भारतीय मानक

Indian Standard

IS 18370 : 2023
ISO 20784 : 2021

संवेदी विश्लेषण — संवेदी और उपभोक्ता उत्पाद दावों के प्रतिस्थापन के लिए मार्गदर्शन

Sensory Analysis — Guidance on Substantiation for Sensory and Consumer Product Claims

ICS 67.240

भारतीय मानक ब्यूरो
BUREAU OF INDIAN STANDARDS
मानक भवन, 9 बहादुर शाह ज़फर मार्ग, नई दिल्ली - 110002
MANAK BHAVAN, 9 BAHADUR SHAH ZAFAR MARG
NEW DELHI - 110002
www.bis.gov.in www.standardsbis.in

September 2023 **Price Group 9**

Test Methods for Food Products Sectional Committee, FAD 28

NATIONAL FOREWORD

This Indian Standard which is identical to ISO 20784 : 2021 ‘Sensory analysis — Guidance on substantiation for sensory and consumer product claims’ issued by the International Organization for Standardization (ISO) was adopted by the Bureau of Indian Standards on the recommendation of the Test Methods for Food Products Sectional Committee and approval of the Food and Agriculture Division Council.

The text of ISO Standard has been approved as suitable for publication as an Indian Standard without deviations. Certain conventions are, however, not identical to those used in Indian Standards. Attention is particularly drawn to the following:

a) Wherever the words ‘International Standard’ appear referring to this standard, they should be read as ‘Indian Standard’; and

b) Comma (,) has been used as a decimal marker while in Indian Standards, the current practice is to use a point (.) as the decimal marker.

In this adopted standard, reference appears to the following International Standard for which Indian Standard also exists. The corresponding Indian Standard which is to be substituted in its place, are listed below along with its degree of equivalence for the edition indicated:

International Standard	*Corresponding Indian Standard*	*Degree of Equivalence*
ISO 5492 Sensory analysis — Vocabulary	IS 5126 : 2016/ISO 5492 : 2008 Sensory analysis — Vocabulary (*second revision*)	Identical

In reporting the results of a test or analysis made in accordance with this standard, if the final value, observed or calculated, is to be rounded off, it shall be done in accordance with IS 2 : 2022 ‘Rules for rounding off numerical values (*second revision*)’.

Contents

Page

Introduction

This document provides a framework to be considered when planning studies designed to support a sensory claim to consumers. The document specifies the principles to be followed, defines the key terms, provides a classification scheme with examples, and also provides case studies for different types of claims.

Product claims are designed to inform the audience of potential buyers or the general population of users about the product characteristics, differentiate the product from its competitors, and influence the buying decision.

Sensory claims in this document refer to the information and messages provided on a pack or label, in printed sales literature, or on television or digital media to communicate to the target users, or potential target users, information about the product's sensory attributes or the user's response to the experiences of using/consuming the product.

In general, countries have guidelines that provide information about claims substantiation testing (for a list of sources, see the Bibliography), or regulatory, legal or media-based requirements that govern advertising messages about products.

Given that the rapid development of new products has resulted in a crowded marketplace in some countries, from a company's standpoint there is more need for sensory claims to differentiate between products. At the same time, however, governments and regulatory bodies also look to protect consumers from misleading claims. Therefore, researchers are more frequently asked to design, conduct and interpret results of sensory claims studies.

This document is intended to guide those researchers to support sensory claims on a scientific basis. Using this guidance will allow sensory professionals to conduct research in a manner that provides competent and reliable evidence to support a claim.

Competent and reliable evidence provides proof that test design, data collection and data analyses are done using sound scientific principles and implemented in a technically competent manner. What constitutes competent and reliable evidence is established by the scientific community but will be debated by the legal community. Establishing that a test result can serve as competent and reliable evidence to support a claim can be done by:

a) qualified persons knowledgeable in the practice of science-based sensory and consumer testing;

b) those persons agreeing that best practices were followed.

In addition, good sensory practice means that the test data are analysed using sound statistical procedures. Product researchers designing such tests should be aware of, and follow, best practices in the sensory and consumer testing community.

Indian Standard

SENSORY ANALYSIS — GUIDANCE ON SUBSTANTIATION FOR SENSORY AND CONSUMER PRODUCT CLAIMS

1 Scope

This document gives guidelines for substantiating sensory claims on food and non-food products and their packaging for advertising consumer-packaged goods.

This document differentiates sensory claims from other types of claims. It provides classification and examples of the different types of sensory claims. It highlights special issues associated with testing to substantiate sensory claims. It includes case studies and references.

This document does not apply to:

— specific or detailed requirements for different test methods that are used to support sensory claims;

— factual claims regarding a product's country of origin, ingredients, processing and nutritional components;

— factual claims regarding the technical features of the product;

— claims regarding a product's health, medical or therapeutic benefits, physiological effects, structure or function benefits when consumed or applied to the human body;

— claims based on instrumental assessments of the attributes or performance of a product (i.e. instrumental assessments; in this case, test methods are used in which no human participant evaluates the product and/or no human participant provides a response to a product);

— claims about services (e.g. a house cleaning service, airline services, automobile services);

— claims about large/slow moving consumer goods (autos, refrigerators, stoves, etc.).

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
sensory claim
advertising message to the consumer about a product's sensory attributes (e.g. "now with roasted flavour"), its functionality (e.g. "removes grease") or performance (e.g. "keeps breath fresh longer") and/or the consumers' affective (e.g. "consumers prefer Brand X") or perceptual responses (e.g. "makes your skin look younger") to it before, during or after use

Note 1 to entry: A sensory claim can be any such message that is used in an advertisement, in any format. This advertising message is made in a public forum. It can appear on the product package, in print or in the media in any format (electronic, television or video). The aim of a sensory message is to inform the potential user/buyer of the product about the product's sensory characteristics, or to highlight what they will perceive while using, or after using, the product. This type of advertising message is to make the potential user/buyer aware of these attributes with the intent to influence purchase, consumption or usage.

3.2
affective claim
message about a user's/potential user's hedonic and/or emotional responses before, during or after using a product

Note 1 to entry: Responses include a consumer's hedonic, attitudinal, cognitive and/or emotional reactions elicited by the product before, during and/or after the use of a product. The most common hedonic responses measured are liking or preference. Attitudinal responses can be that the user is more willing to purchase the product in the future or that the consumer will agree to a statement that the product has the highlighted attribute or provides the specified affective experience.

3.3
perception/performance claim
message about a characteristic describing the perception elicited by the product or its intended effect

EXAMPLE Product XXX is thick (perception) and leaves no residue (performance).

