Review on characteristics of trained sensory panels in food science

Ilija Djekic, José Lorenzo, Paulo Munekata, Mohammed Gagaoua, Igor Tomasevic

▶ To cite this version:

Ilija Djekic, José Lorenzo, Paulo Munekata, Mohammed Gagaoua, Igor Tomasevic. Review on characteristics of trained sensory panels in food science. Journal of Texture Studies, 2021, 52 (4), pp.501-509. 10.1111/jtxs.12616 . hal-04156785

Title: Review on characteristics of trained sensory panels in food science

Authors: Ilija Djekic[a], José M. Lorenzo[b,c], Paulo E. S. Munekata[b] , Mohammed Gagaoua[d], Igor Tomasevic[a]

Running Title: **Sensory panels in food articles**

Article Classification: **Review paper**

Affiliation:

[a] Faculty of Agriculture, University of Belgrade, Nemanjina 6, 11080 Belgrade – Zemun, Serbia

[b] Centro Tecnológico de la Carne de Galicia, rúa Galicia no. 4, Parque Tecnológico de Galicia, San Cibrao das Viñas, 32900 Ourense, Spain

[c] Área de Tecnología de los Alimentos, Facultad de Ciencias de Ourense, Universidad de Vigo, 32004 Ourense, Spain

[d] Food Quality and Sensory Science Department, Teagasc Food Research Centre, Ashtown, Dublin 15, Ireland

Corresponding author:

Dr Ilija Djekic | Full professor
Faculty of Agriculture - University of Belgrade
Nemanjina 6, 11080 | Belgrade – Zemun | Republic of Serbia
Phone: +381 11 4413427 | Cell: +381 65 512 7848
E-mail: idjekic@agrif.bg.ac.rs | idjekic@mts.rs
Scopus Author ID: 57195311310 | ORCID ID: 0000-0002-8132-8299

Abstract:
Sensory analysis has been, is, and will be one of the most important methods in judging food quality. As such, it is an evaluation tool involving human subjects with specific skills to conduct assigned series of tests. This review outlines main characteristics of 179 trained panels published in 16 selected scientific journals in the last 12 months, as well as training methods used for panelists, and type of sensory studies employed. The results reveal that two thirds of the panels have between eight and twelve members, with gender data provided in half of the papers. Overall duration of their initial training is presented only in around 20% of reviewed publications. When provided, duration was below 2 hours per session involving up to 10 sessions. One third of papers confirmed to have conducted training of the panel for methods employed, while almost half used experienced human subjects with no further data. Around 12% of all manuscripts have validated training of their sensory panel, while 20% of papers covered at least one criterion for assessment of their panels' performance. The majority of papers (80%) used descriptive methods, mainly with intensity scales. It is of note that 15% of papers used hedonic tests typical for consumer studies. Almost half of the scholars conducted their research in triplicates (41.3%) while almost one quarter (24%) provided no data on this subject. Type of food analyzed has no effects on the quality of data provided regarding panels, training, sensory methods, and replications.

Keywords: sensory evaluation; panelists' performance; panel training validation; sensory methods

1. Introduction

Sensory analysis is considered as an essential tool in judging food quality (Schiano, Harwood, & Drake, 2017) as taste, texture, flavor and odor have a significant role in food choice and dietary patterns (Drewnowski, 1997). In general there are two types of sensory methods used: (i) analytical ones comprising of difference / discriminative tests and descriptive tests and (ii) affective, divided into acceptance/rating and preference/choice based tests (Meilgaard, Carr, & Civille, 2006). Analytical sensory tests define food profiles and sample variability of food attributes but do not take into account any liking considerations (Schiano et al., 2017). Discrimination tests are considered as "user friendly" since they easily distinguish differences between two or more samples by using reference standards and binomial statistical distributions (Lawless & Heymann, 2010). On the other hand, descriptive analysis is used to profile all important sensory attributes of food but requires trained panel consistent and sensitive to act as an instrument (Chambers, Allison, & Chambers, 2004). Regardless of the type of analytical test, existence of a sensory panel is needed consisting of two phases: training phase and testing phase (Peltier, Mammasse, Visalli, Cordelle, & Schlich, 2018). Acceptance tests are considered as consumer-based, focused to untrained population targeting their food preferences (Schiano et al., 2017).

