Designation: E2610 – 08 (Reapproved 2011)

Standard Test Method for Sensory Analysis—Duo-Trio Test[1]

This standard is issued under the fixed designation E2610; 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 test method covers a procedure for determining whether a perceptible sensory difference exists between samples of two products.

1.2 This test method applies whether a difference may exist in a single sensory attribute or in several.

1.3 This test method is applicable when the nature of the difference between the samples is unknown. It does not determine the size or the direction of the difference. The attribute(s) responsible for the difference are not identified.

1.4 Compared to the triangle test, the duo-trio test is statistically less efficient, but easier to perform by the assessors. For details on how the duo-trio test compares to other three-sample tests, see Refs (**1-4**).[2]

1.5 This test method is applicable only if the products are homogeneous. If two samples of the same product can often be distinguished, then another method, for example, descriptive analysis, may be more appropriate.

1.6 This test method is applicable only when the products do not cause excessive sensory fatigue, carryover or adaptation.

1.7 *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:*[3]

E253 Terminology Relating to Sensory Evaluation of Materials and Products

E456 Terminology Relating to Quality and Statistics

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

E1885 Test Method for Sensory Analysis—Triangle Test

2.2 *ISO Standards:* [4]

ISO 4120 Sensory Analysis—Methodology—Triangle Test

ISO 10399 Sensory Analysis—Methodology—Duo-Trio Test

3. Terminology

3.1 *Definitions*—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*—probability of concluding that a perceptible difference exists when, in reality, one does not. (Also known as Type I Error or significance level.)

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

3.2.3 p_c—probability of a correct response.

3.2.4 p_d *(proportion of discriminators)*—proportion of the population represented by the assessors that can distinguish between the two products.

3.2.5 *product*—material to be evaluated.

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

3.2.7 *sensitivity*—general term used to summarize the performance characteristics of the test. The sensitivity of the test is rigorously defined, in statistical terms, by the values selected for α, β, and p_d.

3.2.8 *triad*—three samples given to an assessor in the duo-trio test; one sample is labeled as a reference the other two samples are labeled with different codes. One of the coded samples is the same product as the reference. The other coded sample is different.

4. Summary of Test Method

4.1 Clearly define the test objective in writing.

[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 Nov. 15, 2011. Published March 2012. Originally approved in 2008. Last previous edition approved in 2008 as E2610 – 08. DOI: 10.1520/E2610-08R11.

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

[3] 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.

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

4.2 Choose the number of assessors based on the level of sensitivity desired for the test. The sensitivity of the test is, in part, a function of two competing risks: the risk of declaring the samples different when they are not (that is, α-risk) and the risk of not declaring the samples different when they are (that is, β-risk). Acceptable values of α and β vary depending on the test objective and should be determined before the test (see for example Appendix X1 and Appendix X2).

4.3 Each assessor receives a triad where one sample is labeled as the reference and the other two samples are labeled with different codes. The assessors are informed that one of the coded samples is the same as the reference and that one is different. The assessors report which of the coded samples they believe to be the same as (or different from) the reference.[5]

4.4 Results are tallied and significance determined by reference to a statistical table.

5. Significance and Use

5.1 The test method is effective for the following test objectives:

5.1.1 To determine whether a perceivable difference results or a perceivable difference does not result, for example, when a change is made in ingredients, processing, packaging, handling or storage; or

5.1.2 To select, train and monitor assessors.

5.2 The test method itself does not change whether the purpose of the duo-trio test is to determine that two products are perceivably different versus that the products are not perceivably different. Only the selected values of p_d, α, and β change. If the objective of the test is to determine if there is a perceivable difference between two products, then the value selected for α is typically smaller than the value selected for β. If the objective is to determine if the two products are sufficiently similar to be used interchangeably, then the value selected for β is typically smaller than the value selected for α and the value of p_d is selected to define "sufficiently similar."