3.4
puffery
extremely broad, vague and subjective statement that is so exaggerated that it is not likely to be believed and is not testable from the standpoint of measurement operations and/or from a practical standpoint

3.5
equivalence claim
message provided when two or more products are stated to be the same in one or several features

3.6
unsurpassed claim
message provided when a product is stated not to be exceeded by any other product in one or several features

3.7
superiority claim
specific type of comparative claim that states a higher level of a product/performance attribute or liking/preference relative to one or several products

3.8
risk
combination of the probability of an occurrence of harm and the severity of that harm

Note 1 to entry: The sensory scientist and the organization's stakeholders should consider the risks associated with making a claim based on sensory testing. Risk refers to the chances/likelihood that making the claim results in negative consequences. These negative consequences can be as ill-defined as consumers making negative posts on social media, or comments within the community of sensory, marketing or legal communities, or extend to a challenge from a competitor, or action taken by a self-policing, regulatory or governmental agency. Risks should be identified, discussed and understood before the claim is made publicly.

[SOURCE: ISO/IEC Guide 51:2014, 3.9, modified — Note 1 to entry has been replaced.]

3.9
sensory analysis methods
set of widely used, science-based methods of sensory analysis, including descriptive, discrimination and performance evaluations

Note 1 to entry: Internal validity and laboratory-type controls are the hallmarks of sensory test methods, especially when the goal is to measure products' attributes. In both product-focused and participant-focused sensory tests, blinding the product's brand identity and obtaining independent judgements are best practices. Test method sensitivity, effect size, and the number and type of assessors are all considerations in sensory test methods (see ISO 6658).

3.10
consumer methods
range of consumer quantitative product testing methods used by most practising sensory/consumer professionals engaged in product testing, which includes both affective testing and perception/ performance testing

3.11
representative sample of consumers
set of people, being a smaller group than the larger population from which it is drawn, used in a test, which provides a test result that covers the range of responses that would be obtained if the entire larger population had been tested

Note 1 to entry: In consumer testing, where affective responses are recorded, the general requirement for sampling participants is usually the following: a) include a large enough number of consumers in order to cover the variability in affective responses present in the larger population; b) use qualified consumers that actually use or consume the product, purchase the product or are product-concept acceptors; c) have a sampling plan that includes relevant demographic, geographic, behavioural or psychological variables.

3.12
representative sample of products
set of products, drawn from the available products in the marketplace, used in a test when the desired claim is about a product's sensory attributes, performance attributes or the hedonic responses it elicits

Note 1 to entry: A recommended practice for researchers is to obtain products to test from a retail channel, after being factory-made and having gone through the usual distribution channels. The principle is that the products that serve in a claims test should be representative of what the consumer would get if they purchased the product in the marketplace. It is also a recommended good practice to include more than one lot, one batch or one factory in the product selection whenever there is a need to include the variability normally present in the product. Bench samples or prototypes developed before scaling up, distribution and market launch can be used to support a claim. If bench prototypes are tested and the results used for claims support, there should be some evidence to show the comparability of the bench sample and the in-market product. The number of products tested and the product variables included in the product selection depends on the extent to which the advertising message conveys that it applies to the entire set of products consumers would use.

3.13
objective test result
test result obtained by using procedures generally accepted in scientific fields as providing a result that did not depend on the experimenter's expectations or interference (and is verifiable)

Note 1 to entry: In objective studies, data are collected without interference by the experimenter, and the study design allows more than one possible outcome. Respondents in such studies also do not have information about the underlying study objectives that might affect their responses. Objective results come from studies that do not depend on the experimenter's or test administrator's opinions or desired outcomes, and where the relevant variables are included, best practices are followed and the results are not a foregone conclusion. Objective research results are usually obtained across different studies and can be corroborated with other related studies or converging test results.

4 General considerations

To substantiate a sensory claim, the following considerations should be taken into account.

a) Sensory claims are based on the recorded responses from the perceiver's direct experience with a product.

b) The product claims are designed to inform the audience of potential buyers about the product characteristics, differentiate the product from its competitors and influence the buying decision.

c) Claims directed at children or children's products are heavily regulated in different countries and are quite restricted.

d) What makes a claim supported by evidence is the proper use of scientific methods along with the weight and relevance of the supporting data.

NOTE 1 Researchers in companies usually test their products frequently. If there is frequent testing of a product, there is often a set of other results against which a single test result can be compared. The support for the claim is stronger when the multiple test results converge. A single test result can be used as the basis of a claim if it is within the range of previously obtained results. If there are no previous tests and no previous test results, other technical information or evidence that is consistent with the single finding can strengthen the support for the claim.

NOTE 2 If a vendor or testing organization performs the test and has no history of previous results and a comparison cannot be made, then it is incumbent upon the company making the claim to establish the risk in making a claim based on a single test.

e) Sensory claims can be based on standardized, scientifically based measurements of:

1) the properties of products/performance established via sensory analysis;

2) users' liking, preferences, feelings, attitudes, or perception of product properties or performance.

f) Claims requiring a clinical study are defined by the Helsinki document[5]: a clinical trial is "any systematic evaluation of medicinal products or devices in human subjects whether in patients or non-patient volunteers, to discover or verify the therapeutic effects of, and/or identify any adverse reactions to them, and/or to study their absorption, distribution, metabolism and excretion *in order to ascertain the efficacy and safety of products*". Claims requiring a clinical study are not sensory claims and are hence are out of scope for this document.

g) In contrast to studies that focus on the therapeutic effects of a product or effects a product might have on underlying structures or processes, a sensory study involves an evaluation of a sensory effect such as a specific mouthfeel, an aroma or an appearance.

EXAMPLE In the case of skincare products, if the main purpose of the sensory study is to demonstrate that there is a change or improvement in the appearance of the skin, visually detectable by the unaided eye and evaluated by a human assessor, this is a sensory study. Demonstrating changes in the tactile properties of the skin, when measured by an external assessor without instrumentation, is also a sensory study.

If the skincare product claim is that there are changes in the underlying structure/function of the skin or in the overall health of the skin surface or underlying dermis, this is a clinical study and is not a sensory study.

NOTE 3 A fine line distinguishes these two types of studies. Undoubtedly, the types of studies that are called "sensory" and "clinical" in the cosmetic and personal care categories vary by country and company. In this document, claims describing a benefit due to underlying structure/function changes within the body are considered to require a clinical study and are hence out of scope.

h) Puffery claims or "hyperbolic claims" (defined by national legislation) are either so vague (e.g. "this perfume will give you wings") or are so exaggerated (e.g. "world's most comfortable shoes") that no one would take them as literally true. Puffery claims cannot be supported by data from science-based testing, either due to their vagueness or due to pragmatic considerations.

i) Researchers planning a sensory claims test must be knowledgeable about the regulatory, governmental, media or potential competitor responses that might be made to a sensory claim. Researchers should address relevant regulations and likely competitive responses in the study design and analyses.