For the purpose of performing sensory analysis, it is important to use a trained sensory panel for objective evaluation of different sensory food properties (Meilgaard et al., 2006). These panelists should be trained on how to distinguish and evaluate each sensory attribute that is of interest for the sensory study (Bratcher, 2013). Also, when designing an experiment with sensory panels included, it is of great importance to define the number of panelists, necessary skills and number of replications (Talsma, 2018). Some authors, like Gómez-Corona, Pohlenz, Cayeux, and Valentin (2020) emphasize that in addition to sensory ability of the panelists, aspects such as memory created by the training panel during training sessions also influence their performance.

Factors in recruiting panelists cover their health status including potential allergies for different types of food and dietary habits; their availability for training sessions and / or testing of food products; verbal creativity in explaining sensory sensations; concentration during sensory testing, previous (sensory evaluation) experience (Murray, Delahunty, & Baxter, 2001). For some food products it may be necessary to have the panelists undergo a medical screening prior to participating in the sensory study (Lawless & Heymann, 2010). When sensory analysis includes oral processing of food, factors like (non)usage of dental wearers (Chen, 2009) and selection of panelists with body mass index in the normal range of 18–25 kg/m^2 (Djekic, Ilic, Lorenzo, & Tomasevic, 2021; Forde, Van Kuijk, Thaler, De Graaf, & Martin, 2013) is necessary or at least advisable. Regarding health issues, specific anosmia or color vision deficiencies may also become a rejecting criteria for performing sensory analysis, like in the research of Tomasevic et al. (2019) when panelists had to successfully pass Ishihara's Test for Color Deficiency to proceed in evaluating color of meat products.

Training of the sensory panel has the potential to improve the panel's internal consensus, and ensure repeatability and discriminative power (Chollet & Valentin, 2001; Wolters & Allchurch, 1994). Depending on type of sensory assessments that are to be used for research activities, it is recommended to develop a set of different tests associated with food products and methods employed (Lawless & Heymann, 2010). Besides tests, typical training procedure should also cover sufficient number of training hours (Chambers et al., 2004) and panel performance evaluation (Labbe, Rytz, & Hugi, 2004). Training of a sensory panel should enable panelists to become familiar with the method and the vocabulary used, improve discrimination ability and enable consensus achievement (Labbe et al., 2004).

Thus, this review aimed to provide an insight into trained sensory panels used across different types of food research studies, covering demographic characteristics of the panelists as well as the main methods of their training, validation, and performance assessment. As an added value, this review further discusses sensory techniques used and number of replications considered among studies published in the last 12 months. Final objective was to provide recommendations to food scholars associated with the use of trained sensory panels.

2. Materials and methods

In order to perform a literature review, it was obvious at first glance that key word "sensory analysis" will raise hundreds of thousands of publications (i.e., in ScienceDirect search engine, over 425,000 results were reported). As such this topic is scientifically dispersed in a very heterogenous way (research and review articles, book chapters, conference papers, editorials, etc.), and to emphasize the attention on sensory panels used for performing sensory analysis of food products, we focused our survey to 16 selected scientific journals covering papers with sensory analyses of food. The selection of scientific journals was based on maturity of the journal, its impact factor, its scope (falling within food science and technology where sensory analysis is only one of many different physic-chemical, microbial and instrumental methods) and indexing by international repositories including Web of Science, Scopus, ScienceDirect, Wiley and MDPI (Spreafico, Russo, & Rizzi, 2017). Journals "Food Quality and Preferences" and "Journal of sensory studies" were deliberately left out from the survey since their main scope is sensory analysis and because materials and methods are significantly better explained in these two sensory-specific journals compared to other journals (Lestringant, Delarue, & Heymann, 2019).

