Designation: ~~E2263 – 04~~ E2263 – 12

Standard Test 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 either a two-alternative ~~forced-choice, which may or may not include~~ forced-choice task, or with the option of choosing no preference. Preference testing is a type of hedonic testing.

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~~ Refer to Test Method E2164~~.~~ for directional difference test.

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,~~ problems associated with its ~~use.~~ use, when testing includes hazardous materials, operations, or equipment. It is the responsibility of the user of this standard to establish appropriate safety and health practices and to 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

E1958 Guide for Sensory Claim Substantiation

E2164 Test Method for Directional Difference Test

~~2.2 *ASTM Publication:*~~

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

2.2 *ISO Standard:*

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

[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~~Oct. 15, 2012. Published ~~March 2004~~December 2012. Originally approved in 2004. Last previous edition approved in 2004 as E2263 – 04. DOI: ~~10.1520/E2263-04.~~10.1520/E2263-12.

[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.

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 *(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 *(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~~assumption concerning the direction of the preference. In this case, the alternative hypothesis will express that a specific product is preferred over another product (that ~~is,~~ is only, 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~~assumption 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 and 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-. 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 = 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 , , and P_{max}. Smaller values for~~of~~ , , 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. Trained panelists or experienced discrimination panelists do not serve as respondents in a paired preference 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~~E1958.)

4.2 Choose the number of respondents (*N*) to be recruited based on the sensitivity level desired for the test (P_{max}, , and). 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, -risk) and the risk of not declaring that a preference exists when there is a preference (that is, -risk). Acceptable values of and 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}, , and 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, or~~ Launching a New Product versus ~~Competition, or Product Improvement versus Current Product—~~Competition—*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 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~~A new or improved product (A) is preferred over the ~~competitors or the current formulation~~ competitor's product (B). The test is one-sided. The value of 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}, , and , as well as the market segment that must be included in the test ~~(see~~ (for issues specific to conducting a paired preference test for an advertising claim, refer to Guide ~~E1858~~E1958).

5.2.2 *Cost Reduction or Reformulation of an Existing Product*—When parity preference is the desired test outcome, values of are increased and values of 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, might be selected at 0.20 and 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}, , and 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~~Herz ~~(~~and Cupchik~~4)~~[4] or Todrank et al.~~(~~[5]~~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~~your 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,~~parameters: the -risk, the -risk, and the maximum allowable 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~~assumption 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 -risk and -risk are = 0.05 and = 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 -risk at a relatively small value and allow the -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 -risk and -risk are = 0.20 and = 0.05. These values

[4] Herz, R. S. and Cupchik, G. C., "An Experimental Characterization of Odor-evoked Memories in Humans," *Chemical Senses*, Vol 17, No. 5, 1992, pp. 519-528.

[5] Todrank, J., Wysocki, C. J., and Beauchamp, G. K., "The Effects of Adaptation on the Perception of Similar and Dissimilar Odors," *Chemical Senses*, Vol 16, No. 5, 1991, pp. 476-482.

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 -risk at a relatively small value and allow the -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 % P_{max} 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: = 0.05, = 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~~X2.1 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 value. The minimum required number of respondents is found in the row corresponding to the selected value of . Alternatively, Table X1.1 can be used to develop a set of values for P_{max}, , and that provide acceptable sensitivity while maintaining the number of respondents within practical limits.

8.2.1 Using the ~~values~~parameters: ~~*from* the example~~= 0.05, ~~*in* 8.1.5, the~~ = 0.20, and P_{max} = 60 %, the researcher would use the section of Table X1.1 corresponding to P_{max} = 60 % and the column corresponding to = 0.20. In the row corresponding to = 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,~~respondents available, quantity of sample, budgetary ~~restraints).~~constraints). The power of the test should then be computed. For this purpose, the following parameters need to be defined: , 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 should be fixed by the experimenter prior to the test c~~onduct. b~~eing conducted. With this information, an exact power computation can be achieved using appropriate software. However, an approximate value can already be inferred by reverse lookup using Table X1.1 or Table ~~X1.2~~X2.1, 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 . 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-). 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)~~(in-home 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,~~ manner: same apparatus, same vessels, same quantities of sample (see Practice E1871~~).~~, ASTM Serving Protocols).