5 Guiding principles for sensory claims substantiation

5.1 General

When performing a sensory claims substantiation by sensory and/or consumer methods, investigators should take into account the nine principles given in 5.2 to 5.10 to obtain test results robust enough to substantiate a claim.

5.2 Reviewing governmental legislation and regulations

The governmental, regulatory, advertising and media requirements and standards for the country(ies) in which the claims will be made should be consulted before beginning a claims test.

5.3 Defining the primary claim(s) and designing the test to address them

The research should be designed to support the primary claim(s) of interest. Determining the wording of the claim in advance of conducting the test is therefore recommended to design the study appropriately. Based on the intended claim(s), the primary endpoint(s) should be selected (e.g. descriptive attributes or consumer questions). These should ideally be limited to a few to avoid the multiplicity effect[7]. It should also be remembered that the more items evaluated in the study, the greater the chances of spurious or contradictory findings. See Annex B for a list of the changes in probabilities as a function of increases in the number of statistical tests in a study.

NOTE In some countries, it is necessary to state the type of sensory claims in advance of conducting research.

5.4 Determining the type of claim: single product or comparative testing

The type of claim should determine whether testing is monadic, paired or uses multiple products. Non-comparative claims should be carried out with monadic testing. Comparative claims should be carried out with product pairs or one product compared to multiple products if the claim refers to the product category in general.

NOTE If a comparative claim is being made, the researcher defines the number(s) and types of products that will serve as a comparison(s). See ASTM E1958 for specific guidance on selecting competitive products in comparative claims tests.

5.5 Defining the decision criteria

Decision criteria should be defined in advance and should be unambiguous.

5.6 Defining the set of relevant products

For comparative claims, the set of relevant products should be defined in advance.

5.7 Defining the population of relevant consumers or assessors

For consumer tests, the population of relevant users, current users, purchasers or potential consumers should be defined in advance of conducting the research. For product-focused tests, assessors with the relevant qualifications and training are defined in advance.

5.8 Defining the strength of evidence

The strength of evidence should be appropriate for the level of challenge expected.

NOTE Researchers need to be aware of how strong the support for a claim is, along with the risk of it being challenged either by a governmental or regulatory organization, or by a competitor. For all claims, the strength of support is determined by:

a) the extent to which the result has been found more than once, which allows the claimant to be more certain the finding is "robust";

b) the extent to which there is converging evidence to support the claims, i.e. checking to see if this result is consistent with other information/results from other tests and measurements;

c) how relevant and representative the respondent sample and the product sample are;

d) how well experimental and confounding variables are managed within the study;

e) the extent to which best practices have been followed in the conduct of the test as described in the preceding principles, including choice of and adherence to (the) primary endpoint(s) prior to conducting the test (see 5.3).

5.9 Ensuring impartiality

The organization that wants to make the claim should ensure the impartiality of the methods, testing conditions and study execution. For example, using same-aged products, prepared and presented in the same manner, with balanced presentation of order, are procedures to follow to ensure both products are tested fairly.

5.10 Ensuring reliability

Best practices should be applied to ensure the research is objective and unbiased, and the finding is robust enough to be obtainable more than once.

6 Classification of sensory claims

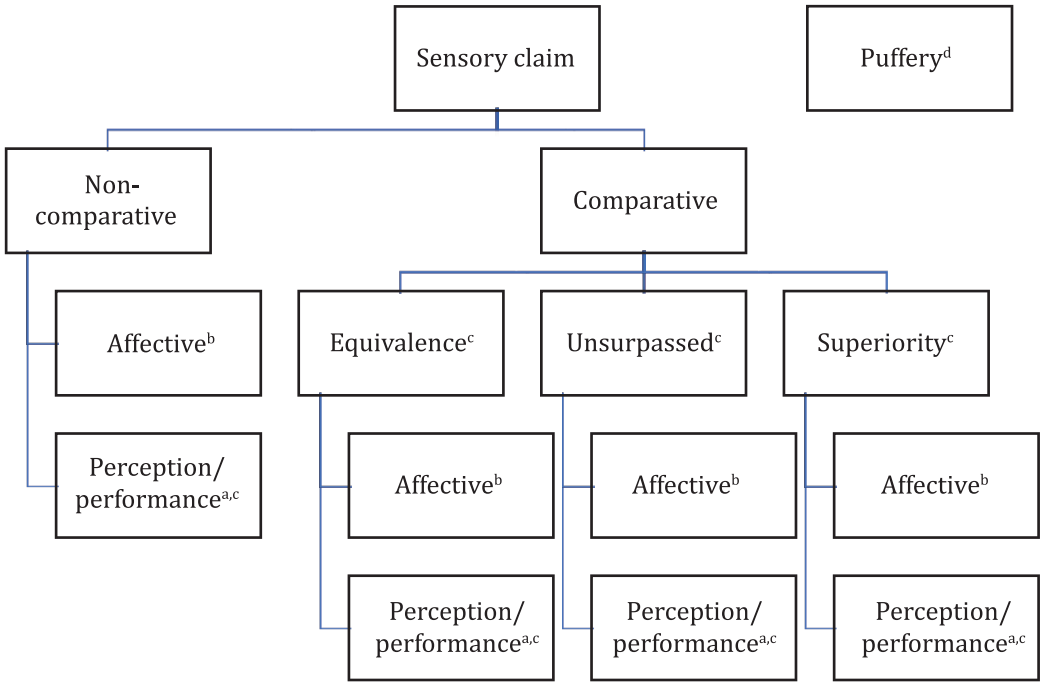
6.1 Classification

The classification of sensory claims helps to develop a clear and rational substantiation and also facilitates the selection of the testing method. The classification schema is given in Figure 1. The main differences in the classification of claims are whether:

a) the claim refers to a single product or a product compared to its competitor(s);

b) the claim refers to the consumer's affective response or perception of the product and/or its performance;

c) the consumers and/or trained assessors can serve as respondents.



a Can be tested by using trained assessors or relevant consumers.

b Relevant consumers (either current purchasers or potential users) serve in affective claims tests.

c The type of respondents used in equivalence, unsurpassed and superiority claims depends on whether the claims are affective or perception/performance.

