Designation: E3009 – 15

StandardTest Method for Sensory Analysis—Tetrad Test[1]

This standard is issued under the fixed designation E3009; 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 or to estimate the magnitude of the perceptible difference.

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. The attribute(s) responsible for the difference are not identified.

1.4 The tetrad test is more efficient statistically than the triangle test (Test Method E1885) or the duo-trio test (Test Method E2610).

1.5 The tetrad method involves the evaluation of four samples. When the products being tested cause excessive sensory fatigue, carryover, or adaptation, methods that involve the evaluation of fewer samples (same-different, triangle test, etc.) may be preferred.

1.6 *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

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

E1885 Test Method for Sensory Analysis—Triangle Test

E2262 Practice for Estimating Thurstonian Discriminal Distances

E2610 Test Method for Sensory Analysis—Duo-Trio Test

2.2 *ISO Standards:*[3]

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.

3.2.1.1 *Discussion*—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.

3.2.2.1 *Discussion*—Also known as Type II Error.

3.2.3 δ—Thurstonian measure of sensory difference (effect size) relative to perceptual noise (standard deviation) (see Practice E2262).

3.2.4 *product*—material to be evaluated.

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

3.2.6 *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 δ.

4. Summary of Test Method

4.1 Clearly define the test objective in writing.

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).

[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 May 1, 2015. Published June 2015. DOI: 10.1520/E3009-15.

[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 International Organization for Standardization (ISO), 1, ch. de la Voie-Creuse, CP 56, CH-1211 Geneva 20, Switzerland, http://www.iso.org.

4.3 Each assessor receives four coded samples where two samples are of one product and the other two samples are of the other product being tested. The assessors are instructed to group the four samples into two pairs based on the level of similarity between samples.

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 perceptible difference results or a perceptible 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 test is to determine that two products are perceptibly different versus that the products are not perceptibly different. Only the selected values of δ, α, and β change. If the objective of the test 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 δ is selected to define "sufficiently similar."

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 STP 913 (**1**).[4]

6.2 Sample preparation and serving sizes should comply with Practice E1871. See Refs (**2**) or (**3**).

7. Assessors

7.1 All assessors must be familiar with the mechanics of the tetrad 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, if the project results are to represent the 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 STP 758 for details (**4**). Allow adequate time between the exposure to the training samples and the actual tetrad 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 sensory difference, δ.

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

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

8.2.1.1 10 to 5 % (0.10 to 0.05) indicates "slight" evidence that a difference was apparent;

8.2.1.2 5 to 1 % (0.05 to 0.01) indicates "moderate" evidence that a difference was apparent;

8.2.1.3 1 to 0.1 % (0.01 to 0.001) indicates "strong" evidence that a difference was apparent; and

8.2.1.4 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" or "was apparent").

8.2.3 For δ: The maximum allowable sensory difference, δ, falls into three ranges:

8.2.3.1 $\delta < 0.5$ represent small values;

8.2.3.2 $0.5 < \delta < 1$ represent medium sized values; and

8.2.3.3 $\delta > 1$ 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 δ 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 δ, α, and β that provide acceptable sensitivity while maintaining the number of assessors within practical limits. The approach is presented in detail in Ref (**5**).

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). Increasing the number of assessors increases the likelihood of detecting small values of δ. Thus, one should expect to use larger numbers of assessors when trying to demonstrate that products are similar compared to when testing for a difference. For comparable sensitivity when testing for similarity, 40 to 50 assessors are needed.

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

9. Procedure

9.1 Prepare worksheets and scoresheets (see Appendix X1) in advance of the test so as to utilize an equal number of the six possible sequences of two products, A and B:

AABB	BBAA
ABAB	BABA
ABBA	BAAB

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

9.2 Present each set of four samples simultaneously if possible, following the same spatial arrangement for each assessor. Within the set of four samples, 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 or leave an aftertaste, present the samples sequentially and do not allow repeated evaluations. In addition, if the samples change over time, for example, cereal with milk, samples should be tested sequentially.

9.3 Instruct the assessors to evaluate the four test samples in the order presented. The assessor should then group the four samples into two groups of two based on similarity. It is critical that the instructions to the assessors say, "Group the four samples into two groups of two based on similarity," and not, "Identify the two samples that are most similar to each other." The latter wording does not correctly represent the tetrad task the assessor is to perform.

9.4 Each scoresheet should provide for a single group 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 from the evaluation just completed should be returned to the test administrator prior to receiving the subsequent set of test samples. 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 grouping of samples into pairs. 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 Manual 26) (**6**). A comment section asking why the choice was made may be included for the assessor's remarks.

9.6 The tetrad 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 group the test samples into two pairs randomly. 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 tetrad 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, δ, α, and β) was selected in determining the number of assessors (Table A1.1).

10.2 If desired, calculate a confidence interval on the sensory difference. This method is described in Appendix X1.

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, 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 is 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; sensory analysis; similarity testing; tetrad test