INTERNATIONAL STANDARD

ISO 11132

Second edition
2021-09

Sensory analysis — Methodology — Guidelines for the measurement of the performance of a quantitative descriptive sensory panel

Analyse sensorielle — Méthodologie — Lignes directrices pour le mesurage de la performance d'un jury descriptif quantitatif

Reference number
ISO 11132:2021(E)

COPYRIGHT PROTECTED DOCUMENT

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: 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.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

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. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 34, *Food products*, Subcommittee SC 12, *Sensory analysis*, in collaboration with the European Committee for Standardization (CEN) Technical Committee CEN/SS C01, *Food Products*, in accordance with the Agreement on technical cooperation between ISO and CEN (Vienna Agreement).

This second edition cancels and replaces the first edition (ISO 11132:2012), which has been technically revised. The main changes compared with the previous edition are as follows:

— the title has been changed to specify that the document is applicable to descriptive sensory panels;

— the Scope has been revised:

  — in order to provide a distinction of application for validation and monitoring, with improved wording to clarify;

  — it has been reduced to measure repeatability only, and reproducibility has been stated to be out of scope;

  — the type of quantitative descriptive sensory panels for which the document is applicable to has been specified;

— the definitions have been revised and new terminological entries have been added;

— the process for the dedicated procedure has been improved;

— experimental designs have been reviewed and augmented;

— statistical analyses related to analysis of variance have been reviewed and augmented to include more models, especially regarding sessions and panellists (fixed or random) effects and interactions;

— the subclauses (specifically the original 6.4.4 and 7.4) and Annexes B and C related to reproducibility have been removed to align with the changes to the Scope.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

A panel of assessors can be used as an instrument to identify products' sensory attributes and to assess the magnitude of sensory attributes.

Performance is the measure of the ability of a panel or an assessor to make reliable and valid attribute assessments across the products being evaluated. It can be assessed at a given time point, typically after a training period (validation) or tracked over time (monitoring). Performance comprises the ability of a panel to detect, identify and measure an attribute, use attributes in a similar way to other panels or between assessors within a panel, discriminate between stimuli, use a scale properly, repeat their own results, and reproduce results in comparison to other panels or assessors.

Measuring performance enables the panel leader to improve panel and assessor output, to identify issues and retraining needs or to identify assessors who are not performing well enough to continue participating.

Sensory analysis — Methodology — Guidelines for the measurement of the performance of a quantitative descriptive sensory panel

1 Scope

This document gives guidelines for assessing the overall performance of a quantitative descriptive panel and the performance of each panel member.

This document is applicable to the validation of the training of individual assessors or panels, as well as to the performance monitoring of established panels.

This document does not apply to the panel performance for descriptive methods where the individual scores of each assessor are not recorded, where there is no single list of attributes that is common to all the assessors, or where dominance rather than intensity is measured. Consequently, the performance of descriptive panels using methods such as consensus profile, free-choice profile, flash profile and temporal dominance of sensations (TDS) are out of scope.

The methods specified in this document are for monitoring and assessing the ability of a panel and its assessors to discriminate between products, the agreement between assessors of the same panel and the repeatability of these assessors in their intensity scoring.

Reproducibility, including both the comparison between panels and the comparison within the same panel of several evaluations conducted under different conditions (i.e. separated in time), is out of scope of this document.

The methods specified in this document can be used, in full or a selection only, by the panel leader to appraise continuously the performance of panels or individual assessors. The methods listed are not exhaustive and other appropriate methods can also be used.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements 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, *Sensory analysis — Vocabulary*

3 Terms and definitions

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

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

3.1
agreement
ability of different panels or assessors to exhibit the same product differences when assigning scores on a given attribute to the same set of products

3.2
panel drift
phenomenon where a panel, over time, changes in sensitivity or becomes susceptible to biases and as a consequence changes the location on the scale where an attribute is rated for a constant, reference product

3.3
performance
ability of a panel or an assessor to make reliable and valid assessments of stimuli and stimulus attributes

3.4
validation
process of establishing that a panel or assessor is able to meet specified *performance* (3.3) criteria

3.5
session
period of time in which products are assessed

Note 1 to entry: In a single session either one or several products may be assessed by one or several assessors. For an assessor, whether alone or as part of a panel, sessions are separated in time.