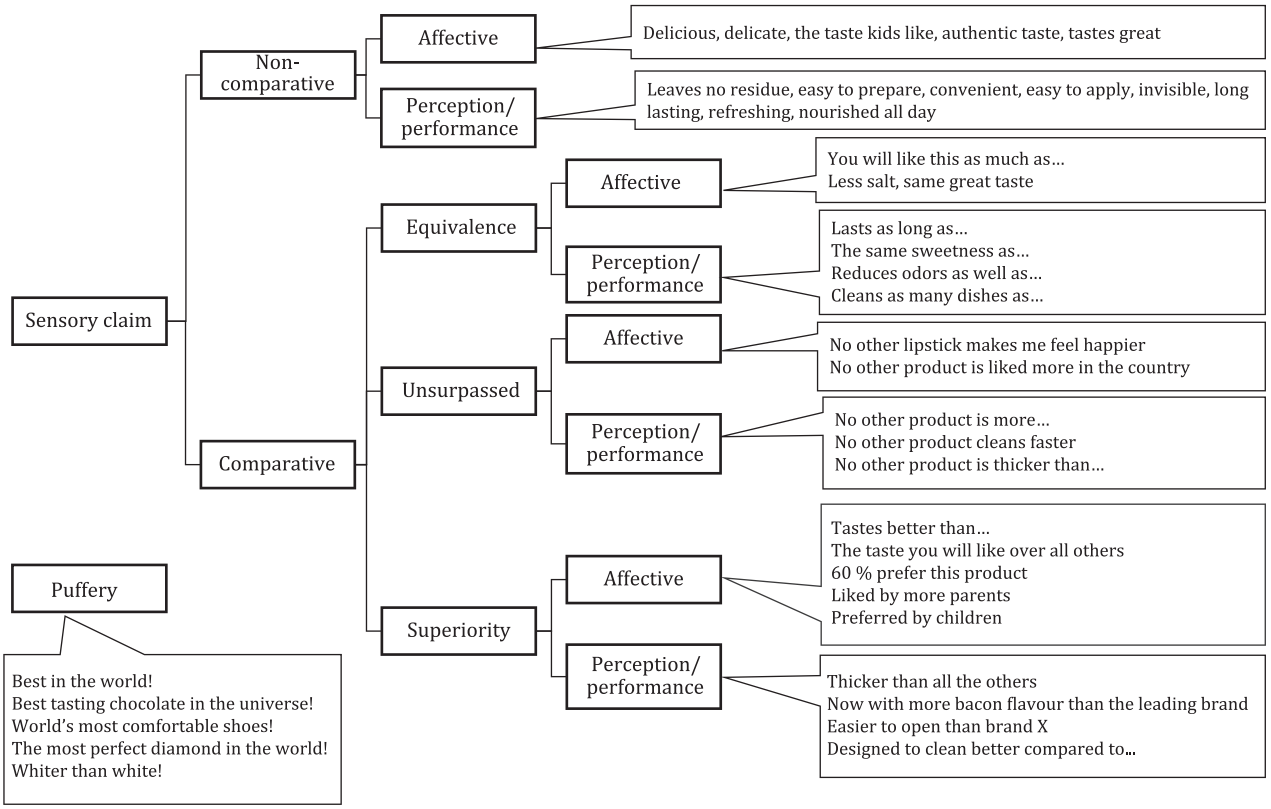
d Puffery is a special type of claim, extremely exaggerated and/or vague, that is not testable using sensory methods.

Figure 1 — Classification schema of different types of sensory claims

6.2 Wording

The examples of wording of the different types of sensory claims are described in Figure 2. The examples in Figure 2 are only a small representation of possible sensory claims.

The wording of a sensory claim will vary with the product category, the desired target consumers and the country where the sensory claim is made. The determining factor is how directly and clearly the sensory test supports the wording of the sensory claim.



NOTE There are many possible sensory claims and many ways to word them. Those listed above are illustrative examples. For different examples of cases of sensory claims substantiation, see Annex A.

Figure 2 — Specific examples of the wording of different types of claims

6.3 Non-comparative sensory claims

This type of claims test involves monadic or single product tests, done with a specified objective and decision criteria, and with quantitative data analysed statistically. For non-comparative claims that are affective, consumer quantitative test methods are used. For non-comparative perception/performance claims, descriptive analysis panel evaluations or consumer's self-reports of perceptions elicited by the product can be used to support a claim. These data are also analysed statistically.

6.4 Comparative sensory claims

Comparative claims refer to claims made to compare two or more products to each other. This comparison can be of an advertiser's product to one or more competitive products, or to a version of the advertiser's own product. The relative positions of the compared products are one of the following: a) superiority, b) equivalence or c) non-inferiority. Comparative claims can highlight either a difference in or something new in a product attribute, affective response or product performance.

If an advertiser states a comparison between two of their own products, e.g. a product revision and a current product, the risk of a challenge is low. However, some companies wish to have some sensory data to support even a low risk claim. They can require test support for a claim highlighting an attribute difference between the new version of the product versus a different version. In this case, information from historical testing of the previous product can be used to inform a success criterion. Because the risk of challenge is low, the success criteria selected can be less stringent than higher risk claims.

Advertising one's own product by comparing it to a competitor's product in publicly made messaging is riskier. The competitor can challenge comparative claims, so it is important that the advertiser conduct rigorous testing to support the claim being made.

Product superiority claims are messages stating that the advertiser's product is more well-liked or preferred, has more of a desirable product attribute or out-performs the product that the claim is being made against.

EXAMPLE 1 "Consumers prefer our brand over competitor X", "Brand X cleans better than the other leading brands".

Equivalence claims are messages stating that the advertiser's product is as good as or is the same as the product that the claim is being made against.

EXAMPLE 2 "Tastes as good as…", "Cleans as well as other leading brands".

Non-inferiority (or unsurpassed) claims are messages stating that the advertiser's product is at least as good as the product(s) that the claim is being made against.

EXAMPLE 3 "No product is liked better than…", "No product has more fruity flavour", "You just can't beat the great taste of…", "Nothing cleans better than…".

An equivalence claim refers to the equality between two or more products. The statistical hypotheses associated with equivalence claims include both upper and lower bounds (i.e. a two-sided alternative hypothesis). The statistical hypotheses associated with superiority and non-inferiority claims, on the other hand, involve only a lower bound (i.e. a one-sided alternative hypothesis). All comparative claims require their own test design and analysis methodologies. It is important to note that equivalence and non-inferiority must not be deduced from the absence of a statistically significant result in a test. Equivalence claims require increased sample sizes compared to claims tests designed to show product superiority or non-inferiority.

All types of comparative claims (superiority, equivalence and non-inferiority) require quantitative test data analysed statistically.

If an affective comparative claim is made, then the data collected are usually liking or preference. If a comparative perception claim or performance attribute claim is made, either a consumer or a trained panel method can be used.

7 Methods

7.1 Many sensory analysis methods can be used to substantiate sensory claims. The specific method used to obtain data to support a sensory claim depends on the type of claim and on the objective of the test, see ISO 6658.

In general, analytical sensory methods such as descriptive and discrimination testing are product-focused. Specifically, the purpose of descriptive or discrimination testing is to determine either what sensory attributes are present in the products and/or if there are sensory differences on those attributes. Often, trained or experienced assessors serve as assessors in descriptive and/or discrimination tests. It is important that sensory researchers wishing to establish a product-based claim are aware of any country-specific regulations around whether or not sensory attribute claims, if targeted to the end consumer/product user, must also be consumer-perceptible.

