INTERNATIONAL STANDARD

ISO 8586

First edition
2012-12-15

Sensory analysis — General guidelines for the selection, training and monitoring of selected assessors and expert sensory assessors

Analyse sensorielle — Lignes directrices générales pour la sélection, l'entraînement et le contrôle des sujets qualifiés et sujets sensoriels experts

Reference number
ISO 8586:2012(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

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. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO 8586 was prepared by Technical Committee ISO/TC 34, *Food products*, Subcommittee SC 12, *Sensory analysis*.

This first edition of ISO 8586 cancels and replaces ISO 8586-1:1993 and ISO 8586-2:2008. The main requirements and criteria for the selection, training and monitoring of selected assessors and of expert sensory assessors have been revised to summarize the information given in ISO 8586-1:1993 and ISO 8586-2:2008.

Introduction

A sensory analysis panel constitutes a true “measuring instrument”, and consequently the results of the analysis depends on its members.

The recruitment of persons willing to participate in a panel therefore needs to be carried out with care and to be considered as a real investment, both in time and money.

Sensory assessment can be performed by three types of assessors:

— sensory assessors;

— selected assessors;

— expert sensory assessors.

“Sensory assessors” are any people taking part in a sensory test. They can be “naive assessors” who do not have to meet any precise criterion, or “initiated assessors” who have already participated in sensory tests (see ISO 5492:2008, 1.5).

“Selected assessors” are chosen for their ability to perform a sensory test (see ISO 5492:2008, 1.6).

“Expert sensory assessors” are selected assessors with a demonstrated sensory sensitivity and with considerable training and experience in sensory testing, who are able to make consistent and repeatable sensory assessments of various products (see ISO 5492:2008, 1.8).

It is necessary to undertake a preliminary selection of the candidates at the recruitment stage, in order to eliminate those who would be unsuited for sensory analysis. However, the final selection can only be made after selection and training. The selection and training methods to be employed depend on the tasks to intend for the “selected assessors” and “expert sensory assessors”.

Sensory assessors work as a panel which is managed by a panel leader. In certain cases (especially for descriptive sensory analysis), the panel may be divided into specialized subgroups.

The recommended procedure involves:

a) recruitment and preliminary screening of naive assessors;

b) familiarization of naive assessors who are to become initiated assessors;

c) selection of initiated assessors in order to determine their ability to perform particular tests, who then become selected assessors;

d) possible training of selected assessors to become expert sensory assessors.

The exact procedures covered by a) and b) and the nature of the tests performed in c) and d) depend on the tasks intended for the panel.

Expert sensory assessors have demonstrated particular acuity and reproducibility in panel work, and have developed a good long-term sensory memory, allowing reliable comparative judgements, possibly in the absence of control samples.

The panel leader is responsible for the general monitoring of the group of expert sensory assessors and for their training. The expert sensory assessors are not responsible for the choice of tests used, the presentation of the samples or for the interpretation of results. These matters are the responsibility of the panel leader who also decides how much information is given to the panel.

The performance of selected assessors should be monitored regularly to ensure that the criteria by which they were initially selected continue to be met.

The entire process is illustrated in Figure 1.

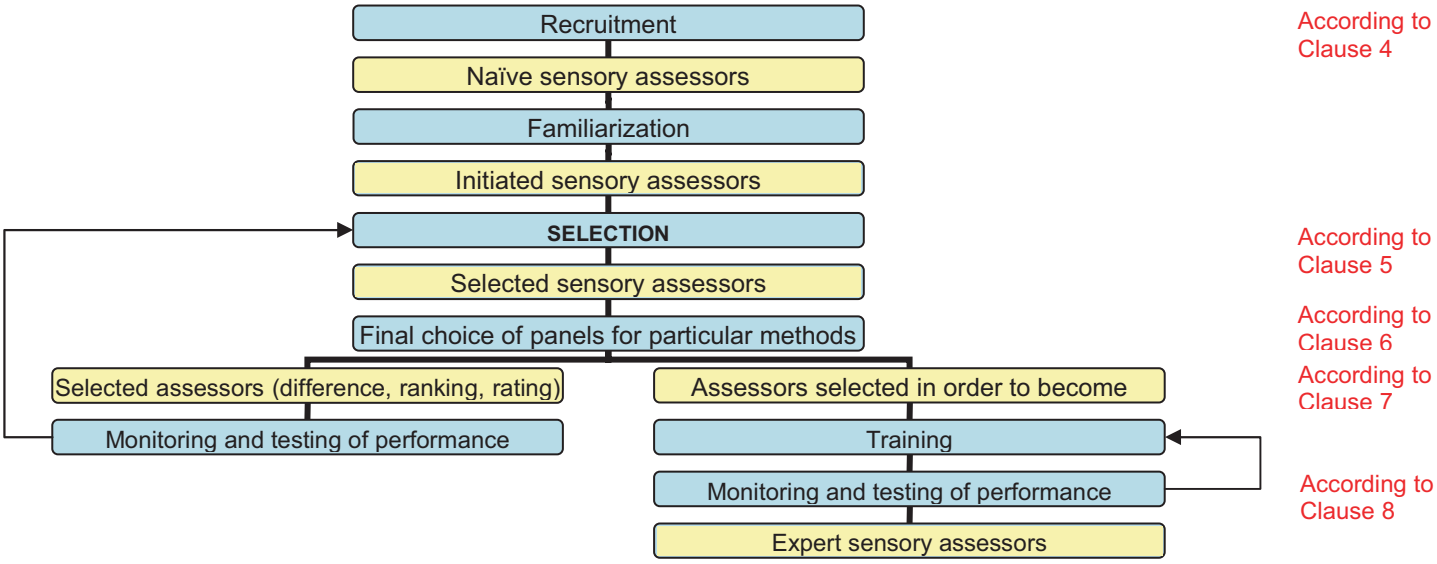

Figure 1 — Entire process

Sensory analysis — General guidelines for the selection, training and monitoring of selected assessors and expert sensory assessors

WARNING — This document does not purport to address all of the safety problems, if any, associated with its use. It is the responsibility of the user to establish appropriate safety and health practices and to ensure compliance with any national regulatory conditions.

1 Scope

This International Standard specifies criteria for the selection and procedures for the training and monitoring of selected assessors and expert sensory assessors. It supplements the information given in ISO 6658.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 5492:2008, *Sensory analysis — Vocabulary*

ISO 6658, *Sensory analysis — Methodology — General guidance*

ISO 8589, *Sensory analysis — General guidance for the design of test rooms*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 5492 and the following apply.

3.1
repeatability
precision under repeatability conditions

Note 1 to entry: Repeatability can be expressed quantitatively in terms of the dispersion characteristics of the results.

[ISO 3534-2:2006, 3.3.5]

Note 2 to entry: Repeatability related to sensory analysis is defined as a measure of the agreement between assessments on the same sample under the same conditions. See Table A.1.

3.2
repeatability conditions
observation conditions where independent test/measurement results are obtained with the same method on identical test/measurement items in the same test facility or measuring facility by the same operator using the same equipment within short intervals of time

Note 1 to entry: Repeatability conditions include:

— the same measurement procedure or test procedure;

— the same operator;

— the same measuring or test equipment used under the same conditions;

— the same location;

— repetition over a short period of time.

[ISO 3534-2:2006, 3.3.6]

Note 2 to entry: Short intervals of time related to sensory analysis are defined as the repetition over the same session.