[SOURCE: ISO 5492:2008/Amd.1:2016, 4.63]

3.6
replicate
occurrence of a particular condition in an experimental design

Note 1 to entry: The term usually implies that the occurrence is one of several of the same kind, but it can refer to a single occurrence. When the condition is performed twice, the wording is "two repetitions", etc.

Note 2 to entry: To specify more than one occurrence of a condition, the terms "replication" or "replicate session" are more explicit.

Note 3 to entry: A "replicate session" is a *session* (3.5) in which the assessors, products, test conditions and task are the same.

3.7
assessor bias
tendency of an assessor to give scores which are consistently above or below the true score when that is known or the panel mean when it is not

[SOURCE: ISO 5492:2008/Amd.1:2016, 1.40]

3.8
order bias
arising from a product's spatial or temporal position relative to a group of products being assessed

Note 1 to entry: The term includes both "position bias" and "sequential bias".

[SOURCE: ISO 5492:2008/Amd.1:2016, 1.42]

3.9
repeatability
agreement (3.1) in assessments of the same products under the same test conditions by the same assessor or panel

Note 1 to entry: Repeatability can be measured within one *session* (3.5) or over several distinctly separate sessions, provided that the *replicate* (3.6) evaluations are conducted under the test conditions that can be considered to be the same. If replicate evaluations are conducted in distinctly separate sessions/sittings, the sessions are generally separated by several days only. In this case, the distinction between repeatability and *reproducibility* (3.10) in the short term is minor and relates to the test conditions being considered the same or not.

[SOURCE: ISO 5492:2008/Amd.1:2016, 1.45, modified — Note 1 to entry has been added.]

3.10
reproducibility
agreement (3.1) in assessments of the same products under different test conditions or by different assessors or panels

Note 1 to entry: Reproducibility may be measured as any of the following:

— the reproducibility of a panel (or an assessor) in the short term, measured between two or more *sessions* (3.5) separated by several days;

— the reproducibility of a panel (or an assessor) in the medium or long term, measured among sessions separated by several months;

— the reproducibility between different panels, in the same laboratory or in different laboratories.

[SOURCE: ISO 5492:2008/Amd.1:2016, 1.46]

4 Principle

4.1 Two possible approaches

4.1.1 General

This document is concerned with sensory panels used to assess the magnitude of one or more sensory attributes in order to make quantitative descriptions or profiles of products (see ISO 13299). Different methods are appropriate to the measurement of the performance of panels used for difference testing.

The performance of a quantitative sensory panel may be evaluated from panel sessions conducted specifically for the purpose of obtaining performance (called "dedicated procedure") by using assessments already available (called "ongoing monitoring").

4.1.2 Performance measurement via a dedicated procedure

A dedicated procedure is the method of choice for the certification of individual assessors and for other validation purposes. For the certification renewal, this dedicated procedure should be repeated at periodic intervals, as needed. Figure 1 is a flow chart for this procedure.

This approach can typically be used at the end of the training phase of a panel to ensure that the panel and the individual assessors have achieved the desired level of performance and can be considered as trained sensory assessors or expert sensory assessors (depending on the performance criteria).

