International Standard ISO 6658

Sensory analysis — Methodology — General guidance

Analyse sensorielle — Méthodologie — Guide général

First edition — 1985-06-15

UDC 543.92   Ref. No. ISO 6658-1985 (E)

Descriptors : agricultural products, food products, sensory analysis, generalities.

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work.

Draft International Standards adopted by the technical committees are circulated to the member bodies for approval before their acceptance as International Standards by the ISO Council. They are approved in accordance with ISO procedures requiring at least 75 % approval by the member bodies voting.

International Standard ISO 6658 was prepared by Technical Committee ISO/TC 34, *Agricultural food products.*

Printed in Switzerland

Contents

Sensory analysis — Methodology — General guidance

0 Introduction

This International Standard constitutes a general introduction to the methodology of sensory analysis and should be read before undertaking the more detailed test procedures described in other International Standards. It covers the general area of methodology and is intended to fulfil the following functions :

    a) to provide a brief background of the essential features of methods of sensory analysis for the user of specific tests;

    b) to provide details of general requirements, procedures and interpretation of results common to all or most tests;

    c) to provide sufficient guidance on requirements, procedures and interpretation of results for the different specific tests to allow choice of the most appropriate procedure(s) for solution of a particular problem.

It comprises three main parts, covered in clauses 4, 5 and 6.

It is essential that clause 4 "General requirements" should be read first. Clause 5 "Test methods" describes, in a general manner, all the main tests, under five headings :

    a) Definitions.

    b) Application.

    c) Assessors.

    d) Procedure.

    e) Interpretation of results.

Clause 6 "Analysis of results" is concerned with the statistical treatment of the results and should be read in conjunction with the cross-references in the text headed "Interpretation of results" in clause 5 "Test methods".

1 Scope and field of application

This International Standard gives general guidance on the methodology of sensory analysis. It describes tests for the examination of food products by sensory analysis, and includes some information on the techniques to be used if statistical analysis of the results is required. The principles embodied in them may apply to the examination of other products by means of the human senses. Generally these tests are intended only for sensory analysis in laboratories, and are not applicable to the determination of the consumer's preference. However, if the test can be used for determining preference, this is indicated in the subclauses entitled "Application".

2 References

ISO 2854, *Statistical interpretation of data — Techniques of estimation and tests relating to means and variances.*

ISO 2859, *Sampling procedures and tables for inspection by attributes.* [1]

ISO 3534, *Statistics — Vocabulary and symbols.*

ISO 3591, *Sensory analysis — Apparatus — Wine-tasting glass.*

ISO 3951, *Sampling procedures and charts for inspection by variables for percent defective.*

ISO 3972, *Sensory analysis — Determination of sensitivity of taste.*

ISO 4120, *Sensory analysis — Methodology — Triangular test.*

ISO 4121, *Sensory analysis — Methodology — Grading of food products by methods using scales and categories.* [2]

ISO 5492/1 to 6, *Sensory analysis — Vocabulary.* [3]

ISO 5494, *Sensory analysis — Apparatus — Tasting glass for liquid products.*

ISO 5495, *Sensory analysis — Methodology — Paired comparison test.*

ISO 5497, *Sensory analysis — Methodology — Guidelines for the preparation of samples for which direct sensory analysis is not feasible.*

ISO 6564, *Sensory analysis — Methodology — Flavour profile methods.* [2]

ISO 8587, *Sensory analysis — Methodology — Ranking.* [2]

ISO 8588, *Sensory analysis — Methodology — "A"-"not A" test.* [2]

---

1) At present at the stage of draft. (Revision of ISO 2859-1974.)

2) At present at the stage of draft.

3) Part 6 is at present at the stage of draft.

3 Definitions

For the purpose of this International Standard, the definitions of the various parts of ISO 5492 apply.

4 General requirements

4.1 Basic information

This clause covers the general requirements common to all situations encountered in sensory analysis. The information basic to these requirements is

a) the human response to one stimulus cannot be isolated from previous experience or from other sensory stimuli received from the environment. Nevertheless, influences arising from these two sources can be controlled and the effect standardized;

b) variability in sensory response is inherent in any group of people used for testing and is unavoidable; however, with training, such a group may show highly consistent individual responses. Recognition of these factors is important in the analysis of results;

c) the validity of the conclusions drawn from the results is dependent upon the test used and the way it is conducted including the questions that have been asked.