Affective tests such as consumer liking or preference should be done using a reasonable and representative sample of product users, or, if the product is new to the market, potential users. Claims based on affective testing are person-based messages, e.g. "consumers prefer this product compared to the competitor's product". The claims made from affective tests are about the responses of the consumers who served in the test. To conduct an affective test, a representative sample of consumers should be drawn. The representative sample of consumers should be large enough and recruiting requirements should be relevant.

Consumers in affective tests are the following:

a) respondents who consume the product, or are purchasers, or potential purchasers/users (if the product is not yet on the market);

b) test participants who are representative of a larger group of potential consumers, purchasers or users;

c) providers of subjective responses to the product, such as liking or preference, or check-all that-apply (CATA) ratings of perceived product attributes or a self-reported agreement.

Like product-focused sensory test methods, blind product exposure and independent judgements are also hallmarks of consumer test methods. External validity, i.e. the extent to which the results can be generalized to the larger population of users/potential users and the extent to which the product was used as consumers would normally, is the most important in consumer methods.

In both analytical sensory tests and consumer tests, the products tested should be representative of the product available in the marketplace. Purchasing products from retail stores is the most direct way of obtaining products representative of what consumers would purchase and consume/use. If a product is not yet launched and available in the marketplace, then the research must provide evidence that the product tested in the claims substantiation test is the same as the product in the marketplace.

If an innovative test method is used (as needed when innovative products are developed), the test method used to support the sensory claims should be based on sound scientific principles that most sensory professionals would agree on.

7.2 Exploratory, hypothesis-generating or discovery research methods should not be used as primary support of a sensory claim. Qualitative test methods when used without any other type of converging evidence should not be used to support claims.

7.3 Different methods can be used to substantiate a sensory claim, including:

a) overall discrimination test (e.g. triangle test, duo-trio test);

b) attribute difference test (directional difference test, attribute rating test);

c) descriptive analysis test;

d) consumer affective or performance test.

7.4 The guiding principles for sensory testing to support sensory claims are provided in Clause 5 and cover the following steps:

a) awareness of the country-driven requirements for making a claim;

b) advance definition of the primary claim;

c) impartial data collection and unbiased measurement;

d) participants and product requirements relevant to the claim.

7.5 Relevant consumers, either current or potential users, should serve in affective claims tests or perceived product functionality. Product attribute or performance attribute claims can be tested by using trained sensory assessors or selected assessors to evaluate product performance characteristics.

Annex A
(informative)

Sensory claims case studies

A.1 Non-comparative — Affective: "Tastes great"

A company that makes salad dressing wants a "tastes great" message on the front of their label. The sensory, marketing and legal team agree that a product test is needed. If a majority of the consumers who taste their product agrees that the product does "taste great", they will put this message on their packaging front label. The researcher reviews the relevant historical data on similar products consumers have rated for "tastes great" to have some context information to set a criterion. Based on this review, the researcher proposes the decision criterion "70 % or more of consumers who taste the salad dressing will agree with the statement that this salad dressing tastes great".

In a central location test, a sample of 120 consumers of this product type tastes this product blind and rates the extent to which they agree with the statement that the product "tastes great". Consumers rate their response on a seven-point Likert scale. The "7" is "agree completely", "1" is "disagree completely" and "4" is "neither agree nor disagree". The researcher recommends the top two box ratings, "agree completely" and "agree very much", are the two ratings that count towards the number of consumers who "agree" that the salad dressing "tastes great". This decision is to ensure only the consumers with strong agreement are counted towards meeting the claims decision criterion.

Results were tabulated and 96 (80 %) of the 120 consumers either "agree completely" or "agree very much" with the statement "this salad dressing tastes great". A 95 % confidence interval for the obtained percentage, 80 %, based on the binomial distribution, is [0,717; 0,867]. Consequently, the percentage of consumers who would give an "agree" response is between 72 % to 87 %. The lower bound of this 95 % confidence interval is higher than the predetermined 70 % criterion. The test results are used to support the message "tastes great" on the packaging.

See ISO 11136 for hedonic test methods with consumers in a controlled area.

NOTE The calculations can be performed, e.g. by using the following line of code in R: binom.test(96,120). Calculations were done using Microsoft Excel 2010.[1)]

A.2 Non-comparative — Performance: "Leaves no residue"

A product developer identifies an ingredient to add to a skin cream that eliminates residue. The company wants to add this ingredient to their current best-selling skincare cream and advertise "leaves no residue". There needs to be sensory data to support this "leaves no residue" claim for the new formulation of skin cream. The sensory scientist proposes a trained and validated descriptive analysis panel be used to evaluate the skin cream with and without the new ingredient. The goal is to determine objectively if this new "no residue" ingredient does provide this benefit. The descriptive panel is trained and routinely evaluates skincare products. As a success criterion, the team aims at showing that the skin cream is more often judged as leaving "no residue" than the opposite, i.e. the proportion of "no residue" answers should be significantly larger than 0,5. (The data for the control, i.e. without the "no residue" ingredient is not used here, but was only used to show that the control is judged to be "with residue".)

The descriptive panel consists of 12 trained assessors who can apply the skin creams to their forearms in a standard way. These assessors also evaluate the appearance of their forearm skin cream after

1) Microsoft Excel is an example of a suitable product available commercially. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO of this product.

application. Assessors using standard amounts and standard application method, apply the two creams, one on each forearm. Each skincare cream is identified only by a three-digit number code. One of the creams contains the "no residue" ingredient and the other does not. In the jar, the two creams are not distinguishable in aroma, appearance, texture and feel when applied.

After three minutes, the trained assessors evaluate each of their forearms for "visible residue".

The same procedure is repeated a second time in the following week to ensure there is robust support for the "leaves no residue" claim. Data from the two trials are shown in Table A.1.

Table A.1 — Data from the two trials

Judgements	Trial 1	Trial 2	Total Trial 1 + Trial 2
Total "no residue" judgements	10	8	18
Proportion of "no residue" judgements	0,83	0,67	0,75

Data from the two trials were combined for the analysis. A total of 18 "no residue" judgements out of 24 was obtained for the skin cream with the added new ingredient. Using the binomial law with the probability of 0,5 to obtain a "no residue" answer, the probability of obtaining 18 "no residue" judgements or more out of 24 by chance is less than 0,023. The company decides to add "leaves no residue" to their printed packaging of their reformulated skin cream to highlight this benefit to consumers.

A.3 Non-comparative — Performance: "Easy to prepare"

A company develops an "easy to prepare" macaroni and cheese that can be prepared with fewer steps than the product the company currently has in the market. The cross-functional team agrees on the success criterion that the new macaroni and cheese product must have a majority of consumers of their current brand rate the new recipe as "easy to prepare". A "majority of consumers" is defined as at least 80 % or more of the consumers in the claims test rate the new recipe as "easy to prepare".