3.3
reproducibility
precision under reproducibility conditions

Note 1 to entry: Reproducibility can be expressed quantitatively in terms of the dispersion characteristics of the results.

Note 2 to entry: Results are usually understood to be corrected results.

[ISO 3534-2:2006, 3.3.10]

Note 3 to entry: Reproducibility related to sensory analysis is defined as a measure of the agreement between assessments on the same sample under different conditions for assessors and panel. See Table A.1.

3.4
reproducibility conditions
observation conditions where independent test/measurement results are obtained with the same method on identical test/measurement items in different test or measurement facilities with different operators using different equipment

[ISO 3534-2:2006, 3.3.11]

Note 1 to entry: Reproducibility conditions related to sensory analysis may include different times (sessions), different environments and different panels. See Table A.1.

4 Selection of assessors

4.1 General

The following general characteristics are desirable for willing participants for training:

a) they shall be motivated and interested in further developing their sensory skills;

b) they shall be willing to participate.

4.2 Recruitment, preliminary screening and initiation

4.2.1 General

To recruit candidates and to select those most suitable for training as selected assessors, follow 4.2.2 to 4.2.5.

4.2.2 Recruitment

Three questions arise when recruiting persons to form a sensory analysis panel.

— From where should people be sought to constitute the group?

— How many people shall be selected?

— How shall the people be selected?

4.2.3 Types of recruitment

4.2.3.1 General

Two types of recruitment are available to organizations:

— through the personnel department of the organization (internal recruitment);

— recruit people from outside the organization (external recruitment).

It is possible to constitute a mixed panel from both types of recruitment.

4.2.3.2 Internal recruitment

Internal candidates are recruited from their office, plant or laboratory staff. It is advisable to avoid those persons who are too personally involved with products or projects being examined, in particular those involved at the technical or commercial level, because they may cause the results to be biased.

In this type of recruitment, it is vital that the organization's general management and hierarchy provide their support and make it known that sensory analysis is considered as forming part of everyone's work. This can be made known at the hiring stage of the personnel.

4.2.3.3 External recruitment

The recruitment is conducted outside the organization.

The most commonly used means for this purpose are:

— by phone and newspaper (recruitment through classified advertisement in the local press, in specialized publication or in newspapers which are distributed free of charge, etc. — in this case, all types of people can reply and it is necessary to carry out a selection);

— opinion poll organizations — some of these organizations can provide the names and addresses of persons likely to be interested;

— in-house "consumer" files, compiled as a result of advertising campaigns or complaints;

— persons visiting the organization;

— personal acquaintances.

4.2.3.4 Mixed panel

A mixed panel may be formed using internal and external recruitment, in variable proportions.

4.2.4 Advantages and disadvantages of internal and external recruitment

4.2.4.1 General

Organizations may wish to use independent internal or external panels for different tasks.

4.2.4.2 Internal recruitment

4.2.4.2.1 Advantages

The advantages are:

— the people are available;

— it is not necessary to make provision for any payment (however, in order to maintain interest, it may be desirable to offer incentives);

— a better confidentiality *vis-à-vis* the results is ensured, which is particularly important for research work;

— assurance of panellist punctuality.

4.2.4.2.2 Disadvantages

The disadvantages are:

— problems related to the hierarchy of the organization;

— candidates are influenced in their judgements by knowledge of the products;

— it is difficult to allow for the evolution of the organization's products (people are influenced by their changing familiarity with the organization's products);

— replacement of candidates is more difficult (limited number of persons in small organizations);

— less choice of people;

— lack of availability;

— conflict of priorities.

4.2.4.3 External recruitment

4.2.4.3.1 Advantages

The advantages are:

— a wide range of choice;

— a subsequent supply of new persons by word of mouth;

— no problems with hierarchy;

— much easier selection, without the risk of offending people if they are unsuitable;

— easy availability.

4.2.4.3.2 Disadvantages

The disadvantages are:

— the method is expensive (remuneration, paperwork);

— this method is better suited to urban communities where there is a sufficient number of inhabitants; in rural areas, it may not be easy to obtain sufficient suitable people in the area, but advantage can be taken of co-operatives (e.g. milk, wine) — in this case, the risk that some candidates are influenced in their judgment because of their experience shall be taken into account;

— since it is necessary that the individuals be available, a disproportionate number of retired people, unemployed women or students are sometimes encountered because it is more difficult to recruit those in full-time employment;

after having paid for the selection and training, there is a risk that people leave at a moment's notice.

4.2.5 Number of persons to be selected

Experience has shown that, after the recruitment, the selection procedures eliminate approximately half the people for reasons such as gustative sensitivity and material conditions.

The number of persons to be recruited varies depending on the following elements:

— the financial means and the requirements of the organization;

— the types and frequency of tests to be conducted;

— whether or not it is necessary to interpret the results statistically.

It is highly desirable that a panel has at least 10 selected assessors. At least two or three times the number of persons actually required to constitute the final panel should be recruited, e.g. in order to obtain a panel of 10 persons, 40 to 60 persons should be recruited and a minimum of 20 persons should be selected.

For specialized purposes and different types of sensory tests, a higher number of assessors can be required.

4.3 Background information and preselection

4.3.1 General

Background information on the candidates may be obtained by submitting them to a combination of clearly understood questionnaires coupled with interviews by persons experienced in sensory analysis. The aspects specified in 4.3.2 to 4.3.5 shall be explored.

4.3.2 General criteria

4.3.2.1 Availability

Candidates shall be available to attend both training and subsequent assessments. Personnel who travel frequently or who have continual heavy workloads are often unsuited for sensory work.

4.3.2.2 Attitudes to foods

Strong dislikes for certain foods and beverages, in particular those which it is proposed to assess, together with any cultural or other reasons for not consuming certain foods or beverages, should be determined. Candidates who are adventurous in their eating habits often make good assessors for descriptive analyses.

4.3.2.3 Knowledge and aptitude

The initial sensory perceptions of the candidates have to be interpreted and expressed, requiring certain physical and intellectual abilities, in particular the capacity to concentrate and to remain unaffected by external influences. If the candidate is then required to evaluate only one type of product, knowledge of all aspects of that product may be beneficial. It is then possible to choose expert assessors from those candidates who have shown an aptitude for sensory analysis of this product.

4.3.2.4 Ability to communicate

The ability of candidates to communicate and describe the sensations they perceive during an assessment is particularly important when considering candidates for descriptive analyses. This ability can be determined at the interview and again during screening tests.

Additional desirable characteristics of candidates include:

a) a good memory for sensory attributes;

b) the ability to verbalize descriptions of products.

4.3.2.5 Ability to describe

Desirable characteristics of candidates include the ability to:

a) describe products and verbalize sensations;

b) develop a memory for the description of sensory attributes.

4.3.3 Health criteria

The candidates shall be in good general health. They shall not suffer from any disabilities, allergies or illnesses which may affect those senses relevant to the sensory analysis to be undertaken and shall not be taking medication which might impair their sensory capacities and thus affect the reliability of their judgements. It may be useful to know whether the candidates have dental prostheses, since they can have an influence in certain types of evaluation involving texture or flavour.

Colds or temporary conditions (e.g. pregnancy) should not be a reason for eliminating a candidate.

4.3.4 Psychological criteria

4.3.4.1 Interest and motivation

Candidates who are interested in sensory analysis and the product or products to be investigated are likely to be more motivated and hence are likely to become better assessors than those without such interest and motivation.