5.3 The test method may change based on the test objective or the assessors' familiarity with the product. The balanced-reference technique (see 9.1.1) typically is used when neither product is more familiar than the other. The constant-reference technique (see 9.1.2) frequently is used when one product is a control/current product or is familiar to the assessors.

6. Apparatus

6.1 Carry out the test under conditions that prevent contact between assessors until the evaluations have been completed, for example, using booths that comply with Ref (**5**).

6.2 Sample preparation and serving sizes should comply with Guide E1871. See Refs (**6**) or (**7**).

7. Assessors

7.1 All assessors must be familiar with the mechanics of the duo-trio test (the format, the task, and the procedure of evaluation). Experience and familiarity with the product and test method may increase the sensitivity of an assessor and may therefore increase the likelihood of finding a significant difference. Monitoring the performance of assessors over time may be useful.

7.2 Choose assessors in accordance with test objectives. For example, to project results to a general consumer population, assessors with unknown sensitivity might be selected. To increase protection of product quality, assessors with demonstrated acuity should be selected.

7.3 The decision to use trained or untrained assessors should be addressed prior to testing. Training may include a preliminary presentation on the nature of the samples and the problem concerned. If the test concerns the detection of a particular taint, consider the inclusion of samples during training that demonstrate its presence and absence. Such demonstration will increase the panel's acuity for the taint but may detract from other differences. See Ref (**8**) for details. Allow adequate time between the exposure to the training samples and the actual duo-trio test to avoid carryover.

7.4 During the test sessions, avoid giving information about product identity, expected treatment effects or individual performance until all testing is complete.

7.5 Avoid replicate evaluations by the same assessor whenever possible. However, if replications are needed to produce a sufficient number of total evaluations, every effort should be made to have each assessor perform the same number of replicate evaluations.

8. Number of Assessors

8.1 Choose the number of assessors to yield the level of sensitivity called for by the test objectives. The sensitivity of the test is a function of three values: the α-risk, and the β-risk, and the maximum allowable proportion of distinguishers, p_d.[6]

8.2 Prior to conducting the test, select values for α, β and p_d. The following can be considered as general guidelines.

8.2.1 *For α-risk*—A statistically significant result at:

10 to 5 % (0.10 to 0.05) indicates "slight" evidence that a difference was apparent
5 to 1 % (0.05 to 0.01) indicates "moderate" evidence that a difference was apparent
1 to 0.1 % (0.01 to 0.001) indicates "strong" evidence that a difference was apparent
Below 0.1 % (<0.001) indicates "very strong" evidence that a difference was apparent

8.2.2 *For β-risk*—The strength of the evidence that a difference was not apparent is assessed using the same criteria as above (substituting "was not apparent" for "was apparent").

8.2.3 *For p_d*—The maximum allowable proportion of distinguishers, p_d, falls into three ranges:

[5] Organizations differ in the instructions they give their assessors. Some organizations instruct their assessors to select the sample that is most similar to the reference. Others instruct their assessors to select the sample that is most different from the reference. Either approach is acceptable.

[6] In this test method, the probability of a correct response, p_c, is modeled as $p_c = 1 \cdot p_d + (1/2) \cdot (1 - p_d)$, where p_d is the proportion of the entire population of assessors who can distinguish between the two products. It is a strictly statistical "guessing model" of the assessor's behavior. It is not a psychometric model of the assessor's decision process, such as the Thurstone-Ura model that could also be applied in discrimination testing.

p_d < 25 % represent small values
25 % < p_d < 35 % represent medium sized values
p_d > 35 % represent large values

8.3 Having defined the required level of sensitivity for the test using 8.2, use Table A1.1 to determine the number of assessors necessary. Enter Table A1.1 in the section corresponding to the selected value of p_d and the column corresponding to the selected value of β. The minimum required number of assessors is found in the row corresponding to the selected value of α. Alternatively, Table A1.1 can be used to develop a set of values for p_d, α and β that provide acceptable sensitivity while maintaining the number of assessors within practical limits. The approach is presented in detail in Ref (**9**).