To test this non-comparative performance claim, the company decides to provide an unbranded box of the new product to a group of 90 consumers who regularly buy macaroni and cheese and for them prepare it at home. The consumers prepare the product following the instructions listed on the box. The consumers rate if this macaroni and cheese is "easy to prepare". They are then asked to complete a questionnaire on their computer about how easy the product is to prepare. "Ease of preparation" is measured on a five-point rating scale, with "5" being "very easy to prepare, "1" being "not at all easy to prepare" and "3" being "neither easy nor hard to prepare".

Eighty (89 %) out of the 90 consumers rate the new recipe as either "very easy to prepare" or "easy to prepare". Using the binomial law, it is shown that the proportion of interest is indeed larger than 80 % (p = 0,035) and the 95 % confidence interval is calculated to be between 0,81 to 0,95 (or 81 % to 95 %) of consumers. This result satisfies the pre-determined success criterion. Marketing prepares the packaging and sales materials with the claim "easy to prepare" for a retail launch.

A.4 Comparative — Affective: "30 % reduced salt, same great taste"

A cracker company wants to make their crackers healthier to appeal to consumers' growing desire to consume healthier snacks. It decides to reduce the salt in their core cracker brand by 30 % while maintaining the taste of the cracker as much as possible. In the 30 % reduced salt crackers, the surface area of the salt particles is changed to provide a more intense perception of salt compared to the salt used in the current crackers. Sensory data are needed to establish the similarity of the saltiness intensity in the 30 % reduced salt and the current cracker. To address this need, the sensory scientist plans to do a directional paired comparison test.

The cross-functional team also wants some consumer data to support the claim "same great taste". The team agrees on the criterion that if at least 70 % of the consumers who taste the 30 %-reduced-salt cracker agree that it has a "great taste", the "less salt, same great taste" claim is supported.

After several rounds of revision and internal panel testing of the product, the product development team has a recipe that is relatively similar in perceived saltiness compared with the current cracker. The team reduced the salt by 30 % while maintaining only a small difference in perceived saltiness and delivering a taste profile they consider indistinguishable from the current product.

To determine if the 30 %-reduced-salt cracker delivers a saltiness intensity that is not lower than the current cracker, 68 experienced assessors are recruited to perform a directional paired comparison test for similarity. The objective is to show that the salt intensity from the 30 % reduced salt cracker is only considered as lower than the current cracker by "few" assessors, if any. The team decides that a maximum of 20 % (p_d = 0,2) of the assessors should be able to correctly detect the difference, i.e. the probability for a correct answer should be lower than p = 0,6.

Thirty of these 68 assessors (46 %) select the current cracker as "saltier" than the 30 % reduced salt cracker. This proportion is significantly lower than the acceptable proportion of 0,6 (one-sided p = 0,006, one-sided upper confidence bound 0,55) as the current cracker.

The team proceeds to a consumer test to determine if the claim "same great taste" can be used.

Ninety consumers of the current crackers blind taste both the 30 %-reduced-salt cracker and the current cracker and indicate if each has a "great taste" on a five-point scale. The top two categories are considered as agreement on "great taste", and the success criterion is to show that at least 2/3 of the consumers agreed with that statement. A one-sided approach is chosen for the analysis. The data from the current cracker is not used.

Sixty-five of the 90 consumers (72 %) rated the 30 % reduced salt cracker at least "4" on the five-point scale. A one-sided binomial test shows this rate not to be significantly larger than 2/3 = 0,667 (p = 0,157, lower confidence bound 0,63). The team concludes that the "30 % less salt, same great taste" claim is not supported. The "30 % less salt, same great taste" claim is not supported for the new 30 % reduced salt cracker.

A.5 Comparative — Unsurpassed — Performance: "Cleans as well as the leading brand"

A laundry detergent manufacturer has a brand of laundry detergent, Brand A, that sells in the national market. This Brand A laundry detergent is threatened by another more expensive brand from company B that advertises its superiority: "Brand B cleans better than all the other leading brands". Brand A and Brand B comprise 91 % of the laundry detergent brands sold nationally.

Researchers in the Brand A company feel their Brand A laundry detergent cleans as well as the competitor's detergent, and the company wants to make this claim in their printed materials to retailers to maintain their retail shelf space. They also want to advertise on television that "[Brand A] cleans as well as the leading brand" to defend against Brand B's superiority claim.

Brand A's success criteria are determined as:

a) in the lab, with standard soils applied, expert assessors' mean ratings on a 10-point cleanliness scale show that the two laundry detergents are comparable in cleaning performance, i.e. differ by no more than 0,5 points;

b) consumers who use Brand A and Brand B at home for two consecutive weeks rate the cleaning performance of the two laundry detergents comparably, i.e. they do not differ by more than 0,5 points on a 10-point overall cleanliness scale.

To evaluate the laboratory cleaning performance of Brand A versus Brand B, standard loads of laundry are prepared with the same soils applied. One load of washing is cleaned with a standard amount of Brand A's detergent. A second load is treated with the same soils and with the same standard amount of Brand B. The company's group of 10 expert assessors examine each load of clean clothes, blind to which detergent was used. The mean cleanliness rating for Brand B is 8,1. The mean cleanliness ratings for Brand A is 7,9. The standard error of these paired ratings is 0,4. The two one-sided tests (TOST) procedure for equivalence tests is used with an equivalence margin of δ = 0,5, yielding p values

of 0,020 1 and < 0,001. As both p values are below the significance level of 5 % based on the TOST methodology, the sensory team concludes the two detergents perform sufficiently similarly.

A sample of 150 consumers use laundry detergents Brand A and Brand B at home. Brand A users and Brand B users are represented according to the proportionate market share of the two brands and they use one of the laundry detergents for two weeks. They complete a short questionnaire evaluating the detergent's performance. After a one-week washout period where their usual detergent was used, these consumers receive the second detergent to use at home for two weeks. Consumers complete a second questionnaire about the detergent's performance. Product usage order is counterbalanced within each (user) group and the products are coded with three-digit numbers. The consumers' mean rating of the overall cleanliness of their laundry cleaned with Brand A is 7,4 on a 10-point scale. For Brand B, consumers rate the overall cleanliness of the laundry as, on average, 7,7. The standard error of the difference is 1,4. As previously, a TOST using paired t-tests was employed with an equivalence margin of δ = 0,5, yielding p = 0,041 and p < 0,001. As both p values are below the significance level of 5 %, based on the TOST methodology (which exceptionally does not require a multiplicity correction), the team concludes that Brand A performs comparably to Brand B based on the predefined non-inferiority margins.

Company A researchers conclude that Brand B is not better. A decision is made to communicate to retailers that "no other brand of clothes detergent cleans better" in the sales literature, presentations to retailers and on television.