4.3.4.2 Sense of responsibility and power of concentration

Candidates shall show interest and motivation for the tasks and shall be willing to persevere with tasks demanding prolonged concentration. They shall be punctual in attending sessions and shall be reliable and honest in their approach.

4.3.4.3 Ability to judge

Assessors shall come to a decision, stand by it without any personal preferences, be self-critical and know their limitations.

4.3.4.4 Willingness to co-operate

Assessors should be willing to learn and not be dominant in a group discussion.

4.3.5 Other factors

Other information which may be recorded during recruitment is name, age group, sex, nationality, educational background, current occupation, and experience in sensory analysis. Information on smoking habits may also be recorded, but candidates who smoke shall not be generally excluded.

IMPORTANT — Any files on individual persons shall comply with the legal requirements of the country concerned.

4.4 Screening

4.4.1 General

Various tests which may be used for screening purposes are described in 4.4.2.

The choice of the tests and of the materials to be used is conducted on the basis of the applications envisaged and the properties to be assessed.

4.4.2 Types of screening test

All the tests described have the dual function of familiarizing the candidates with both the methods and the materials used in sensory analysis. They are divided into three types as those aimed at:

a) determining impairment;

b) determining sensory acuity;

c) evaluating a candidate's potential for describing and communicating sensory perceptions.

Test results should only be carried out after previous experience followed by familiarization.

The tests shall be conducted in an appropriate environment in accordance with the recommendations specified in ISO 8589. They shall be followed by interviews. Several tests described in this International Standard are based on those specified in ISO 6658.

The selection of assessors should take into account the intended application, the performance of the candidates at the interviews, and their potential rather than their current performance. Candidates with high success rates are to be expected to be more useful than others, but those showing improving results with repetition are likely to respond well to training.

4.4.3 Colour vision

4.4.3.1 General

Candidates with abnormal colour vision are unsuitable for tasks involving judgement or matching of colours. Assessment of colour vision can be carried out by using an effective test, e.g. the Ishihara test (see Reference [11]) or the Farnsworth Munsell 100 hue test.

4.4.3.2 Reagents and materials

Use only reagents of recognized analytical grade, unless otherwise specified, and distilled or demineralized water or water of equivalent purity.

4.4.3.2.1 Yellow **quinoline yellow** (E 104; CAS No. 8004-92-0; CI 47005).

4.4.3.2.2 Blue **patent blue V** (E 131; CAS No. 3536-49-0; CI 42051).

4.4.3.2.3 Red **carmoisine red** (E 122; CAS No. 3567-69-9; CI 14720).

4.4.3.2.4 Graphite (CAS No. 7782-42-5) and **cornstarch** (CAS No. 9005-25-8).

4.4.3.3 Preparation of stock solutions and mixture

Prepare the test series out of two stock solutions. For the colour test series from yellow to green and blue, put 1 g quinoline yellow into a 500 ml volumetric flask and 0,1 g patent blue V in a 1 000 ml volumetric flask, make up to volume with water.

For the colour test series from red over violet to blue, weigh 1 g carmoisine red into a 1 000 ml volumetric flask and 0,1 g patent blue V into a 1 000 ml volumetric flask and make up to volume with water.

For the test with grey colour from light to dark, make a homogenous mixture from 90 % mass fraction cornstarch (native low water content) and 10 % mass fraction graphite.

4.4.3.4 Preparation of test solutions

For each of test samples 1 to 11, in a 100 ml volumetric flask, mix the volume(s) of stock solution(s), in millilitres, listed in Table 1 with water, and transfer the solutions into a series of test tubes. Close the tubes.

Table 1 — Volumes of stock solution (colour solution) in 100 ml dilutions

Values in millilitres

Colour solution volume	Sample No.										
	1	2	3	4	5	6	7	8	9	10	11
Yellow or red	25	23,5	21,5	19,0	16,5	12,5	7,0	3,5	1,5	0,5	0
Blue	0	1,5	3,5	6,0	8,5	12,5	18,0	21,5	23,5	24,5	25
NOTE Green = yellow + blue; violet = red + blue											

For each of test samples 1 to 10, add the masses of cornstarch and graphite listed in Table 2.

Table 2 — Amount of stock mixture mixed with white cornstarch

Values in grams

Substance	Sample No.									
	1	2	3	4	5	6	7	8	9	10
Cornstarch	19,9	19,7	19,5	19,3	19,1	18,9	18,7	18,5	18,3	18,1
Graphite/cornstarch stock mixture	0,1	0,3	0,5	0,7	0,9	1,1	1,3	1,5	1,7	1,9

4.4.3.5 Procedure

The test tubes are presented in randomized order and have to be sorted from yellow through green to blue or red through violet to blue or from light to dark grey.

4.4.3.6 Minimum requirements

In every test series of 10 to 11 samples two mistakes involving two adjacent samples are allowed.

4.4.4 Ageusia and anosmia

It is desirable that candidates be tested to determine their sensitivity to substances which may be present in small concentrations in products, in order to detect ageusia, anosmia or possible lack of sensitivity (see ISO 3972[2]).

Samples of sapid and/or olfactory materials (see Table 3) at well above threshold levels are prepared. Each sample is attributed a different, random, three-digit code number. Candidates are presented with one sample of each type and are allowed to familiarize themselves with them (see ISO 6658).

They are then presented with a series of the same materials labelled with different random numbers. They are asked to match each of them to one of the original set and to describe the sensation they are experiencing.

Approximately twice as many new samples as original samples shall be presented. None of the samples shall be so intense as to produce strong carry-over effects and hence to influence subsequent tasting. Odourless flavourless water shall be made available for cleansing the palate between samples.

Examples of materials that may be used are given in Table 3. For these substances and concentrations, it is generally accepted that candidates who make fewer than 80 % correct matches should not be chosen as selected assessors. A correct description of the sensations produced by the samples is desirable but less important.

Table 3 — Examples of taste or odour materials and concentrations for screening tests

Taste or odour	Material	Chemical Abstracts Service (CAS) No.	Concentration in water at room temperature g/l	Concentration in ethanol[a] at room temperature g/l
Taste				
Sweet	Sucrose	57-50-1	10 (1 %)	—
Acid	Citric acid	77-92-9	0,3 (0,03 %)	—
Bitter	Caffeine	58-08-2	0,3 (0,03 %)	—
Salty	Sodium chloride	7647-14-5	2 (0,2 %)	—
Umami/broth	Monosodium glutamate or mixture mass fractions (50 % monosodium glutamate, 25 % disodium 5'-guanylate, 25 % disodium 5'-inosinate)	142-47-2 5550-12-9 4691-65-0	0,6 or 0,18	
Astringent	Tannic acid[b] or quercetin or potassium aluminium sulfate (Alum)	1401-55-4 117-39-5 7784-24-9	1 0,5 0,5	—
Metallic	Iron(II) sulfate heptahydrate, $FeSO_4 \cdot 7H_2O$[c]	7782-63-0	0,01	—
Odour				
Lemon, fresh	Citral ($C_{10}H_{16}O$)	5392-40-5	—	1×10^{-3}
Vanilla	Vanillin ($C_8H_8O_3$)	121-33-5	—	1×10^{-3}
Thyme	Thymol ($C_{10}H_{14}O$)	89-83-8	—	5×10^{-4}
Floral, lily of the valley, jasmine	Benzyl acetate ($C_9H_{10}O_2$)	140-11-4	—	1×10^{-3}

[a] Stock solutions are prepared with ethanol, but the final dilution is made with water and shall not contain more than 2 % volume fraction of alcohol.

[b] This material is not very soluble in water.