8.4 Often in practice, the number of assessors is determined by material conditions (for example, duration of the experiment, number of available assessors, quantity of product). However, increasing the number of assessors increases the likelihood of detecting small proportions of distinguishers. Thus, one should expect to use larger numbers of assessors when trying to demonstrate that products are similar compared to when one is trying to prove they are different. Often 20 to 36 assessors are used when testing for a difference. For comparable sensitivity when testing for similarity, 40 to 78 assessors are needed.

9. Procedure

9.1 If neither product is more familiar than the other, use the balanced reference technique (9.1.1). If the product is familiar to the assessors (for example, a control sample from the production line), use the constant reference technique (9.1.2).

9.1.1 *Balanced-Reference Technique*—Prepare worksheets and scoresheets (see Appendix X1) in advance of the test so as to utilize an equal number of the four possible sequences of two products, A and B:

A_RAB A_RBA
B_RAB B_RBA

9.1.1.1 Distribute these at random among the assessors so that serving order is balanced.

9.1.2 *Constant-Reference Technique*—Prepare worksheets and scoresheets (see Appendix X2) in advance of the test so as to utilize an equal number of the two possible sequences of two products, A and B:

A_RAB A_RBA

9.1.2.1 Distribute these at random among the assessors so that serving order is balanced.

9.2 Present each triad simultaneously if possible, following the same spatial arrangement for each assessor (on a line to be sampled always from left to right, in a triangular array, etc.). Within the triad, assessors are typically allowed to make repeated evaluations of each sample as desired. If the conditions of the test require the prevention of repeat evaluations 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. In addition, if the samples change over time, for example, chewing gum or cereal with milk, samples should be tested sequentially.

9.3 Instruct the assessors to evaluate the reference sample first and then evaluate the two coded samples in the order in which they were presented. The assessor should then indicate which of the two coded samples is the same as the reference.

9.4 Each scoresheet should provide for a single triad of samples. If a different set of products is to be evaluated by an assessor in a single session, the completed scoresheet and any remaining product should be returned to the test administrator prior to receiving the subsequent triad. The assessor cannot go back to any of the previous samples or change the verdict on any previous test.

9.5 Do not ask questions about preference, acceptance, or degree of difference after the initial selection of the sample that matches the reference. The selection the assessor has just made may bias the reply to any additional questions. Responses to such questions may be obtained through separate tests for preference, acceptance, degree of difference, etc. (see Ref (**10**)). A comment section asking why the choice was made may be included for the assessor's remarks.

9.6 The duo-trio test is a forced-choice procedure; assessors are not allowed the option of reporting "no difference." An assessor who detects no difference between the samples and requests to report "no difference" should be instructed to randomly select one of the coded samples as being the same as the reference. In such situations the assessor can indicate that the selection was only a guess in the comments section of the scoresheet.

10. Analysis and Interpretation of Results

10.1 Use Table A1.2 to analyze the data obtained from a duo-trio test. The actual number of assessors can be greater than the minimum value given in Table A1.1. If the number of correct responses is greater than or equal to the number given in Table A1.2, conclude that a perceptible difference exists between the samples. If the number of correct answers is less than the number given in Table A1.2, conclude that the samples are sufficiently similar. Again, the conclusions are based on the risks accepted when the level of sensitivity (that is, p_d, α and β) was selected in determining the number of assessors (Table A1.1).

10.2 If desired, calculate a confidence interval on the proportion of the population that can distinguish the samples. This method is described in Appendix X3.

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, age, lot number, packaging, where obtained, 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 differ only due to the variable of interest, if at all);

11.1.3 The number of assessors, the number of correct selections, and the result of the statistical evaluation;

11.1.4 *Assessors*—Age, gender, experience in sensory testing, experience with the product category, experience with the samples in the test;

11.1.5 Any information and any specific instructions given the assessor in connection with the test;

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

11.1.7 The location and date of the test and the name of the panel leader.

12. Precision and Bias

12.1 Because results of sensory difference tests are functions of individual sensitivities, a general statement regarding the precision of results that is applicable to all populations of assessors cannot be made. However, adherence to the recommendations stated in this standard should increase the reproducibility of results and minimize bias.