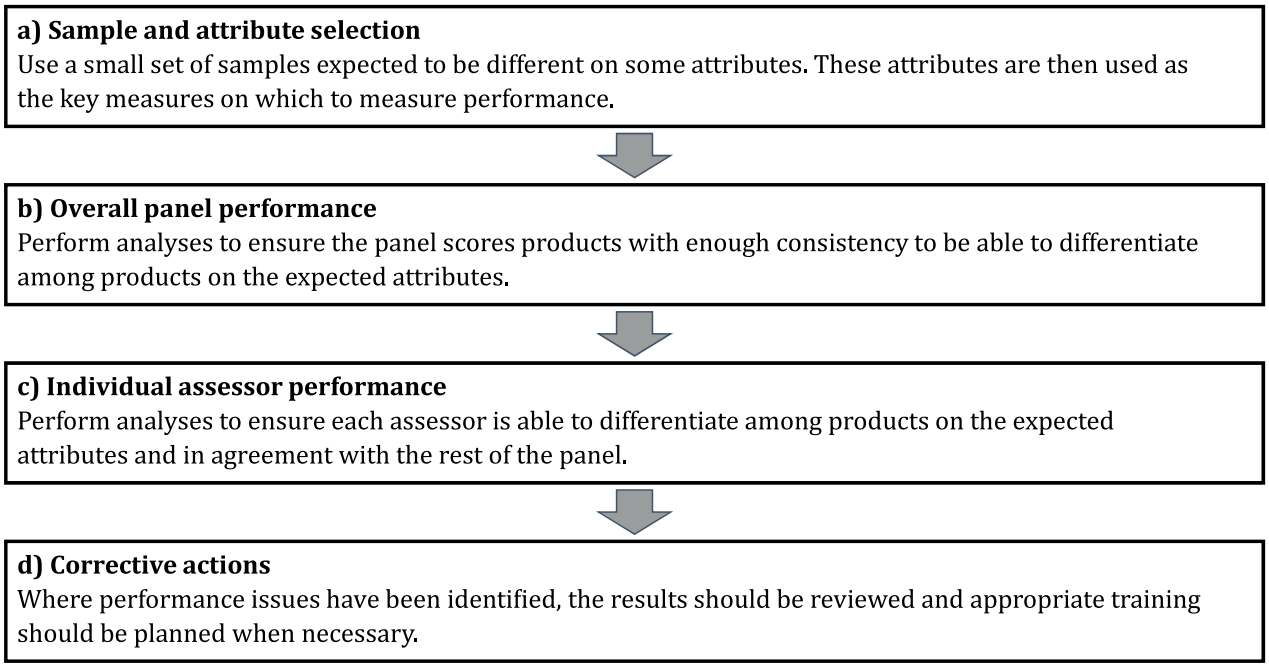

Figure 1 — Process steps for the performance measurement via a dedicated procedure

4.1.3 Ongoing monitoring via routine product profiling

Another approach consists of monitoring profile data that was already collected. To review ongoing profile data generated by a panel, it can be appropriate to use data that originated from quite different profiling experiments using different product types, product numbers, etc. The procedure is the same as that shown in Figure 1. However, as there are no predefined differences, it is recommended that attributes for which the products are significantly discriminated by the panel as a whole for a given profile be used as the key measures to check the performance of individual sensory assessors. Attributes that result in no significant difference cannot be reliably used to check consistency since the lack of agreement within and between sensory assessors probably means that the products are very similar for those characteristics.

In this case, over a given period, it will be necessary to check on a set of products more different that the panel is indeed capable of highlighting difference in these characteristics.

4.2 Indicators of panel or individual assessor performance

For one assessment, the following indicators can be determined:

— discrimination of the panel, measured as the ability of the panel to exhibit significant differences among products;

— discrimination of an assessor, measured as the ability of the assessor to exhibit significant differences among products;

— agreement of an assessor, measured as the degree of alignment between the assessor's average product scores and the ones of the panel;

— agreement of the panel, measured as the degree of alignment between the assessors' average product scores.

For replicate assessments:

— repeatability of an assessor, measured as the degree of homogeneity between replicated assessments of the same product;

— repeatability of a panel, measured as the average degree of homogeneity between replicated assessments of the same product for each assessor.