[c] To avoid the appearance of a yellow coloration due to oxidation, it is necessary to use a freshly prepared solution.

5 Training of assessors

5.1 Principle

To provide assessors with rudimentary knowledge of procedures used in sensory analysis and to develop their ability to detect, recognize, describe and discriminate sensory stimuli. To train assessors to use this expertise so that they may become proficient in the use of such methods with particular products.

5.2 General

A number of assessors one and a half to two times greater than that finally required in the panel shall be trained. To ensure the development of the correct approach to sensory analysis, all training shall be conducted in a suitable environment in accordance with the recommendation given in ISO 6658. It is also useful to train assessors in basic knowledge of the products they assess, e.g. by giving information on the manufacturing process or by organizing visits to plants.

Assessors shall be instructed and trained to be objective and to disregard their likes and dislikes.

Results shall be discussed and assessors shall be given the opportunity to reassess samples and to check their replies where disagreement exists.

Assessors shall be instructed not to use perfumed products prior to or during sessions. They shall also be asked to avoid contact with tobacco or with intensive tastes or odours for at least 60 min prior to such sessions. Soap used for washing shall not leave any odour on the hands.

It shall be emphasized to assessors that if they carry any odour into the test room, tests may be invalidated.

5.3 Assessment procedure

At the start of any training programme, assessors shall be taught the correct way to assess samples. In all assessments, instructions shall be read thoroughly before any task and adhered to throughout the analysis. The temperature of samples shall be specified. Unless told to concentrate on specific attributes, the assessors should usually examine attributes in the following order:

— appearance (mainly colour);

— odour;

— texture;

— flavour (comprising aroma and taste);

— after-taste.

When assessing odour, assessors shall be taught to take short rather than long sniffs and not to sniff too many times lest they become confused and fatigued.

With both liquid and solid samples, the assessors shall be told in advance the procedure to be followed. The problem of adaptation and the advantages of using a palate cleanser and of standard time intervals between samples shall also be discussed. Any procedure finally agreed upon shall be stated clearly so that all assessors assess products in the same way. The interval between samples shall be sufficient to permit recovery, but not so long that assessors lose their ability to discriminate.

5.4 Training of colour, taste, odour and texture

5.4.1 Tests for detection of a stimulus

These tests are based on the triangle test according to ISO 4120.[3]

One material at a time is tested. Two samples of the test material and one sample of water or other neutral medium, or one sample of the test material and two of water or other neutral medium, are presented to each candidate. The concentration of the test material shall be at the supra-threshold level.

The test materials, their concentrations and the neutral medium (if used) shall be chosen by the organizer in relation to the types of assessment for which the candidates are to be used. Preferably candidates should have 100 % correct responses.

An inability to detect differences after several repetitions indicates unsuitability for this type of test.

Examples of materials which may be used in detection tests are described in Table 4. In case of repeated training the concentrations may be reduced.

Table 4 — Examples of materials which may be used in detection tests

Material	CAS No.	Taste	Mass concentration or volume fraction in water at room temperature
Caffeine	58-08-2	bitter	0,2 g/l
Citric acid	77-92-9	sour	0,2 g/l
Sodium chloride	7647-14-5	salty	1,3 g/l
Sucrose	57-50-1	sweet	6 g/l
Monosodium glutamate	142-47-2	umami	0,3 g/l
Iron(II) sulfate heptahydrate	7782-63-0	metallic	0,005 g/l
(*Z*)-Hex-3-en-1-ol	928-96-1	Green, grassy, unripe	0,4 ml/l

5.4.2 Tests for discrimination between levels of intensity of a stimulus

These tests are based on the ranking test according to ISO 8587.[7] The tests are carried out using stimuli for taste, odour (only for very small concentrations), texture (mouth and hand), and colour.

For each test, four samples having different intensities of the property are presented in a random order to the candidates, who are required to put them in order of increasing intensity. This random order shall be the same for all candidates, to ensure that comparisons of their performance are not influenced by the effects of different orders of presentation.

A satisfactory level of success in this task can be specified only in relation to the particular intensities used.

Examples of products that may be used are given in Table 5; for these concentrations, candidates who invert the order of more than one adjacent pair of samples shall be considered unsuitable as selected assessors for this type of analysis.

Table 5 — Examples of products which may be used in discrimination tests

Test	Product[a]	Description	Mass concentration/ mass fraction in water at room temperature
Taste discrimination Paired comparison or ranking	Citric acid	Sour	0,1 g/l; 0,2 g/l; 0,3 g/l; 0,5 g/l
Odour discrimination Paired comparison or ranking	Isoamyl acetate	Fruity	5 mg/l; 10 mg/l; 20 mg/l; 40 mg/l; ethanol dilution)
Texture discrimination	To suit the industry concerned (e.g. cream cheese, purée, gelatine)	Creamy, hard, viscous, etc.	—
Colour discrimination	Cloth, colour scales, etc.	Red, green, etc.	Intensity of a colour ranging, for example, from dark red to light red

[a] Other appropriate products showing a graduation in characteristics may also be used.

5.4.3 Descriptive ability

These tests are aimed at determining a candidate's ability to describe sensory perceptions. Two tests are advocated, one covering odour stimuli and the other textural stimuli. The tests are conducted as combined assessments and interviews.

5.4.4 Odour description test

Candidates are presented with between five and 10 olfactory stimuli, preferably related to the product or products intended to be evaluated. The set shall contain some samples which are easy to recognize and others which are less common. The intensity shall be well above the recognition threshold, but not greatly above the levels that might be encountered in the products ultimately of interest.

Several methods of sample preparation exist which are either direct or retronasal in nature.

In direct methods, flasks, smelling strips or capsules containing odours are employed.

In retronasal methods, the odours may be evaluated by ingestion of aqueous solutions.

The method most commonly used is still that of the evaluation of odours from flasks. This method is described as follows:

— Samples are absorbed in odourless paraffin wax or odourless cotton wool which is placed in odourless flasks which do not a visual recognition of the colour and can be capped. Sufficient material shall be allowed to evaporate into the headspace of the flasks and the intensity shall be checked before presentation of the flasks to candidates.

Samples may also be presented on smelling strips or pads.

— Samples are presented one at a time, and the candidate is asked to describe or record what is perceived. Following the initial reaction, the organizer may, if desired, discuss the sample in order to bring forth further comments and to explore more fully the candidate's capability of discussing the stimuli.

Candidates are graded according to performance on a scale such as the following:

— 3 points for a correct identification or a description of the most frequent association;

— 2 points for a description in general terms;

— 1 point for an identification or description of an appropriate association following discussion;

— 0 points for no response or a totally wrong response.

A satisfactory level of success in this task can be specified only in relation to the materials used. Examples of olfactory materials which may be used are given in Table 6. See also ISO 5496.[5]

5.4.5 Texture description test

Candidates are provided with a series of products in random order and are asked to describe their textural characteristics.

Solid sample products should be presented as uniformly sized blocks and liquid sample products shall be presented in vessels hiding any possible differences.

Candidates are graded according to performance on a scale such as the following:

— 3 points for a correct identification or a description of the most frequent association;

— 2 points for a description in general terms;

— 1 point for an identification or description of an appropriate association following discussion;

— 0 points for no response or a totally wrong response.

A satisfactory level of success in this task can be specified only in relation to the products used. Examples of products that may be used are given in Table 7.