After selecting "sensory analysis" as a key word, the papers were sorted by relevance and only "research articles" were included. Finally, to make the literature review most current, we refined the search by year (only papers indexed in 2020 and 2021 were analyzed) and processed papers that occurred on the first page of the search database (presenting 20-30 most relevant and latest papers). Such a search revealed 395 papers. When the literature review began, papers using "untrained panels", "semi-trained panels" and "consumer panels" were further excluded. The final set comprised of 174 papers with 179 trained panels. Figure 1 depicts the 16 targeted journals and number of papers selected per journal. For the purpose of performing data analysis and comparisons of categories or count data associated with type of food tested (animal origin, plant origin, beverages and other types of food), χ^2 tests were used similar to review performed by Lestringant et al. (2019). Figure 2 provides all sequential steps used to perform this review.

3. Results and discussion

3.1 Demographic structure of sensory panels

Table 1 provides basic information on the demographics of the panels. Almost two thirds had between eight and twelve panelists, 3.4% of the papers missed to provide this data. When it comes to age and gender, it seemed that more than half of the papers did not report any information. From the ones that gave data, the majority of them provided range of years (38%) rather than average age of the panelists (3.3%). When it comes to gender, number of females in trained sensory panels were slightly prevailing although most of the sensory panels didn't follow any "gender balance" procedure. There are conflicting studies associated with gender perception of tastes from the ones confirming differences (Hirokawa, Yamazawa, & Shimizu, 2006) to the ones failing to raise such a conclusion (Yang, Williamson, Hasted, & Hort, 2020). Giovannelli et al. (2016) assume gender differences in the processing of time associated with olfactory stimulation while Proserpio, Laureati, Invitti, Cattaneo, and Pagliarini (2017) associate body-mass index and gender as combining factors influencing differences.

Descriptive panels usually consist of 8-16 panelists that are selected and trained in a proper way (Tabary, Miège, Brémaud, Carvalho, & Vincenzi, 2021; Wolters & Allchurch, 1994). A recent review on milk narrows this range from 6 to 12 (Schiano et al., 2017). Lestringant et al. (2019) calculated in a review covering papers published between 2010 and 2015, that in average 11 panelists are used in sensory analysis using descriptive methods. It is worthy to note that the ISO 12399 standard recommends to use 12 panelists (ISO, 2016). Overall, there were no significant differences associated with size of panel and type of food analyzed ($p > 0.05$; χ^2 test).

The analysis of geographic coverage shows that 16 papers (8.9%) had co-authors from three or more countries, 34 papers (19.0%) with co-authors from two countries and 129 papers (72.1%) with authors from the same country. Table 2 summarizes the geographic distribution of affiliations. Some authors, like Garcia-Bailo, Toguri, Eny, and El-Sohemy (2009), indicate that cultural / national background of a panelist may influence taste perception, since common polymorphisms in genes that are engaged in taste perception are in charge of interindividual differences associated with food preferences within and in-between populations. Also, Yang et al. (2020) confirmed that ethnicity may be one of influential factors in perceiving intensity of five basic tastes. When it comes to different cross-cultural panels engaged, this factor can be pronounced if higher level of variability between panels exists (Teo et al., 2018). However, there were no data clarifying potential distribution of panelists from different countries.

Accordingly, Gagaoua and co-workers reported in the field of beef sensory quality traits including tenderness that the evaluation depends on the country origin of the panelists (Gagaoua et al., 2016). However, the authors who also considered the end-point cooking temperature of meat showed that irrespective of the sensory protocol, trained panelists from France and the United Kingdom expressed the same perceptions for tenderness and juiciness when assessing beef cooked to 55°C or 74 °C. The authors explained these differences by the different habits and preferences of the panelists in each country. Indeed, another phenomenon known as the "halo effect" is an important factor to consider as it was described to explain differences among panelists and their origin. On another hand, Gagaoua and co-workers further revealed in another study that the associations between protein biomarkers and beef tenderness depend not only on the know factors impacting final scores of tenderness such as animal type, gender, breeds, muscle, rearing practices but also on how sensory scores by panelists are determined, which further depend on the country of origin of the panelists (Gagaoua, Terlouw, Richardson, Hocquette, & Picard, 2019).