4.2 Statement of problem

In sensory analysis, a given problem frequently requires appreciable discussion or thought before an appropriate test is selected. This is because the initial concept of the problem may require clarification. It would, for example, be meaningless to carry out a preference test, with a large number of assessors, without having first established that there is a significant difference, which can be established with a much smaller number of assessors using a difference test.

Similarly, where many samples are concerned, it is useful to carry out preliminary tests to establish

— appropriate orders of magnitude, such as the concentration of the reference solution to be used

— the number of samples to be evaluated

— the relevance of the attributes (judged to be representative of the product to be analysed).

There are two main types of problem : those in which the primary aim of the test is to describe the product(s), and those in which the aim is to distinguish between two or more products. For the second type, it is important to distinguish between the need to know whether there is a difference at all, how great the magnitude of the difference, the direction (or quality) of that difference, the influence of that difference, for example, with regard to preference, and whether all or only part of a population is detecting a difference.

4.3 Choice of test

When the nature of the problem has been established, the choice of the appropriate test will also be related to the degree of confidence deemed necessary, the nature of the samples and the availability of assessors.

For each test, an attempt is made below to give guidance as to its relevance. Preliminary tests may be necessary.

Because of sensory fatigue and the effects of adaptation, only a limited number of samples, depending on the nature of the test and the type of product, can be assessed during a session.

The statistical plan should always be determined before commencing tests. This is especially recommended if the number of samples to be evaluated requires more than one session. Details of statistical plans should be selected from specialized texts.

Whilst the use of control samples is essential in most cases, their use naturally limits the number of samples which can be assessed during any given session.

Once it has been decided how to approach a particular problem, the choice of the tests to be used should be related to the number and type of assessors available.

4.4 Choosing and training assessors [1]

Guidance is given, for each test, on the minimum number of experts, selected assessors, or assessors required. Training will improve precision and some degree of selection can also be valuable. Note that the selection of assessors for their ability to discriminate and describe foods is quite different from that used for preference tests. The former tasks require selection and training whereas the latter require only that the panel be representative of a specified sector of the population, for example a group of consumers.

If a selection procedure is to be carried out, the important criteria are as follows :

a) management support in the firm;

b) availability, with respect to normal employment;

c) motivation (willingness and interest);

d) good health (including the absence of specific allergies or treatment with medications) and good dental and general hygienic condition.

According to the aim assessors should be selected and, if necessary, trained by repetition of a test method with, ideally, the full range of samples likely to be encountered. Any type of test may be used.

It is important that

— the selection method should be relevant to the test situation

— the same test should be carried out more than once to ensure that an assessor has not given the correct answer by chance.

---

1) The choosing and training of assessors will form the subject of a future International Standard.

This, in turn, requires a homogeneous sample which, in the case of solid foods, is not always easy to obtain.

With two homogeneous samples of known properties, difference testing is possible; with three or more, ranking or rating may be used.

The assessors should be selected according to their ability to discriminate and according to the consistency of this discrimination. This selection should, of course, apply to all who are to participate, with no exceptions on the basis of status or prior knowledge. The number of assessors selected for a given test should normally exceed (for example by 50 %) that required for any given panel, to allow for the normal level of absence.

The selection procedure should be based on inspection of whatever data are obtained with the criterion for selection of consistent, correct performance. It should be borne in mind that a smaller panel of good discriminators may be statistically less efficient than a larger panel including poorer discriminators. However, the inclusion of less discriminating assessors merely to increase panel size does not improve the sensory analysis. Similarly, if the recommended number of assessors cannot be obtained, repeated use of a small panel does not necessarily give the same result as the use of an equivalent large panel.

The subsequent participation of assessors in a given panel should be on the basis of random selection, but this should be balanced in such a way as to result in equal participation over a number of panels, because frequent participation is necessary for motivation and to guarantee performance. The performance of individuals should be frequently assessed in order to detect changes in ability.

4.5 Sensitivity tests

These tests are frequently used in the selection and training of assessors. They may be divided into three types :

a) those involving the use of very dilute aqueous solutions of single compounds or simple mixtures, intended to establish the different thresholds of the assessors, for example appearance, recognition or difference threshold (see ISO 3972);

b) those involving food substances at normal concentrations with or without low concentrations of other materials, for example tests for detection of tainting;

c) those involving dilution techniques where food substances (or chemical mixtures) are examined at decreasing or increasing concentrations.

4.6 Material to be tested

General principles for product sampling (in accordance with International Standards relating to the product under test) should be applied for test samples; valid conclusions can be drawn for a product as a whole only if the samples tested are representative.