Table 6 — Examples of olfactory materials for odour description test

Material	CAS No.	Name most commonly associated with the odour
Benzaldehyde	100-52-7	Bitter almonds, cherry
Octen-3-ol	3391-86-4	Mushroom
(*Z*)-Hex-3-en-1-ol	928-96-1	Fresh grass
(*S*)-(+)-Carvone	2244-16-8	Caraway
γ-Nonalactone	104-61-0	Coconut
Diacetyl	431-03-8	Butter
Cinnamaldehyde	104-55-2	Cinnamon
Phenyl acetate	122-79-2	Floral
Diallyl sulfide	2179-57-9	Garlic
Camphor	76-22-2	Camphor, medicine
Menthol	1490-04-6	Peppermint
Eugenol	97-53-0	Clove
Anethol	104-46-1	Aniseed
Vanillin	121-33-5	Vanilla
β-Ionone	79-77-6	Violets, raspberries
Butyric acid	107-92-6	Rancid butter
Acetic acid	64-19-7	Vinegar
Isoamyl acetate	123-92-2	Fruit, acid drops, banana, pear
Dimethylthiophene	638-02-8	Grilled onions
NOTE It is possible to use food products, spices, extracts, infusions or chemical odorants. Materials chosen shall be suited to local needs and shall be free from other odorous materials (ISO 3972[2]).		

Table 7 — Examples of products for texture description test

Product	Texture most commonly associated with the product
Oranges	Juicy, cellular particles
Breakfast cereals (cornflakes)	Crispy, crunchy
Pears (Passe Crassane)	Gritty, juicy
Granulated sugar	Crystalline, coarse
Marshmallow topping	Sticky, malleable
Chestnut purée	Pasty
Semolina	Grainy
Double cream	Unctuous, creamy
Edible gelatine	Gummy
Corn muffin cake	Crumbly
Cream toffee	Tacky
Calamari (squid)	Elastic, springy, rubbery
Celery	Fibrous
Raw carrots	Crunchy, hard

5.4.6 Recognition of differences in texture

5.4.6.1 Method: Ranking test according to ISO 8587.[7]

5.4.6.2 Test samples: gelatine at different firmness levels (see Table 8).

Substance: gelatine Type A, 240 bloom, particle size 0,5 mm; saccharose; demineralized water; red food colour, mass concentration 5 g/500 ml).

Table 8 — Composition of the gelatine samples

Sample No.	Demineralized water ml	Saccharose g	Food colour g	Gelatine g
1	500	60	1,20	20,5
2	500	60	1,25	25,0
3	500	60	1,30	27,5
4	500	60	1,40	32,5
5	500	60	1,45	42,5
6	500	60	1,55	57,5
7	500	60	1,60	67,5
8	500	60	1,70	82,5
9	500	60	1,75	95,5
10	500	60	1,85	100,0

Preparation: Weigh all ingredients into a pot, to achieve the desired mass which then should be recorded and heat the content to 60 °C. Add water to return to the mass recorded before heating. Fill the solution into small containers of similar size and shape. Keep at room temperature for 4 h and in refrigerator for 24 h to harden the gelatine.

Application: Each assessor gets all samples in a randomized order and tests each encoded sample by touching and reorders the samples by firmness. At least 80 % of the samples shall be ordered correctly.

5.5 Training in detection and recognition of special tastes and odours

Matching, recognition, paired comparison, triangle and duo–trio tests (see ISO 6658 and specific International Standards) shall be used to demonstrate differences in tastes at high and low concentrations and to train assessors to recognize and describe them correctly (see ISO 3972[2]). Identical tests shall be used to develop assessors' acuity for odour stimuli (see ISO 5496[5]). Stimuli shall initially be presented singly as an aqueous solution, but as experience is gained the base may be replaced by actual foods or beverages. Mixed samples in which the proportions of two or more components vary may also be introduced.

Alteration of the appearance of a sample (e.g. by the use of coloured light) is particularly useful in demonstrating the need to be objective when trying to detect differences in other sensory characteristics.

Samples used for training and testing shall be characteristic of their origin, style and quality, and representative of the range generally found on the market.

Samples shall be presented in the quantity and at the temperatures generally met with in the trade or in use.

Exceptions may be made when demonstrating excellence, imperfections or faults.

Care shall be taken to ensure that sensory fatigue does not arise owing to the testing of an excessive number of samples.

Table 9 gives examples of materials which may be used during this phase of training. If possible, stimuli should be chosen to relate to the material or materials it is intended to assess.

Table 9 — Examples of materials to be used during training in detection and recognition

Example	Material	Description
1	Table 2 materials	
2	Table 4 products	
3	Saccharin (100 mg/l)	Sweet
4	Quinine sulfate (0,20 g/l)	Bitter
5	Grapefruit juice	Bitter, astringent, grapefruit
6	Apple juice	Sweet, fruity, apple
7	Sloe juice	Sloe, bitter, astringent
8	Cold tea	Tea
9	Sucrose (10 g/l; 5 g/l; 1 g/l; 0,1 g/l)	Sweet
10	(*Z*)-Hex-3-en-1-ol (CAS-No. see Table 6) (15 mg/l)	Grassy
11	Benzyl acetate (10 mg/l)	Almond
12	Items 4 to 7 with varying sucrose contents (see item 9)	Intensity of sweetness
13	Tartaric acid (0,3 g/l) plus hexanol (30 mg/l), tartaric acid (0,7 g/l) plus hexanol (15 mg/l)	Intensity of astringency
14	Orange-flavoured drink coloured yellow; orange-flavoured drink coloured orange: lemon-flavoured drink coloured yellow	Orange and lemon
15	Succession of caffeine (0,8 g/l), tartaric acid (0,4 g/l), and sucrose (5 g/l)	Bitter, astringent, sweet
16	Succession of caffeine (0,8 g/l), sucrose (5 g/l), caffeine (1,6 g/l), and sucrose (1,5 g/l)	Bitter, sweet, bitter, sweet

5.6 Training in the use of scales

Assessors shall be introduced to the concepts of rating and/or classification and/or interval and/or ratio scales (see ISO 6658 and ISO 4121[4]) depending on the scales to be used in the future. The various rating procedures are then used to attach meaningful magnitudes to the samples. As indicated in 5.4, the base shall initially be water, but actual foods and beverages, with mixed stimuli, both of which may vary independently, may then be introduced.

Table 10 gives examples of materials which may be used during this phase of training.

If possible, stimuli shall be chosen to relate to the product or products it is intended to assess.

Table 10 — Examples of materials which may be used for training in the use of scales

Example	Material				
1	Table 4 products and Table 9, example 9				
2	Caffeine	0,15 g/l	0,22 g/l	0,34 g/l	0,51 g/l
3	Tartaric acid	0,05 g/l	0,15 g/l	0,4 g/l	0,7 g/l
4	Hexyl acetate	0,5 mg/l	5 mg/l	20 mg/l	50 mg/l
5	Cheeses, e.g. mature hard cheese such as Cheddar or Gruyère, ripened soft cheese such as Camembert				
6	Pectin gels				
7	Lemon juice and diluted lemon juice		10 ml/l	50 ml/l	

Intensity training of basic tastes, astringency, unripe, texture, hardness of cheese and jellies, lemon taste.

5.7 Training in the development and use of descriptors (profiles)

Assessors shall be introduced to the idea of profiling by being presented with a series of simple products and asked to develop vocabularies for describing their sensory characteristics, in particular terms which allow samples to be differentiated. Terms shall be developed individually and then discussed and an agreed list of at least 10 devised. This list shall then be used to produce profiles of the products, first by attributing the terms appropriate to each sample and then by scoring their intensities using the various types of scales discussed in 5.6. The organizer shall produce profiles of the products using the results to illustrate the value of descriptive analysis. Examples of products which may be used in this training exercise could be commercial fruit juice and blends, breads, cheeses, fruit and vegetables.