3.2 Training of the sensory panel

Table 3 depicts training data obtained from the analyzed database. Issue that draws attention is that in over 80% of the studies, no data are provided regarding duration of training and number of sessions. This is opposed to a review report covering descriptive analysis in the period 2010-2015 where below 20% of the studies omitted training details regarding their duration (Lestringant et al., 2019).

From the ones that provided information, training consisted of up to 10 sessions with one session lasting around one and a half hour. There is no strict rule how long should panel trainings last since they depend on many factors - previous experience of the panel and each individual panelist including the sensory method used. Recommendation vary from 10 to over 100 hours (Dairou & Sieffermann, 2002; Meilgaard et al., 2006), but main criteria should be made based on panel performance assessment (Labbe et al., 2004), rather than excuses in terms of time/resource constraints. Lestringant et al. (2019) categorize different types of panel training as familiarization (1 session); short – long trainings (consisting of several session and covering a period of up to six months) and continual training.

When it comes to training contents, only one third of studies reported that they have trained their panels for the sensory methods that were to be used for the research. In general, most of the reported training for descriptive analysis followed guidelines outlined in ISO 13299 (ISO, 2016) developing sensory vocabulary, raising agreements on descriptors and sensory profiling and reference used. This enables reaching consensus among panelists in terms of definitions and corresponding reference standards which best describe attributes and differences between the samples. Sensory profiling is considered as a descriptive method that both qualifies and quantifies sensory properties of food (Labbe et al., 2004). For these tests extensive and in-depths tests are needed compared to applying simple discrimination tests (Lawless & Heymann, 2010).

When it comes to 'experience and competence' of the panel in almost half of the studies (44.1%) the only stated criterion was that "panel was trained". Different principles that were reported comprised of number of years serving as a panelist, sum of training hours (in the near or distant past), or possession of certificates associated mostly with wine, coffee or olive oil testing. Lestringant et al. (2019) categorize panel selection for descriptive methods based on: (i) sensory performance, (ii) prior experience, (iii) professional expertise, with the fourth category defined as „no information provided ". However, this criterion may be categorized more as "panel selection" rather than "panel training".

Standard ISO 8586 has been identified as an additional training criterion in 15% of analyzed papers (mentioned joint with training methods, as competence criteria or per se). However no data were provided about parts of the standards that have been used: training of tastes/odors, detection of specific tastes/odors, use of scales/ranks/scores, development of descriptors and other issues outlined in the standard (ISO, 2012a). Finally, 6.7% papers provided no data on any type of training. There were no differences associated with training methods used and type of food analyzed ($p > 0.05$; χ^2 test).

When using trained panelists, Tabary et al. (2021) recommend verification of sensory panels at least once a year. Since various authors highlighted ISO 8586 as criteria for justifying competence of the sensory panel, it is of note that this standard highlights maintenance of skills through selective verifications twice a year with ideal scenario of weekly participations, monthly meetings and required retraining of panelists after six or more weeks (ISO, 2012a). None of the papers reported these conditions.

Finally, an interesting pattern has been observed in 12 papers providing no data on training panels but refereeing to some previous studies. However, tracking back to those studies still provide limited to no data since those papers were also refereeing to the older ones and as such, traceability of tracking backwards didn't justify the purpose.