4.3 Statistical analyses

A single, consistent approach to statistical analysis of the results is described in this document. However, some indicators of panel performance can be assessed by more than one measure. For instance, error mean square and error standard deviation (SD) (its square root) both express variability in the evaluation of a product. The measures used should be those that are usual in the field of application.

Other relevant measures of agreement between assessors in the use of the scale for an attribute are the interaction of assessor and product and the coefficient of correlation between an assessor's scores and the panel means. An assessor may have no bias but may use the scale in a different way. A correlation close to 1, a regression slope close to 1 and a regression intercept close to 0 indicate good agreement between an assessor and the rest of the panel.

When each assessor evaluates a small number of samples (fewer than six), the correlation coefficient should be interpreted with caution, as it can be high (up to 0,7) by chance alone.

5 Prerequisites

5.1 Experimental conditions

The test facilities should be in accordance with ISO 8589.

5.2 Qualification of assessors

The panel should have the level of qualification and experience of selected/screened assessors in accordance with ISO 8586 or higher.

6 Performance measurement via a dedicated procedure

6.1 Sample and attribute selection

At each dedicated study, the panel of assessors should be presented with a set of samples similar to those the panel are to assess when evaluating products and for which statistically significant differences between at least one pair of the samples are expected for each of the relevant attributes.

In order to ensure that all key aspects of the products are examined, an adequately diverse set of attributes should be included in the test.

These relevant attributes are used as key measures against which to assess panel performance. The sample set should include replicates. There should be the same number of replicates of each sample. The replicates can be evaluated within a single session or over two or more sessions. The number of assessors, samples and replicates depends on the products, sensory attributes assessed and purpose of the procedure. For example, two or three replicates of three or four samples might be used. Care should be taken to limit the number of assessments required in a session, so as to avoid sensory fatigue. The attributes of the samples should be similar to the range of values that the panel assesses when evaluating products.

6.2 Experimental designs

6.2.1 General

Several types of experimental designs can be used in the dedicated procedure, depending on the most important objective to answer.

6.2.2 Randomized block design

A randomized block experimental design can be used, in which the assessors are the “blocks”. This design is appropriate when no carry-over effect from one sample to the next is expected. Otherwise, a balanced design should be used instead (see 6.2.3).

6.2.3 Balanced and random designs

If a carry-over effect is expected from one sample to the next, a suitable experimental design is the Williams Latin square[12]. Table 1 shows the Williams Latin square design with four assessors and four samples.

Table 1 — Williams Latin square design for four assessors and four samples

Assessor	Session	Order			
		1	2	3	4
1	1	A	B	C	D
2	1	B	D	A	C
3	1	C	A	D	B
4	1	D	C	B	A
1	2	B	D	A	C
2	2	C	A	D	B
3	2	D	C	B	A
4	2	A	B	C	D

In this design, each assessor samples the four products in a different order in a given session and any particular product is followed by a different one for each assessor. For example, in session 1, A is followed by B for assessor 1, C for assessor 2, D for assessor 3 and none for assessor 4.

For each replicate of the products’ evaluation, it is recommended to use a different product order for each assessor, in order to reduce the order effect and the carry-over effect.

If multiples of four assessors are available, the same design can be repeated for each set of four.

It is also possible to choose a random product order design, i.e. to randomly affect each product to each position in each session.

The advantage of these approaches is to minimize the carry-over effect at panel level and therefore get better estimates of the product means at panel level for performance evaluation. However, if the product order does have an impact, the agreement between assessors will be impacted because each assessor will not experience the same product order. In order to compare the assessors on the exact same task, the same product order can be used for all sensory assessors (see 6.2.4).

6.2.4 Same order design

In order to focus on individual assessor performance and, in an effort to compare assessors under the most similar conditions, an alternative design is proposed, whereby all assessors evaluate the products in the same order, see Table 2.