Other products may be selected to suit the application.

5.8 Practice

The formal training sessions outlined in 5.4 to 5.6 shall be interspersed with exercises to provide assessors with further experience.

5.9 Specific product training

After basic training, assessors may undergo a period of product training, the exact nature of this depending on whether it is intended to use the panel for difference or descriptive testing (visual, odour, textural and flavour evaluations).

5.9.1 Difference assessment

Samples similar to those that are intended to be assessed are presented to the assessors who evaluate them using one of the difference assessment procedures. (See ISO 6658 and the International Standards describing individual difference tests).

5.9.2 Descriptive assessment

For descriptive analyses which are not intended for a specific product, experience should be gained with a wide range of different products. The number of samples for evaluation during the training depends closely on the variability of the products which are to be assessed by the panel. For assessors who are to assess one specific product type, several samples of this type of product should be presented.

EXAMPLE During training, 10 to 15 samples of a type of a product can be presented.

Descriptors are proposed to describe the various organoleptic characteristics.

The organizer then leads a discussion to help the panel to put similar descriptors into groups and to rationalize the vocabulary by selecting a single descriptor to replace each group of terms. The process is assisted by examining external standards and samples with particular characteristics.

The agreed descriptors are then incorporated into a score sheet. Several further samples are examined and the terminology is further improved. The meaning of intensity scales for each attribute shall be discussed and rationalized by reference to actual samples.

6 Final choice of panels for particular methods

6.1 Principle

Those assessors most appropriate for a given method are chosen to make up pools from which panels of assessors for particular tests can be assembled.

The number of assessors required for each purpose shall be at least the size recommended in ISO 6658. Candidates selected as appropriate for one purpose are not necessarily appropriate for another, and candidates excluded for one purpose are not necessarily unsuitable for others.

6.2 Difference assessments

Final panel selection is based on repeat examination of actual samples. If the panel is to be used for the detection of a particular characteristic, the ability to detect adulterated samples at decreasing concentrations can also be used as a criterion for selection. Assessors selected shall perform consistently and be able to differentiate correctly the samples presented.

6.3 Ranking assessment

Final panel selection is based on repeat examination of actual samples. Assessors selected shall perform consistently and be able to rank correctly the samples presented. See also ISO 8587.[7]

6.4 Rating and scoring

Assessors shall assess approximately six different samples in triplicate, presented in random order, and, if possible, at more than one session. The results should be tabulated as shown in Tables B.1 and B.2.

The data should be analysed by analysis of variance (ANOVA, as shown in Tables B.1 and B.2, Friedman or Page test) to examine the individual results of each assessor.

Assessors who have a high residual standard deviation, indicating inconsistency, or for whom the variation between the samples is not significant, indicating poor discrimination, should be considered for rejection. However, if most of the assessors are poor in one or both of these respects, it may be because the samples are not sufficiently different to be reliably distinguished.

The combined data should also be analysed by ANOVA as shown in Tables B.2 and B.3. The statistical significance of the variation between assessors, the variation between samples, and the assessors–samples interaction should be determined.

Significant variation between assessors indicates the presence of bias, i.e. one or more assessors give(s) scores consistently higher or lower than the others. Significant variation between samples indicates that the assessors as a panel successfully differentiate between the samples. A significant assessors–samples interaction indicates that two or more of the assessors have a different perception of the dissimilarities between two or more samples. In some cases, an assessors–samples interaction may even reflect a disagreement about the ranking of the samples.

Although ANOVA is appropriate for scoring, it is not appropriate for some forms of rating. If, for instance, a ranking procedure is used, then non-parametric methods like the Friedman test might be more appropriate (see ISO 8587[7]).

6.5 Qualitative descriptive analysis

No additional specific selection procedure is advocated among those already outlined.

Assessors are chosen on the basis of their performance in the various exercises, in particular those specified in 5.6 and 5.9.2.

6.6 Quantitative descriptive analysis

6.6.1 Assessors are chosen for their skills and their aptitude to follow an intensive training in order to become expert sensory assessors.

Panel leaders should evaluate the performance of the selected assessors, over a period of time and on the products concerned. Any selected assessors who show good repeatability, good acuity or particular

aptitude regarding specific attributes (e.g. a taint) of classes of materials, should be considered for use on panels of expert assessors.

Moreover desirable characteristics of candidates include:

a) a memory for sensory attributes;

b) an ability to communicate with other assessors;

c) an ability to verbalize descriptions of products.

The extent to which selected assessors possess these characteristics varies. So a complementary selection could be performed or the training programme adjusted accordingly.

6.6.2 If controls or reference samples (see ISO 5492) have been provided, candidates shall be tested for their ability to recognize and describe them.

6.6.3 Assessors shall assess approximately six samples using the vocabulary and score sheet developed as specified in 5.9.2. The samples shall be presented in triplicate in a properly balanced order. Each descriptor for each assessor shall then be subjected to analyses similar to those specified in 6.4 and described in Annex B or to other multidimensional methods of analysis (ISO 13299[8]).

6.7 Assessors for particular assessments

Despite being selected as the most suitable candidates, selected assessors may fluctuate in their performance during training. With descriptive analysis, it can often prove an advantage to select the better performers or to divide the assessors into subgroups following a programme of assessments and before additional training or any complex statistical examination of the data. For this purpose, the procedures used are the same as those specified in 6.4.

7 Training of experts

7.1 General

One of the aims is to optimize the technical knowledge of the selected assessors by training and development of their sensory potential to become expert sensory assessors. Assessors should possess knowledge of the physiology of taste and smell.

Training is aimed at optimizing the sensory knowledge of the assessors and especially at enabling them to memorize the descriptors of the sensory profile and their intensities, as well as acquiring the required qualities for producing sensory profiles (repeatability, trueness, discriminatory ability).

7.2 Sensory memory

Expert sensory assessors need an above-average ability to keep sensory impressions in mind. The tests used to train a selected assessor rely largely on short-term sensory memory, whereas long-term sensory memory is essential for the expert sensory assessor. Characteristics noted in a current assessment may need to be related to experience of earlier assessments.

NOTE The attributes assessed during a test are naturally related to the experiences acquired during training.

The tests used for training expert sensory assessors are aimed more at exploiting long-term memory.

7.3 Semantic and metric learning of sensory descriptors

Training normally comprises two phases:

— the generation, the definition and the recognition of each descriptor, the objective of which is to identify the words which enable the description of the product or the object (either by means of an existing list or via descriptors generated by the group) and to associate them with the corresponding sensory perceptions, to define each of the descriptors on the basis of these sensory perceptions and to learn to identify its presence or absence in the product or object;

— the assessment of the intensity and the memorization of the scale, the objective of which is to learn to evaluate the intensity of each descriptor and to memorize intensity levels for each of the selected descriptors.

NOTE 1 Training can call upon group facilitation techniques by alternating both individual and collective exercises. This approach requires the assessors to make a great effort to concentrate and memorize under the guidance of the panel leader or facilitator.

Training may initially consist of assessing samples having as a descriptor more or less pronounced intensities and in producing a classification based on this descriptor. Subsequently, the assessors learn to express the intensities in the form of notes by means of references or products or materials having different levels of intensity for a given descriptor.