13. Keywords

13.1 difference testing; discrimination test; duo-trio test; sensory analysis; similarity testing

ANNEX

(Mandatory Information)

A1. NUMBER OF ASSESSORS AND CORRECT RESPONSES NEEDED FOR A DUO-TRIO TEST

TABLE A1.1 Number of Assessors Needed for a Duo-Trio Test (10)

NOTE 1—Entries are the minimum number of assessors required to execute a duo-trio test with a prespecified level of sensitivity determined by the values chosen for p_d, α and β. Enter the table in the section corresponding to the chosen value of p_d and the column corresponding to the chosen value of β. Read the minimum number of assessors from the row corresponding to the chosen value of α.

		β							
α		0.50	0.40	0.30	0.20	0.10	0.05	0.01	0.001
	p_d = 50 %								
0.40		2	4	4	6	10	14	27	41
0.30		2	5	7	9	13	20	30	47
0.20		5	5	10	12	19	26	39	58
0.10		9	9	14	19	26	33	48	70
0.05		13	16	18	23	33	42	58	82
0.01		22	27	33	40	50	59	80	107
0.001		38	43	51	61	71	83	107	140
	p_d = 40 %								
0.40		4	4	6	8	14	25	41	70
0.30		5	7	9	13	22	28	49	78
0.20		5	10	12	19	30	39	60	94
0.10		14	19	21	28	39	53	79	113
0.05		18	23	30	37	53	67	93	132
0.01		35	42	52	64	80	96	130	174
0.001		61	71	81	95	117	135	176	228
	p_d = 30 %								
0.40		4	6	8	14	29	41	76	120
0.30		7	9	13	24	39	53	88	144
0.20		10	17	21	32	49	68	110	166
0.10		21	28	37	53	72	96	145	208
0.05		30	42	53	69	93	119	173	243
0.01		64	78	89	112	143	174	235	319
0.001		107	126	144	172	210	246	318	412
	p_d = 20 %								
0.40		6	10	23	35	59	94	171	282
0.30		11	22	30	49	84	119	205	327
0.20		21	32	49	77	112	158	253	384
0.10		46	66	85	115	168	214	322	471
0.05		71	93	119	158	213	268	392	554
0.01		141	167	207	252	325	391	535	726
0.001		241	281	327	386	479	556	731	944
	p_d = 10 %								
0.40		10	35	61	124	237	362	672	1124
0.30		30	72	117	199	333	479	810	1302
0.20		81	129	193	294	451	618	1006	1555
0.10		170	239	337	461	658	861	1310	1905
0.05		281	369	475	620	866	1092	1583	2237
0.01		550	665	820	1007	1301	1582	2170	2927
0.001		961	1125	1309	1551	1908	2248	2937	3812

TABLE A1.2 Number of Correct Responses Needed for Significance in a Duo-Trio Test (10)

NOTE 1—Entries are the minimum number of correct responses required for significance at the stated α-level (that is, column) for the corresponding number of respondents, *n* (that is, row). Reject the assumption of "no difference" if the number of correct responses is greater than or equal to the tabled value.

NOTE 2—For values of *n* not in the table, compute the missing entry as follows: Minimum number of correct responses (*x*) = nearest whole number greater than $x = n/2 + z\sqrt{n/4}$, where *z* varies with the significance level as follows: 0.25 for α = 0.40; 0.52 for α = 0.30; 0.84 for α = 0.20; 1.28 for α = 0.10; 1.64 for α = 0.05; 2.33 for α = 0.01; 3.09 for α = 0.001.