A.6 Comparative — Superiority — Affective: "The ketchup taste consumers prefer on their burger"

A condiment company develops and launches a ketchup with a more complex flavour profile and which is more versatile than other ketchup brands in the market. Several of the large fast-food chains have adopted this ketchup as the standard condiment served on their burgers. The condiment company has done considerable consumer sensory testing during the development process and feels that their ketchup is superior in taste compared to the other main brand in the market. These two brands comprise 87 % of brands sold nationally.

The condiment company decides to test consumers' preference for their new ketchup versus the market leader. The success criteria decided by the team is that preference for the new ketchup needs to be established in a blind paired preference test between the two ketchup brands using a geographically representative sample of consumers. For this preference test, 213 consumers are asked to come to a central location taste test with hamburgers and ketchup in four different sites across the country, warranting a power of 90 % (β = 0,1) to detect a difference of p_d = 0,2 (p = 0,6) in a one-sided preference test at level α = 0,05. The four different locations represent major census regions in the country. Consumers who participate in the test eat hamburgers with ketchup regularly. Consumers are given two hamburger halves, each half with a different ketchup applied. The order of tasting of the two ketchups is balanced. Consumers are asked which ketchup they prefer for taste.

The results are that 123 preferred the new ketchup and 89 prefer the existing national brand. The p value from the one-sided binomial test is 0,010 with a lower confidence bound of 0,52. The results do meet the success criteria of achieving a preference for the new ketchup. The condiment company decides to make the claim of "the ketchup taste consumers prefer on their burger" in print and television advertisements.

A.7 Comparative — Superiority — Performance: "Less clumping than other comparably-priced brands"

A cosmetics company wants to launch a new formulation of mascara with the claim "less clumping than other comparably-priced brands". The company has developed a mascara that contains a unique ingredient that adheres to eyelashes in such a way that few if any clumps form. Therefore, there is a strong technical basis underlying the potential claim. Category sales data show that three other leading brands of mascara are comparably priced.

The success criterion is that a panel of 12 trained assessors rate the company's product significantly lower in clumping intensity compared to the other three comparably priced brands.

The company decides to perform a sensory test with 16 experienced consumers as test subjects. Each test subject applies one mascara each day following a standard protocol. Each assessor performs a visual evaluation of each subject's lashes and rates the intensity of clumping on a 10-point scale anchored at the low end by "no clumping" and at the high end by "extreme clumping". Both subjects and assessors are blind as to the brand of mascara they were applying and evaluating. The order of applying the mascaras is balanced across days. The assessors perform their evaluations of each subject within two hours of mascara application on each of the three days of the test. The results are shown in Table A.2.

Table A.2 — Results of evaluations

Parameter	New mascara	Brand B	Brand C	Brand D
Average intensity of clumping	1,8	5,4	3,8	6,2
p value (comparison of proportion against new)	—	0,004	0,022	< 0,001

A three-way analysis of variance, with the factors: product, subject and assessor, is performed on the clumping intensity data. (F_{product} = 3,4, df = 3 and 738, p = 0,017). More importantly comparing the average clumping intensity ratings of the company's product to each of the comparably priced competitors using one-sided, post-hoc t-tests for pairwise comparisons reveals that the company's mascara has significantly lower clumping intensity than each of the other three products.

Based on the results of the study, the company decides to move forward with the claim against their comparably priced competitors.

NOTE A correction for multiplicity is not required here as the category-wide claim is only made if all pairwise comparisons are simultaneously statistically significant. This is different from a situation in which the company tests against three competing brands and decides based on the test results against which competitors to make the claim. In that case, the claim would be made based on any significant results, thereby requiring a multiplicity correction such as, for example, Dunnett's multiple comparisons procedure.

A.8 Non-comparative — Attribute: "Now less bitter" or "now with more roasted flavour"

A coffee company changes its processing methods to produce a coffee that is less bitter and sour, and has more intense roasted flavour notes. The company wishes to highlight these changes to the consumer by putting either the claim "now less bitter" or the claim "now with more roasted flavour" on the packaging, if such a claim could be supported. If both claims are supported, the package can claim "now less bitter and with more roasted flavour". The sensory researcher advises the team and her descriptive analysis panel is well-versed in evaluating the sensory attributes of different coffees. The success criterion to support the claim would be that the trained descriptive panel found differences (in the direction claimed) in either the bitter or the roasted intensity between the current process coffee and the new process coffee, and that observed mean differences should be at least 1 cm on the 15 cm line scale, as earlier studies indicated this to be the threshold to warrant consumer noticeability. The overall significance level is set to 5 %.

The sensory research used 11 trained panellists to evaluate the two coffees, current process and new process on five key attributes: coffee impact, roasted, sour, sweet and bitter. The focus of the analysis to support the claim was on roasted and bitter. These five key attributes, identified in drivers of liking study done on the coffee and a competitive set done in the previous year, correlate the most with consumer liking. The two coffees were prepared in a standard manner, and the assessors evaluated the intensities of the five attributes in each coffee on a 15 cm line scale, anchored "none" to "extreme". Two trials of these blind taste evaluations were performed. The mean intensity ratings and the respective one-sided p-values obtained are listed in Table A.3.

Table A.3 — Mean intensity ratings and the respective one-sided *p*-values

Parameter	Coffee impact	Roasted	Sour	Sweet	Bitter
New process coffee	8,8	7,9	4,7	5,6	7,1
Current process coffee	8,1	6,4	6,1	5,3	7,9
p value	0,092	0,016	0,006	0,538	0,037
p value (Bonferroni correction)	—	0,032	—	—	0,074

Data were analysed using analysis of variance using the product, assessors and product assessor interaction as independent variables. One-sided tests on products were performed at a significance level of 5 %. As the claim to be used is decided upon based on test results, a correction for multiplicity is required. Here, the Bonferroni correction was applied on the two focal attributes, namely roasted and bitter. Corrected *p* values (multiplied by two as two attributes were tested for a potential claim) are provided in Table A.3. Alternatively, the two (uncorrected) *p* values of primary interest can be compared against the halved significance level of 0,025.

The results show that the new coffee has significantly higher roasted intensity (mean difference 1,5 cm). However, after multiplicity correction, bitter is not statistically significantly lower for the new versus the old process coffee (corrected *p* value 0,074 > 0,05, the uncorrected *p* value of 0,037 > 0,025), and the mean difference is only 0,8 cm. Therefore, the claim “less bitter” cannot be supported, but the claim “more roasted flavour” can be. Marketing begins preparing the new packaging materials with the “now with more roasted flavour” claim.

Annex B
(informative)

Type I error rates as a function of the number of statistical tests in a study

Table B.1 shows how the alpha (Type 1 error rate) increases with the number of product or attribute comparisons within the same study. This increase in error rates with an increasing number of statistical tests means that the probability to observe at least one false positive (significant) result increases substantially with the number of tests performed. Thus, it becomes more likely to conclude a difference exists when, in fact, it does not exist for any attribute.