NOTE 2 Discriminatory and matching tests can be used to highlight the different characteristics of the products and materials or to check whether the characteristics have been memorized.

7.4 Building up a thesaurus of descriptors

Trainees need to understand the role of sensory descriptors as an aid to developing long-term sensory memory, and also as a means of communicating with clients and other experts.

It is necessary for them to acquire knowledge and command of the specific terminology used.

7.5 Training on assessment conditions

The trainee needs to learn to assess large numbers of samples at a single occasion. The trainee also needs to learn to assess a wide range of samples of a product.

8 Monitoring and testing of performance of selected assessors and expert sensory assessors

8.1 Objectives

The objectives of the monitoring of the assessors' performances are to check that their assessments are:

— repeatable;

— discriminatory;

— homogeneous;

— reproducible.

8.2 Principle

The principles of monitoring performance are based on:

— participation in various sensory tests (depending on their specificities) for expert sensorial assessors;

— production of product or material profiles with one or more inter- or intrasession repetitions for expert sensory assessors;

— participation in interlaboratory tests according to ISO 5725[6] within the same sector of activity (suppliers or subcontractors working on the profiles of the same products).

8.3 Analysis of the results

8.3.1 General

The analysis of the results obtained allows the assessment of the performance of the group as a whole as well as the individual performances of the assessors (see Annex A).

8.3.2 Assessment of the performance of the group as a whole

Different methods can be used, e.g. ANOVA:

— single factor ANOVA (products) in order to assess the discriminatory ability;

— three factor ANOVA (products, assessors, sessions) as well as superposing of the profiles of the two or three sessions in order to check reproducibility by studying the session factor and the session–product interaction;

— three factor ANOVA (products, assessors session)) in order to ensure homogeneity by studying the product–assessor interaction.

Other statistical techniques, e.g. principal component analysis (PCA), discriminant factor analysis (DFA) generalized Procrustes analysis, calculation of the RV coefficients (coefficient enabling the assessment of the degree of similarity of two matrices), allow the agreement of the assessors with one another and with respect to the group conclusion to be examined.

8.3.3 Assessment of the individual performance

The data may be represented graphically, or statistical tests may be carried out, e.g:

— comparison of the scorings of each individual with respect to the mean of the group;

— visual representation of the magnitude of the standard deviations;

— homogeneity of the scoring with respect to the group (overscoring, underscoring);

— assessment of the differentiation of the products;

— individual repeatability or reproducibility.

9 Management and follow-up of the group

9.1 Motivation

It is important to maintain the group's motivation:

— by providing information concerning the exploitation of the results;

NOTE 1 Be careful not to bias future work.

— by providing feedback concerning the individual results;

— by a reward.

NOTE 2 "All effort merits a reward."

9.2 Maintaining of skills

So that the group functions efficiently and does not lose the benefit of its training, it shall be called upon on a regular basis. Weekly participation is desirable, and monthly participation shall be a minimum.

It is highly recommended that selective verifications of the group's performances be conducted, approximately twice a year.

Moreover, it may be necessary to retrain the assessors after long periods of interruption (>6 weeks).

Ideally, the group should be benchmarked with respect to other groups by participating in intercomparison studies:

— participation in interlaboratory tests;

— comparison with respect to suppliers or subcontractors working on the same products.

9.3 Renewal

Account being taken of the unavoidable departures of certain group members (moving house, illness, etc.), it may be necessary to recruit new people.

Specific training shall therefore be envisaged in order to bring the new assessors up to a satisfactory level of performance.

The process of integration into the group may be progressive, taking into account the ability of the new assessors to give reliable responses.

9.4 Retraining

If the nature of the products or materials changes, new training sessions shall be implemented in order to take account of possible new descriptors or of the modification of the intensity scales (see ISO 13299[8]).

Annex A
(informative)

Repeatability and reproducibility of assessors and panels

Table A.1 — Repeatability and reproducibility

Parameter	Definition	Determined by (see ANOVA principles in Reference [10])
Repeatability	A measure of the agreement between assessments on the same sample under the same conditions, i.e: — same assessors (panel); — same time (session); — same environment.	**Assessor**: — Standard deviation (SD) of that assessor's scores on replicate samples within a session — Error SD from one-way ANOVA of panel means **Panel**: — SD of panel means of replicate scores within a session — Error SD from one-way ANOVA of panel means (pooled over sessions)
Reproducibility	A measure of the agreement between assessments on the same sample under different conditions, i.e: For an assessor: — same assessor — different times (sessions) — different environments For a panel: — same panel — different times (sessions) — different environments	**Assessor**: Combination of within sessions SD and between sessions SD from ANOVA **Panel**: Combination of within sessions SD and between sessions SD from ANOVA
Reproducibility between panels (trained, for example, according to different methods) or between assessors	**Between panels:** A measure of the agreement between assessments on the same sample under different conditions, i.e: — different panels — different times (sessions) — different environments **Between assessors:** Consistency of asssessors within sessions: Agreement between assessments made by different assessors in a panel on the same sample	Consistency of SD between sessions and SD between panels Between assessors SD from two-way ANOVA of scores in one session

Annex B
(informative)

Use of analysis of variance in the choice of selected assessors for scoring

Table B.1 lists the results of the assessors.

In Table B.1, Y_{ijk} is the score given by the *j*th assessor to the *k*th replicate of sample *i*, and there are *p* samples, *q* assessors and *r* replicates.

For the more specific case of the final choice of panels for scoring and rating (see 6.4), $p = 6$ and $r = 3$. In this case, Table B.2 tabulates the ANOVA for the *j*th assessor.

Table B.1 — Results of the assessors

Sample	Assessor								Mean
	1		2		*j*		*q*		
	Score	Mean	Score	Mean	Score	Mean	Score	Mean	
1									
2									
i					Y_{ij1} Y_{ijk} Y_{ijr}	$\bar{Y}_{ij\cdot}$			$\bar{Y}_{i\cdot\cdot}$
p									
Mean					$\bar{Y}_{\cdot j\cdot}$				$\bar{Y}_{\cdots}$

Table B.2 — Analysis of variance per assessor — Data not combined

Source of variation	Degrees of freedom *v*	Sum of squares *S*	Mean square *MS*	*F*
Between samples	$v_1 = p - 1$	$S_1 = r \sum_{i=1}^{p} \left(\bar{Y}_{ij\cdot} - \bar{Y}_{\cdot j\cdot}\right)^2$	$MS_1 = S_1/v_1$	$F = MS_1/MS_2$
Residual	$v_2 = p(r - 1)$	$S_2 = \sum_{i=1}^{p} \sum_{k=1}^{r} \left(Y_{ijk} - \bar{Y}_{ij\cdot}\right)^2$	$MS_2 = S_2/v_2$	
Total	$v_3 = pr - 1$	$S_3 = \sum_{i=1}^{p} \sum_{k=1}^{r} \left(Y_{ijk} - \bar{Y}_{\cdot j\cdot}\right)^2$		

In Table B.2, the mean for sample *i* is given by

$$\bar{Y}_{ij\cdot} = \frac{\sum_{k=1}^{r} Y_{ijk}}{r} \tag{B.1}$$

and the overall mean is given by

$$\bar{Y}_{\cdot j\cdot} = \frac{\sum_{i=1}^{p}\sum_{k=1}^{r} Y_{ijk}}{pr} \tag{B.2}$$

The residual standard deviation is calculated as follows:

$$\sqrt{MS_2} \tag{B.3}$$

For the combined data, the ANOVA table is constructed as shown in Table B.3.