	α							α							
n	0.40	0.30	0.20	0.10	0.05	0.01	0.001	*n*	0.40	0.30	0.20	0.10	0.05	0.01	0.001
								31	17	18	19	20	21	23	25
2	2	2	—	—	—	—	—	32	18	18	19	21	22	24	26
3	3	3	3	—	—	—	—	33	18	19	20	21	22	24	26
4	3	4	4	4	—	—	—	34	19	20	20	22	23	25	27
5	4	4	4	5	5	—	—	35	19	20	21	22	23	25	27
6	4	5	5	6	6	—	—	36	20	21	22	23	24	26	28
7	5	5	6	6	7	7	—	40	22	23	24	25	26	28	31
8	5	6	6	7	7	8	—	44	24	25	26	27	28	31	33
9	6	6	7	7	8	9	—	48	26	27	28	29	31	33	36
10	6	7	7	8	9	10	10	52	28	29	30	32	33	35	38
11	7	7	8	9	9	10	11	56	30	31	32	34	35	38	40
12	7	8	8	9	10	11	12	60	32	33	34	36	37	40	43
13	8	8	9	10	10	12	13	64	34	35	36	38	40	42	45
14	8	9	10	10	11	12	13	68	36	37	38	40	42	45	48
15	9	10	10	11	12	13	14	72	38	39	41	42	44	47	50
16	10	10	11	12	12	14	15	76	40	41	43	45	46	49	52
17	10	11	11	12	13	14	16	80	42	43	45	47	48	51	55
18	11	11	12	13	13	15	16	84	44	45	47	49	51	54	57
19	11	12	12	13	14	15	17	88	46	47	49	51	53	56	59
20	12	12	13	14	15	16	18	92	48	50	51	53	55	58	62
21	12	13	13	14	15	17	18	96	50	52	53	55	57	60	64
22	13	13	14	15	16	17	19	100	52	54	55	57	59	63	66
23	13	14	15	16	16	18	20	104	54	56	57	60	61	65	69
24	14	14	15	16	17	19	20	108	56	58	59	62	64	67	71
25	14	15	16	17	18	19	21	112	58	60	61	64	66	69	73
26	15	15	16	17	18	20	22	116	60	62	64	66	68	71	76
27	15	16	17	18	19	20	22	122	63	65	67	69	71	75	79
28	16	16	17	18	19	21	23	128	66	68	70	72	74	78	82
29	16	17	18	19	20	22	24	134	69	71	73	75	78	81	86
30	17	17	18	20	20	22	24	140	72	74	76	79	81	85	89

APPENDIXES

(Nonmandatory Information)

X1. DUO-TRIO TEST TO CONFIRM THAT A DIFFERENCE EXISTS: FRAGRANCE FOR FACIAL TISSUE BOXES

X1.1 *Background*—An equipment supplier claims that their new fragrance delivery technology, applying the fragrance to the inside of the box, is superior to the current fragrance delivery method, applying the fragrance directly to the tissues. A product development fragrance chemist needs to confirm the claim before ordering a trial of the new equipment.

X1.2 *Test Objective*—To determine if the two methods of fragrance delivery produce any difference in the perceived fragrance of tissues after they have been stored for a period of time comparable to normal product age at time of use.

X1.3 *Number of Assessors*—To protect the fragrance chemist from falsely concluding that a difference exists, the sensory analyst proposes α = 0.05. Also, in order to keep the number of evaluations within reasonable limits, she suggests setting p_d at 40 % with β = 0.20. These values are agreed to by all parties concerned with the test. The analyst consults Table A1.1 in the section corresponding to p_d = 40 % and the column corresponding to β = 0.20. Then reading from the row corresponding to α = 0.05, she finds that a minimum of 37 assessors are needed for the test. In order to balance the orders of presentation of the samples, the analyst decides to use 40 assessors.

X1.4 *Conducting the Test*:

X1.4.1 Sufficient quantities of product are prepared using both of the fragrance delivery technologies. Products are stored under the same conditions for three months (the typical age of use of tissues). Sixty (60) samples of "A" tissues (tissues with fragrance applied directly to them) and 60 samples of "B" tissues (tissues with fragrance applied to the inside of the box)

are prepared for the test. Ten of each of the four possible triads: A_RAB, A_RBA, B_RAB and B_RBA, are prepared.