Methods of preparation and presentation of samples should be appropriate to the product and to the problem concerned. A product which is normally consumed hot should be prepared in the usual manner and tested hot. Similarly, a product which is normally consumed in discrete pieces should not be homogenized. Care is needed, however, to ensure maximum uniformity between subsamples for each assessor.

Carriers may sometimes be used for tests relating to the evaluation of products for which direct tasting is not feasible (see ISO 5497).

All variation of the test samples (except those required by the test) shall be minimized.

When the test concerns only differences in flavour, the interactions which can exist shall be masked (for example by the use of light of a suitable colour).

Containers should be chosen so as not to affect the test, and lighting should be specified when appearance is being assessed.

Most of these factors relate to the particular product for testing.

4.7 Test room [1]

Sensory analysis should be conducted in a special test room.

The aim should be to create for each assessor a separate environment with minimum distraction so that each can quickly adjust to the nature of the new task(s). Extraneous activities, including preparation of the samples, should not be allowed during the tests, as these may lead to biased results. The room should be at a comfortable temperature and should be ventilated with odour-free air; limited air flow is desirable to avoid excessive temperature fluctuations. Persistent odours, such as tobacco or cosmetics, should not be allowed to contaminate the environment of the test room.

Sound should be restricted. A low background noise is usually more tolerable than a fluctuating level of noise. Conversation is more distracting than background noise. Interruptions cause the greatest distraction.

It is usually helpful to have control over both the colour and the intensity of the lighting, although coloured lights rarely succeed in completely masking differences in appearance.

The surfaces with which food comes into contact should be non-absorbent. The dimensions of the tasting booths are important; very low ceilings and very narrow booths can be oppressive or can give rise to a feeling of claustrophobia. Comfortable seating is necessary.

---

1) The design of test rooms will form the subject of a future International Standard.

4.8 Apparatus

Two general problems are those of maintaining the temperature of the samples at a constant level and avoiding taint. The temperature at which the samples are served should be appropriate for the product and any changes during the test should be minimized. Disposable containers are frequently used, but some are not suitable for hot products; it cannot be assumed that they are taint-free. Similarly, it is important to ensure adequate rinsing of reusable containers because detergents used for cleaning may leave a residual odour or flavour.

If standardized containers exist and are suitable for the test, they should preferably be used (for example wine-tasting glasses in accordance with ISO 3591, or tasting glasses for liquid products in accordance with ISO 5494).

Checks on the water supply are desirable to ensure that it is bland. For particular purposes, deionized water, glass-distilled water, low mineral content spring water, carbon-filtered water or boiled tap water may be used, but it should be noted that they are likely to have differing flavours.

4.9 Conduct of test

The conduct of the test will depend upon the decisions taken in the briefing of the assessors prior to the test. The design of the questionnaire (proforma, test form) may take account not only of the results given by the assessors but also of how the data are to be handled.

The manner and order of presentation of the samples are important aspects of the test. The samples should be coded, for example by random three-digit numbers, to minimize bias. It is important to vary the codes. The order of assessment can also be a source of bias and, in general, the order is specified. With a small number of samples and assessors, the order can be balanced so that every possible order occurs an equal number of times. In larger experiments, the order can be randomized.

The collation of the results comprises three aspects :

   a)   checking that all data have been recorded;

   b)   verification that any additional relevant information which may aid or cast doubt on the interpretation of the results has been noted;

   c)   checking that the assessors are motivated to continue participating at a desirable level of interest.

Hunger and satiety can influence an assessor's performance, and, if panels are held too frequently, performance may deteriorate. Expectoration of samples may be recommended with trained panels. If it is possible, assessors should be asked to refrain from smoking and from consuming anything but water for 1 h before a test. Similarly, the use of odoriferous cosmetics is undesirable. Foreign odours carried by assessors should be taken into account, together with their influence on the rest of the panel.

Assessors suffering from colds, emotional upsets, etc. should be excluded from tests until they recover.

The time of day at which the test is conducted is important. Maximum sensitivity is found at mid-morning and mid-afternoon, depending on local circumstances, but away from mealtimes.

5 Methods of test

5.1 Types of tests

Beside sensitivity tests described in 4.5 used for selection and training of assessors, the most commonly used tests are divided into three groups :

   a)   difference tests used to determine whether a sensory difference exists between two products (see 5.2);

   b)   tests using scales and categories, to estimate the order or size of differences or the categories or classes to which samples should be allocated (see 5.3);

   c)   analytical or descriptive tests, used to identify the specific sensory attributes present in a sample; the tests may also be quantitative (see 5.4).

