with chocolate flavor "A," given the sensitivity chosen for the test (P_{max} = 70 %, α = 0.05, β = 0.01). The analyst concludes that product with chocolate flavor "A" would be the better candidate for in-home testing.