X1.4.2 The test is conducted with 40 assessors who have some experience in odor evaluation. The samples are prepared by the fragrance chemist, using the same fragrance and the same tissues on the same day. The boxed tissues are then stored under identical conditions for 3 months. Test tissues are taken from the center 50 % of the box; each tissue is placed in a sealed glass jar 1 h prior to evaluation. This allows for some fragrance to migrate to the headspace, and the use of the closed container reduces the amount of fragrance buildup in the testing booths. Each of the two samples is used as the reference in half (20) of the evaluations. Fig. X1.1 shows the scoresheet used.

X1.5 *Analysis and Interpretation of Results*—Only 21 out of the 40 subjects chose the correct match to the designated reference. According to Table A1.2, 26 correct responses are required at an α-risk of 5 %. In addition, when the data are reviewed for possible effects from the position of each sample as reference, the results show that the distribution of correct responses is even (see Ref (**10**)). This indicates that the quality or quantity, or both, of the two fragrances have little, if any, additional biasing effect on the results.

X1.6 *Report and Conclusions* —The sensory analyst informs the fragrance chemist that the odor duo-trio test failed to detect any significant odor differences between the two fragrance delivery technologies given the fragrance, the tissue, and the storage time used in the study at $\alpha = 0.05$ and $\beta = 0.20$.

DUO-TRIO TEST **Test No.** 230S

Panelist No. 21 Name: _____ Date: _____

Type of Sample: Facial tissue in a glass jar

Instructions

1. Please sniff each sample, starting at the left. Remove the cap only briefly and take short, shallow sniffs.
2. The left hand sample is a reference. Determine which of the two coded samples matches the fragrance of the reference.
3. Indicate the matching sample by placing an X in the corresponding box.

If no difference is apparent between the two unknown samples, you must guess.

Reference **Code** _____ **Code** _____

☒ ☐ ☐

Comments: _____

_____

FIG. X1.1 Scoresheet for Duo-Trio Test
Example X1: Balanced Reference Mode

X2. DUO-TRIO TEST WITH BALANCED RISKS: NEW CAN LINER

X2.1 *Background*—A brewer is faced with two supplies of cans, "A" being the regular supply he has used for years and "B" a proposed new supply said to provide a slight advantage in shelf life. He wants to know whether any difference can be detected between the two cans. The brewer feels that it is important to balance the risk of introducing an unwanted change to his beer against the risk of passing up the extended shelf life offered by can "B."

X2.2 *Test Objective*—To determine if any sensory difference can be perceived between the two beers after 8 weeks of shelf storage at room temperature.

X2.3 *Number of Assessors*—The brewer knows from past experience that if no more than $p_d = 30\ \%$ of his panel can detect a difference it is safe to assume that no meaningful difference exists. He is slightly more concerned with introducing an unwanted difference than he is with passing up the slightly extended shelf life offered by can "B." Therefore, he decides to set the β-risk at 0.05 and his α-risk at 0.10. Referring to Table A1.1 in the section for $p_d = 30\ \%$, the column for $\beta = 0.05$ and the row for $\alpha = 0.10$, he finds that 96 respondents are required for the test.

X2.4 *Conducting the Test*—A duo-trio test in the constant reference mode is appropriate because the company's beer in can "A" is familiar to the tasters. A separate test is conducted at each of the brewer's three testing sites. Each test is set up with 32 subjects, with "A" as the reference; 64 glasses of beer "A" and 32 of beer "B" are prepared and served to the subjects

in 16 combinations A_RAB and 16 combinations A_RBA, the left-hand sample being the reference.

X2.5 *Analysis and Interpretation of Results*—From the three test sites, 18, 20, and 19 subjects correctly identified the sample that matched the reference. In this test, all cans were obtained from the same lot and the subjects were from the same panel, so combination of the three test results is permissible: 18 + 20 + 19 = 57 correct out of 96 trials. From Table A1.2, the number of correct responses for significance at the 10 % risk level with 96 assessors is 55. The brewer concludes that a perceptible difference exists. Next, he examines the comments made by panelists to determine if there is a consistent description of the difference. If none is found, he may submit the samples to a descriptive panel. Ultimately, if a difference is found a consumer test may be required to determine if there is preference for one can or the other.