In Table B.3, the mean for sample *i* is given by

$$\bar{Y}_{i\cdot\cdot} = \frac{\sum_{j=1}^{q}\sum_{k=1}^{r} Y_{ijk}}{qr} \tag{B.4}$$

and the mean for assessor *j* is given by

$$\bar{Y}_{\cdot j\cdot} = \frac{\sum_{i=1}^{p}\sum_{k=1}^{r} Y_{ijk}}{pr} \tag{B.5}$$

The mean of scores given by assessor *j* to sample *i* is

$$\bar{Y}_{ij\cdot} = \frac{\sum_{k=1}^{r} Y_{ijk}}{r} \tag{B.6}$$

and the overall mean is

$$\bar{Y}_{\cdots} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r} Y_{ijk}}{pqr} \tag{B.7}$$

The statistical significance of interaction of assessors and samples is determined by testing the ratio MS_6/MS_7 against critical values in tables of the *F*-distribution with v_6 and v_7 degrees of freedom.

The statistical significance of the variation between assessors is determined by comparing the ratio MS_5/MS_7 (the factors samples and assessors are considered as fixed factors) with critical values of the *F*-distribution with v_5 and v_7 degrees of freedom.

Table B.3 — Analysis of variance for all assessors and all products — Combined data

Source of variation	Degrees of freedom, v	Sum of squares, S	Mean square, MS
Between samples	$v_4 = p - 1$	$S_4 = qr \sum_{i=1}^{p} \left(\bar{Y}_{i..} - \bar{Y}_{...}\right)^2$	$MS_4 = S_4/v_4$
Between assessors	$v_5 = q - 1$	$S_5 = pr \sum_{i=1}^{q} \left(\bar{Y}_{.j.} - \bar{Y}_{...}\right)^2$	$MS_5 = S_5/v_5$
Interaction	$v_6 = (p - 1)(q - 1)$	$S_6 = r \sum_{i=1}^{p} \sum_{j=1}^{q} \left(\bar{Y}_{ij.} - \bar{Y}_{...}\right)^2 - S_4 - S_5$	$MS_6 = S_6/v_6$
Residual	$v_7 = pq(r - 1)$	$S_7 = \sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{k=1}^{r} \left(\bar{Y}_{ijk} - \bar{Y}_{ij\cdot}\right)^2$	$MS_7 = S_7/v_7$
Total	$v_8 = pqr - 1$	$S_8 = \sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{k=1}^{r} \left(Y_{ijk} - \bar{Y}_{...}\right)^2$	

Annex C
(informative)

Example of practical application

An assessment by each assessor of three samples from six batches of fish stored in ice for different lengths of time gave the results shown in Table C.1 using a 10 point scoring system (individual scores and means).

The ANOVA table is then constructed as shown in Table C.2.

The overall ANOVA is then calculated as shown in Table C.3.

It would be concluded that assessors 1 and 4, having low residual standard deviations and statistically significant variation between the samples, were suitable. Assessor 2, having very high residual standard deviation and consequently no significant variation between the samples, would not be suitable, nor would assessor 3, who had no significant variation between the samples.

Variation between assessors is significant, and it can be seen that assessors 2 and 3 give lower scores than do assessors 1 and 4. On the other hand, the assessors–samples interaction is not significant, and it is not possible to assert that the assessors have disagreements about the ranking of the samples.

Table C.1 — Assessor's scores

Sample	Assessor								Mean
	1		2		3		4		
	Score	Mean	Score	Mean	Score	Mean	Score	Mean	
1	8		5		6		9		
	8	8,3	8	7,3	7	6,0	8	8,3	7,5
	9		9		5		8		
2	6		6		5		7		
	8	7,0	7	5,7	4	5,3	7	6,7	6,2
	7		4		7		6		
3	4		5		4		5		
	5	4,7	2	3,3	3	4,0	5	5,0	4,2
	5		3		5		5		
4	6		6		4		6		
	6	5,7	4	5,3	2	3,3	5	5,3	4,9
	5		6		4		5		
5	4		3		4		4		
	5	4,0	2	3,0	4	4,3	5	4,3	3,9
	3		4		5		4		
6	5		4		5		7		
	6	5,7	2	4,3	4	5,0	5	6,3	5,3
	6		7		6		7		
Mean	5,9		4,8		4,7		6,0		5,4

Table C.2 — Analysis of variance — Data not combined

Source of variation	Degrees of freedom, v	Assessor							
		1		2		3		4	
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Between samples	$v = 5$	7,42	13,36[a]	7,83	2,66[b]	2,80	2,40[b]	6,13	13,80[a]
Residual	$v = 12$	0,56		2,94		1,17		0,44	
	Residual standard deviation	0,75		1,71		1,08		0,67	

[a] Significant at the level $\alpha = 0{,}001$.

[b] Not significant at the level $\alpha = 0{,}05$.

Table C.3 — Analysis of variance — Combined data

Source of variation	Degrees of freedom v	Sum of squares S	Mean squares MS	F
Between assessors	$v = 3$	26,04	8,68	6,79[a]
Between samples	$v = 5$	104,90	20,98	16,42[a]
Interaction	$v = 15$	16,04	1,07	0,84[b]
Residual	$v = 48$	61,33	1,28	
Total	71	208,31		

[a] Significant at the level $\alpha = 0{,}001$.

[b] Not significant at the level $\alpha = 0{,}05$.

The statistical significance of interaction of assessors and samples is determined by testing the ratio MS_6/MS_7 against critical values in tables of the F-distribution with v_6 and v_7 degrees of freedom.

The statistical significance of the variation between products is determined by comparing the ratio MS_5/MS_7 (the factors samples and assessors are considered as fixed factors) with critical values of the F-distribution with v_5 and v_7 degrees of freedom.

Bibliography

[1] ISO 3534-2:2006, *Statistics — Vocabulary and symbols — Part 2: Applied statistics*

[2] ISO 3972, *Sensory analysis — Methodology — Method of investigating sensitivity of taste*

[3] ISO 4120, *Sensory analysis — Methodology — Triangle test*

[4] ISO 4121, *Sensory analysis — Methodology — Evaluation of food products by methods using scales*

[5] ISO 5496, *Sensory analysis — Methodology — Initiation and training of assessors in the detection and recognition of odours*

[6] ISO 5725 (all parts), *Accuracy (trueness and precision) of measurement methods and results*

[7] ISO 8587, *Sensory analysis — Methodology — Ranking*

[8] ISO 13299, *Sensory analysis — Methodology — General guidance for establishing a sensory profile*

[9] ISO 22935-1, *Milk and milk products — Sensory analysis — Part 1: General guidance for the recruitment, selection, training and monitoring of assessors*

[10] Lea P., Naes T., Rødbotten M. *Analysis of variance for sensory data*. Chichester: Wiley, 1997, 102 p.

[11] Ishihara S. *Tests for colour blindness.* Tōkyō: Kanehara, 1994. 38 plates

[12] Busch-Stockfisch M.ed. *Praxishandbuch Sensorik in der Produktentwicklung und Qualitätssicherung*. Hamburg: Behr, 2003–2010. [Practical handbook: sensory analysis in product development and quality assurance]

[13] Höhl K., Schönberger G.U., Busch-Stockfisch M. Water quality and taste sensitivity for basic tastes and metallic sensation. *Food Qual. Prefer.* 2010, **21** pp. 243–249

ICS 67.240, 03.100.30

Price based on 28 pages