3.3 Training validation and panelists performance

When setting a sensory panel, first step is its validation to confirm that the panel can perform and be used in different sensory studies. Also, validation of training (prior to sensory analysis), may result in exclusion of panelists due to discriminating problems. However, based on our study, limited information is provided regarding validation methods used and resulting activities applying to sensory panels. Around 12% of all papers included in this review did not mention any validation activities. In general, the following methods were used:

- Selection of panelists (and removal of the ones with unsatisfactory results) was based on different discriminating capacity, reproducibility capacity, reliability and / or individual consensus during training sessions with the development of different thresholds such as having $p < 0.05$ (or $\alpha = 0.05$) or that the results evaluated by a single panelist does not differ by more than 20% in line with Poisson and Schieberle (2008) or by setting acceptance eligibility between 60% and 80%. One-way analysis of variance (ANOVA) was used as the tool for different statistical calculations.
- In some cases, final selection of panelists (from a larger group) was performed by using triangle tests (ISO, 2004), 2-AFC (a two-alternative forced-choice) tests (ISO, 2005) or sequential analysis of different discrimination test (ISO, 2019).

Within the 12% of papers, in some of them validation was only mentioned as being performed with no specific data. Standard ISO 11132 recommends a procedure on how to validate a panel in terms of number of products (three to four), number of attributes (at least eight), significant difference between at least two products and conducting two to three evaluation sessions (ISO, 2012b). Although some authors believe that testing of panelists should be kept to a reasonable amount and avoid to over-test panelists since too many screening tests decrease enthusiasm and motivation (Lawless & Heymann, 2010), still training remains the only tool that enables consistency during sensory assessment.

Besides validating the panel (prior to sensory analysis), it is also important to assess panelists performance. Criteria for evaluating the attributes of a trained sensory panel and evaluation of the panel performance cover the following aspects (Lawless & Heymann, 2010; Talsma, 2016):

(i) is the panel capable of showing products differences / discriminate in-between samples;
(ii) are the scores of panelists reliable (in-between replicates and over time in case of evaluating products over time);
(iii) are results valid in terms of visible consensus between panelists and scoring in a similar way; and
(iv) are they able to specify specific sensory attributes and sensations.

Outcome of this review shows that similar to validation of the panel, below 20% of the analyzed manuscripts mentioned assessment performance of the panel and results. In general, the following were applied:

- Use of basic descriptive statistics such as mean and the corresponding standard deviation, subject's bias, variability and variation coefficient for panel performance assessment.

- Use of statistical analysis, mainly ANOVA in three combinations: one-way (among panelists); two-way (panelists *versus* products), and three-way (panelists *versus* products *versus* sessions). This aligns to the generalizability theory where four factors influence data obtained from a sensory panel – products, panelists, replications and interaction in-between them (Talsma, 2016).
- Performance assessment covering analysis of maximum two of the following principles: reproducibility, repeatability, discriminatory ability and homogeneity of the panel.
- Overall, except using ANOVA, two software were mentioned – © XLSTAT and © PanelCheck.

Standard ISO 11132 outlines all four criteria as of equal importance: discriminability (linked with differences between products), homogeneity (consensus of the panel), repeatability (within sessions), and reproducibility (between session) as well as two-way ANOVA for panel performance (discrimination, homogeneity and repeatability) and one-way ANOVA for assessor performance (discrimination and repeatability) (ISO, 2012b). Tabary et al. (2021) recognize the four criteria as key performance indicators in evaluating the quality of a panel while Talsma (2016) stresses reliability analysis as most beneficial when performing quantitative descriptive analysis and/or sensory profiling studies.

3.4 Sensory methods used

Based on the analyzed studies, 179 panels evaluated four groups of products: plant origin products including both fresh and processed fruits and vegetables covering 40% of the sampled papers. Animal origin food has a similar share of 34.6% comprising of all types of meat and dairy products. Beverages were in the focus of sensory panels in 20% of the selected papers analyzing coffee, juices, and alcoholic drinks. Finally, the last group named “other” consisted of different gels, semi-solid food or specific protein enriched products with the share below of 10%.