5.2 Difference tests

5.2.1 General

The following tests are used to determine whether there is a sensory difference between two products :

   a)   paired comparison test (see 5.2.2);

   b)   triangular test (see 5.2.3);

   c)   duo-trio test (see 5.2.4);

   d)   two-out-of-five test (see 5.2.5);

   e)   "A"-"not A" test (see 5.2.6).

For all these tests, there are different manners of analysing the results, and these are described in 6.2.

5.2.2 Paired comparison test (see also ISO 5495)

5.2.2.1 Definition

paired comparison test : A test in which samples are presented in pairs for comparison and detection of differences on the basis of some defined criteria.

5.2.2.2 Application

The paired comparison test is recommended :

   a)   to determine if there is a difference, and if so, the direction of a difference between two samples;

   b)   to establish if there is a preference;

   c)   for the selection and training of assessors.

Advantages of the test over other difference tests are simplicity and less sensory fatigue.

The disadvantage of the test is that, as the number of samples to be compared increases, the number of inter-comparisons required rapidly becomes impracticable.

5.2.2.3 Assessors

The recommended number of assessors is

   7 or more experts;

   20 or more selected assessors;

   30 or more assessors.

More comprehensive studies, for example consumer tests, require a much larger number of assessors, of the order of hundreds.

5.2.2.4 Procedure

The assessor is given one or more pairs of coded samples, presented in a controlled or random order. The two samples in each pair may be the same or different. The assessor is asked specific relevant question(s) referring to the difference, the direction of difference or preference [see 5.2.2, a) and b)]. Questions of difference and preference shall not be combined.

5.2.2.5 Analysis of results

See 6.2.2.

5.2.3 Triangular test (see also ISO 4120)

5.2.3.1 Definition

triangular test : Difference test involving three coded samples, two of which are identical, presented simultaneously. The assessor is asked to select the odd sample.

5.2.3.2 Application

The triangular test is recommended

   a) to detect slight differences between samples;

   b) when only a limited number of assessors is available;

   c) for the selection and training of assessors.

The test shall not be used for the determination of preference.

Some disadvantages of the test are that

   a) it is uneconomical for the assessment of a large number of samples;

   b) it may be more affected by sensory fatigue than the paired comparison test with intensely flavoured samples;

   c) it may be difficult to ensure that two samples are identical.

5.2.3.3 Assessors

The recommended number of assessors is

   6 or more experts;

   15 or more selected assessors;

   25 or more assessors.

5.2.3.4 Procedure

The assessor is presented with one set of three coded samples, two of which are identical, and is asked to select the odd sample. Samples should be presented an equal number of times in each of the two sets of three distinct permutations of order which are

   BAA   ABB

   AAB   BBA

   ABA   BAB

5.2.3.5 Analysis of results

See 6.2.3.

5.2.4 Duo-trio test

5.2.4.1 Definition

duo-trio test : Difference test in which the reference sample is presented first. It is followed by two samples, one of which is identical to the reference sample and which the assessor is asked to identify.

5.2.4.2 Application

This duo-trio test is used to determine whether there is a sensory difference between a given sample and a reference. It is especially suitable when the reference sample is well known to the assessors, for example a sample of regular production.

If there are after-tastes, this test is less suitable than the paired comparison test (5.2.2) or the "A"-"not A" test (5.2.6).

5.2.4.3 Assessors

The recommended number is 20 or more assessors.

5.2.4.4 Procedure

The assessors are first presented with the identified reference sample. This is followed by two coded samples, one of which is identical to the reference sample. The assessor is asked to identify this sample.

5.2.4.5 Analysis of results

See 6.2.4.

5.2.5 Two-out-of-five test

5.2.5.1 Definition

two-out-of-five test : Difference test involving five coded samples, two of which are of one type and three of another. The assessor is asked to group the two sets of samples.

5.2.5.2 Application

The two-out-of-five test is recommended :

a) when only a small number of selected assessors (for example 10) is available;

b) to establish a difference more economically than by other tests (the method is statistically more efficient).

The disadvantages of this test are the same as those of the triangular test (5.2.3), except that it is more strongly affected by sensory fatigue and memory effects. Its principal use is in visual, auditory or tactile applications.

5.2.5.3 Assessors

The recommended number is 10 or more selected assessors.