Table B.1 is a concrete illustration of what is meant by "multiplicity"[7]. Users of descriptive analysis to support a claim should be aware of this effect. If, for example, 10 products or attributes are tested at a significance level of 5 % each, the probability for at least one of them to become statistically significant just by chance (i.e. even if the products do not differ at all on any of these attributes) is about 40 %.

The best way to mitigate multiplicity effects in sensory testing is to establish the reproducibility or reliability of a product or an attribute difference and the strict identification of primary endpoints prior to execution of a study, or to provide other evidence that supports a finding of product or attribute differences. It is worth noting that the same effect occurs whenever multiple tests are performed: multiple pairwise product comparisons, multiple breakouts by subgroups (e.g. age, gender) or comparisons of multiple attributes for the same products.

Table B.1 was derived using the following line of code in R 3.6.3: cbind(x<-1:30, round(1-0.95^x,2), round(1-0.99^x,2)).

Table B.1 — Changes in probabilities as a function of increases in comparisons (on products or attributes) in a study

Number of statistical tests (on products or attributes)	Significance level for each comparison		Number of statistical tests (on products or attributes)	Significance level for each comparison	
1	0,05	0,01	16	0,56	0,15
2	0,10	0,02	17	0,58	0,16
3	0,14	0,03	18	0,60	0,17
4	0,19	0,04	19	0,62	0,17
5	0,23	0,05	20	0,64	0,18
6	0,26	0,06	21	0,66	0,19
7	0,30	0,07	22	0,68	0,20
8	0,34	0,08	23	0,69	0,21
9	0,37	0,09	24	0,71	0,21
10	0,40	0,10	25	0,72	0,22
11	0,43	0,10	26	0,74	0,23
12	0,46	0,11	27	0,75	0,24
13	0,49	0,12	28	0,76	0,25
14	0,51	0,13	29	0,77	0,25
15	0,54	0,14	30	0,79	0,26

Bibliography

[1] ISO/IEC Guide 51:2014, *Safety aspects — Guidelines for their inclusion in standards*

[2] ISO 6658, *Sensory analysis — Methodology — General guidance*

[3] ISO 11136, *Sensory analysis — Methodology — General guidance for conducting hedonic tests with consumers in a controlled area*

[4] ASTM E1958, *Standard Guide for Sensory Claim Substantiation*

[5] Declaration of Helsinki. *Bulletin of the World Health Organization*. 2001, **79**(4), p. 373

[6] Corbin R. et al. *A Practical Guide to Comparative Advertising: Dare to Compare*. Elsevier, 2018

[7] Martin G. Munchausen's statistical grid, which makes all trials significant. *The Lancet*. 1984, 22;2(8417–8418), p. 1457

[8] Advertising Standards of Canada. *Guidelines for the Use of Comparative Advertising: Guidelines for the Use of Research and Survey Data to Support Comparative Advertising Claims*, 2010

[9] Schneider-Häder B., Hamacher E., Beeren C. *Sensory Claims – Methodological approach to development and substantiation*. DLG-Expert report 15/2015. DLG e.V., Frankfurt, 2015. Available from: https://www.dlg.org/fileadmin/downloads/lebensmittel/themen/publikationen/expertenwissen/lebensmittelsensorik/e_2015_15_Expertenwissen_SensoryClaims.pdf

[10] ASAI. *Code of Standards for Advertising and Marketing Communications in Ireland*. Advertising Standards Authority for Ireland (ASAI), Dublin, Ireland

[11] NAD: http://www.asrcreviews.org/asrc-procedures

[12] UK. Advertising Standards Authority (ASA) and Committees of Advertising Practice (CAP) guidelines: https://www.asa.org.uk/news/a-quick-guide-to-comparative-advertising.html#.VVDGXvCy7-I with specific codes at: https://www.asa.org.uk/codes-and-rulings/advertising-codes.html

[13] NBC Universal Advertising Standards. Advertising Guidelines: https://nbcuadstandards.com/files/NBC_Network_Advertising_Guidelines.pdf

[14] ABC Television Network. Advertising Standards and Guidelines: https://abcallaccess.com/app/uploads/2016/01/2014-Advertising-Guidelines-.pdf

[15] Committee of Advertising Practice (CAP): www.cap.rg.uk

Bureau of Indian Standards

BIS is a statutory institution established under the *Bureau of Indian Standards Act*, 2016 to promote harmonious development of the activities of standardization, marking and quality certification of goods and attending to connected matters in the country.

Copyright

Review of Indian Standards

Amendments are issued to standards as the need arises on the basis of comments. Standards are also reviewed periodically; a standard along with amendments is reaffirmed when such review indicates that no changes are needed; if the review indicates that changes are needed, it is taken up for revision. Users of Indian Standards should ascertain that they are in possession of the latest amendments or edition by referring to the website- www.bis.gov.in or www.standardsbis.in

This Indian Standard has been developed from Doc No.: FAD 28 (21628).

Amendments Issued Since Publication

Amend No.	Date of Issue	Text Affected

BUREAU OF INDIAN STANDARDS

Headquarters:

Manak Bhavan, 9 Bahadur Shah Zafar Marg, New Delhi 110002
Telephones: 2323 0131, 2323 3375, 2323 9402 *Website*: www.bis.gov.in

Regional Offices:		*Telephones*
Central	: 601/A, Konnectus Tower -1, 6th Floor, DMRC Building, Bhavbhuti Marg, New Delhi 110002	2323 7617
Eastern	: 8th Floor, Plot No 7/7 & 7/8, CP Block, Sector V, Salt Lake, Kolkata, West Bengal 700091	2367 0012 2320 9474
Northern	: Plot No. 4-A, Sector 27-B, Madhya Marg, Chandigarh 160019	265 9930
Southern	: C.I.T. Campus, IV Cross Road, Taramani, Chennai 600113	2254 1442 2254 1216
Western	: Plot No. E-9, Road No.-8, MIDC, Andheri (East), Mumbai 400093	2821 8093

Branches : AHMEDABAD. BENGALURU. BHOPAL. BHUBANESHWAR. CHANDIGARH. CHENNAI. COIMBATORE. DEHRADUN. DELHI. FARIDABAD. GHAZIABAD. GUWAHATI. HIMACHAL PRADESH. HUBLI. HYDERABAD. JAIPUR. JAMMU & KASHMIR. JAMSHEDPUR. KOCHI. KOLKATA. LUCKNOW. MADURAI. MUMBAI. NAGPUR. NOIDA. PANIPAT. PATNA. PUNE. RAIPUR. RAJKOT. SURAT. VISAKHAPATNAM.

Published by BIS, New Delhi