X3. CONFIDENCE INTERVALS FOR DUO-TRIO TESTS

X3.1 *Background*—If desired, analysts can calculate a confidence interval on the proportion of the population that can distinguish the samples. The calculations are as follows, where c = the number of correct responses and n = the total number of assessors:

p_c (proportion correct) = c/n
p_d (proportion distinguishers) = $2p_c - 1$
s_d (standard deviation of p_d) = $2\sqrt{p_c(1-p_c)/n}$
upper confidence limit = $p_d + z_\beta s_d$
lower confidence limit = $p_d - z_\alpha s_d$

X3.1.1 z_α and z_β are critical values of the standard normal distribution. For a one-sided 90 % confidence interval, $z = 1.28$; for a one-sided 95 % confidence interval, $z = 1.64$; and for a one-sided 99 % confidence interval, $z = 2.33$. For a two-sided 90 % confidence interval, $z = 1.64$; for a two-sided 95 % confidence interval, $z = 1.96$; and for a two-sided 99 % confidence interval, $z = 2.58$.

X3.2 *Analysis and Interpretation of Results*—Consider the data from Appendix X2, where $c = 57$, $n = 96$, $\alpha = 0.10$, and $\beta = 0.05$. It follows that:

p_c (proportion correct) = 57/96 = 0.59375
p_d (proportion distinguishers) = 2(0.59375) – 1 = 0.1875
s_d (standard deviation of p_d) = $2\sqrt{0.59375(0.40625)/96}$ = 0.10025
upper confidence limit = 0.1875 + 1.64(0.10025) = 0.352
lower confidence limit = 0.1875 - 1.28(0.10025) = 0.059

X3.2.1 Considered individually, the brewer can be 90 % confident that at least 5.9 % of the assessors can distinguish the new can liner from the current one and he can be 95 % confident that the proportion of distinguishers may be as large as 35.2 %. The finding that at least 5.9 % of the assessors can distinguish the samples supports the conclusion drawn in Appendix X2, that is, that there is a perceptible difference between the samples. The upper limit of the confidence interval ($p_d \leq 35$ %) also supports the conclusion reached in Appendix X2 because it is greater than the proportion of distinguishers that the brewer was concerned with. The introduction of the new can liner may pose a real risk in the market place.

REFERENCES

(1) Ennis, D. M., "The Power of Sensory Discrimination Methods," *Journal of Sensory Stud.*, Vol 8, 1993, pp. 353-370.
(2) Frijters, J. E. R., "Three-Stimulus Procedure in Olfactory Psychophysics: An Experimental Comparison of Thurstone-Ura and Three-Alternative Forced-Choice Models of Signal Detection Theory," *Perception and Psychophysics*, Vol 28, No 5, 1980, pp. 390-397.
(3) MacRae, S., "The Interplay of Theory and Practice Sensory Testing," *Chem and Ind.*, Jan. 5, 1987, pp. 7-12.
(4) O'Mahony, M. and Odbert, N., "A Comparison of Sensory Difference Testing Procedures: Sequential Sensitivity Analysis and Aspects of Taste Adaptation," *Journal of Food Science*, Vol 50, 1985, pp. 1055-1058.
(5) *Guidelines for Physical Requirements for Sensory Evaluation Laboratories*, ASTM STP 913, ASTM International.
(6) 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.
(7) 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.
(8) *Guidelines for the Selection and Training of Sensory Panel Members*, ASTM STP 758, ASTM International.
(9) Schlich, P., "Risk Tables for Discrimination Tests," *Food Quality and Preference*, Vol 4 1993, pp. 141-151.
(10) Meilgaard, M., Civille, G. V., and Carr, B. T., *Sensory Evaluation Techniques*, 4th Edition, CRC Press, Inc., Boca Raton, FL, 2006.