5.2.5.4 Procedure

The assessor is presented with one set of five coded samples, two of which are of one type and three of another. The assessor is asked to group the two sets of samples. When the number of assessors is less than 20, the order of presentation should be selected at random from the following 20 distinct permutations :

AAABB	BBBAA
AABAB	BBABA
ABAAB	BABBA
BAAAB	ABBBA
AABBA	BBAAB
ABABA	BABAB
BAABA	ABBAB
ABBAA	BAABB
BABAA	ABABB
BBAAA	AABBB

5.2.5.5 Analysis of results

See 6.2.5.

5.2.6 "A"-"not A" test (see also ISO 8588)

5.2.6.1 Definition

"A"-"not A" test : Test in which a series of samples, which may be "A" or "not A", is presented to the assessor after he has learnt to recognize sample "A". The assessor is asked to indicate which sample is "A".

5.2.6.2 Application

This test is a difference test which may be used for the assessment of samples having variations of appearance or leaving a persistent after-taste.

It is especially useful when strictly similar repeated samples cannot be obtained.

5.2.6.3 Assessors

The recommended number is

20 selected assessors;

30 assessors.

5.2.6.4 Procedure

The assessor is presented with samples one at a time. The assessor is first presented with the reference sample "A" several times, until he can recognize it. He is then given several samples, each of which may be "A" or "not A", at random, and he has to determine which they are. An appreciable time interval (for example 2 to 5 min) should be allowed between receipt of samples, and only a few samples should be examined during one session.

5.2.6.5 Analysis of results

See 6.2.6.

5.3 Tests using scales and categories (see also ISO 4121)

5.3.1 Types of tests

The following tests are used to estimate the order or size of differences, or the categories or classes to which samples should be allocated :

a) ranking (see 5.3.2);

b) classification (see 5.3.3);

c) rating (see 5.3.4);

d) scoring (see 5.3.5);

e) grading (see 5.3.6).

5.3.2 Ranking (see also ISO 8587)

5.3.2.1 Definition

ranking : Test in which a series of three or more samples is presented to an assessor at the same time and which are to be arranged in order of intensity or degree of some specified attributes.

5.3.2.2 Application

Ranking has a wide application, but it is not very discriminating. It is recommended for use :

a) as a screening test, to aid the planning of a more precise assessment;

b) for selection of products;

c) as a consumer test for acceptance and determination of the order of preference;

d) for training assessors.

The test is rapid and of value in the assessment of a small number of samples (about six) with complex attributes (for example quality and flavour) and also when a large number of samples (of the order of 20) is to be assessed for appearance.

5.3.2.3 Assessors

As a function of the purpose of the test, described previously, the recommended number of assessors is

2 or more experts;

5 or more selected assessors;

10 or more assessors (or 100 or more for a consumer test).

5.3.2.4 Procedure

Before the test, it is necessary to ensure that the assessors understand and agree on the attribute or criterion to be evaluated. In the test, each assessor examines the coded samples in a prescribed order and as a result assigns a preliminary ranking. This ranking can be checked and adjusted by re-examination of the samples.

5.3.2.5 Analysis of results

See 6.3.2.

5.3.3 Classification

5.3.3.1 Definition

classification : Test in which samples are assigned to predefined categories (classes). The categories used in the classification are only nominal.

5.3.3.2 Application

Classification is recommended for use in a situation such as assessing the defects present in a product.

5.3.3.3 Assessors

The recommended number of assessors is

3 or more experts;

3 or more selected assessors.

5.3.3.4 Procedure

The classification to be used should be clearly defined and understood by the experts or selected assessors. Each assessor examines the samples and assigns each sample to one of the categories.

5.3.3.5 Analysis of results

See 6.3.3.

5.3.4 Rating

5.3.4.1 Definition

rating : Method of classification involving categories. Each category is composed of an ordered scale. The points on each scale are of an ordinal nature.

5.3.4.2 Application

Rating is recommended for use as a means of evaluating

a) the intensity of one or more attributes;

b) the degree of preference.

The test may give a more informative answer than ranking because it estimates the magnitude of attributes or preference.

5.3.4.3 Assessors

As a function of the purpose of the test, described previously, the recommended number of assessors is

a) for determination of intensity of attributes :

1 or more experts,

5 or more selected assessors,

20 or more assessors;

b) for determination of degree of preference :

50 or more assessors (for 2 samples) or

100 or more (for 3 or more samples).

7