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).~~ etc.). Respondents are typically allowed to ~~try~~evaluate each sample more than once. If the conditions of the samples restrict ~~retrying~~reevaluating 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~~ the preference question be asked about the samples, because the selection the ~~assessor~~respondent has made on the initial question may bias the ~~reply~~response to subsequent questions. Responses to additional questions may be obtained through separate tests for acceptance, degree of difference, ~~and so forth (see Manual 26).~~etc. See Manual 26[6]. A section soliciting open-ended comments may be included following the initial preference question.

9.6 The paired preference test can ~~be~~ either be 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~~A respondent 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.

[6] *MNL26-2ND Sensory Testing Methods: Second Edition*, Chambers, E. and Wolf, M.B., Eds., ASTM International, 1996.

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~~are handled, it is always important to report the percentage of no preference responses and take those into account for your final action steps. (Refer to Guide E1958 for decision rules regarding handling of no preference votes and specific claims.)

~~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~~whether the number of respondents is equal to or greater than planned or fewer than planned.

10.2.1 ~~When~~If the actual number of respondents is equal to or greater than planned, refer to Table ~~X1.3~~X1.2 ~~(one-tailed)~~(one-sided) or Table ~~X1.4~~X2.2 ~~(two-tailed)~~(two-sided) 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"~~ "preference" or ~~"no preference,"~~ "no preference," are based on the predetermined , , 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 -risk is now larger than the value chosen because a smaller number of respondents participated in the test. A result of "no preference" becomes more likely as N decreases.

10.3 *Analysis for Parity*—Different analyses are used depending on ~~if~~whether the number of respondents is equal to or greater than planned or fewer than planned. There is a direct relationship between sample size (N) and test sensitivity in parity testing.

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 , , 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 (standard deviation of P_c) $= \sqrt{P_c(1-P_c)/n}$~~

~~Confidence Limit $= P_c + z\,S_c$~~

Proportion of common responses

$P_c = c/n$

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

Confidence Limit $= P_c \pm z\,S_c$

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

-risk	z
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 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 -level of significance). If the confidence limit is greater than P_{max}, then conclude that the products are not at parity. Understand that the -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 Test sensitivity parameters: ,, and P_{max} levels, one-tailed or two-tailed test, critical value, decision risk;

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

11.1.6 ~~Any information or instructions given to the assessor in connection with the test; including how the product was identified when presented;~~The test environment: use of booths, simultaneous or sequential presentation and lighting conditions;

11.1.7 ~~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.~~administrator;

11.1.8 Next steps.

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~~cannot be 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.~~bias if the same target population is sampled from over repeated preference tests and the underlying population is homogeneous in its preferences.

13. Keywords

13.1 paired preference; preference; sensory; test method

APPENDIXES

(Nonmandatory Information)

X1. EXAMPLE ~~1—PRODUCT IMPROVEMENT:~~ X1: PAIRED PREFERENCE TEST: BEVERAGE FLAVORING 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"~~ that is sweeter and more "chocolatey" is preferred when used in a milk alternative beverage prior to fielding ~~a~~ more expensive in-home consumer ~~test. It~~ testing. Chocolate flavor "A" is a new, less expensive flavor that was determined by descriptive analysis to be higher in Sweetness and Chocolate Flavor impact. It is hypothesized by the development team that this sweeter flavor system will also be preferred and is intended to replace chocolate flavor "B," which is the current product. 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~~one-sided 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 = 0.05, and a P_{max} of 70 % with = 0.01. The analyst enters Table X1.1 in the section corresponding to P_{max} = 0~~.70~~ 70 % and the column