When it comes to sensory methods, analytical tests both descriptive and difference have prevailed (Figure 3), although some papers reported use of various acceptance tests typical for consumer studies. As presented in Figure 3, the most dominant type of sensory method used by panelists is descriptive analysis (over 80%). This type of analysis is considered as one of the main methodologies due to the use of rating scores allocated by the panelists by using a “unstructured scale” or a “n”-point scale (Bi, 2003). Typical methods reported were Flavor Profile (Cairncross & Sjöström, 1997), Quantitative Descriptive Analysis (Stone, Sidel, Oliver, Woolsey, & Singleton, 2008) and generic descriptive analysis (Lawless & Heymann, 2010) that are in line with a review on descriptive analysis methods used in sensory analysis performed by Lestringant et al. (2019).

Within the analyzed papers, two thirds used intensity scale up to 10 points. Among this group of tests, occasionally scholars used temporal dominance of sensation, catch-all-that-apply and quality grading. Since tasting is a dynamic activity, when data are collected simultaneously, use of check that apply or temporal dominance of sensation is gaining more attention (Visalli, Mahieu, Thomas, & Schlich, 2020). When it comes to quality grading, it is important to define quality parameters measured and the use of a scientifically validated method (Feria-Morales, 2002).

Besides descriptive methods, limited number of papers reported use of difference tests, namely duo-trio analyzing if a perceptible sensory difference or similarity exists between two samples (ISO, 2017), alternative forced choice test (2-AFC) determining if a difference or similarity between two samples exists associated with the intensity of a sensory attribute forcing choice between two alternatives (ISO, 2005) and ranking tests allowing evaluation of differences between several samples focused on intensity of one or several attributes or of an overall impression (ISO, 2006).

It is interesting to see that 15.6% of papers used different acceptance tests, mainly hedonic scale with up to 9 intensity scores or other types of hedonic tests. However, this type of tests is most directed to consumers asking them to evaluate food and score how much they like it (Wichchukit & O'Mahony, 2015; Xia et al., 2021). Although there are papers that extrapolate consumer likings using acceptance tests and sensory panels (Schiano et al., 2017), many papers confirm large deficiencies in the capability of trained panelists to envisage the preferences of consumers as discussed half a century ago by Ellis (1969) to one of the latest research of Maheeka, Godfrey, Ellis, and Hort (2021) confirming that consumer and expert panels can’t be considered as interchangeable and depend on the objective of the sensory study. There were no differences associated with type of sensory method used and type of food analyzed ($p > 0.05$; χ^2 test).

When designing a sensory experiment, researchers are confronted with two opposed criteria: resources (in terms of number of panelists, products, sessions and associated costs) and quality of data striving them to be as high as possible (Pagès & Périnel, 2004). Table 4 gives an overview of number of replicates that have been reported in the studies. Sensory experiments in triplicates were mostly stated (41.3%) while almost one quarter (24%) did not report any data on replications.

When it comes to discussing optimal number of replicates, an interesting debate has been initiated upon publication of Meyners, Carr, and Hasted (2020) assuming that in many cases one replicate is typically sufficient. Rationale for such a conclusion is that if a trained panel is used, further replications add limited value to the research. This has been opposed in four commentaries where in one Berget (2020) states that "sensory descriptive analysis without replication will have less power" since replicates enable various quality controls of variations between and within panelists. Christensen (2020) also argues that trials with replications are needed not only for analyzing between to within panelists, but also for utilizing future research. In one of previous review on number of replicates, Peltier et al. (2018) state that one replicate may be sufficient for getting a picture of the data, two replicates are optimal for trained panels with more than two replicates being useless. Moser, Lepage, Pineau, Fillion, and Rytz (2018) highlight well-trained panel as the one and only criteria in decreasing number of replicates to one assessment, based on their study with Nestlé R&D panel. Still, due to possible lack of data on training capacity of the panel, ISO 12399 procedure recommends two replicates in order to obtain robust average estimation of the data (ISO, 2016). Based on the number of replicates reported, our survey showed an average of 2.69 ± 0.98, which is in concurrence with research from Lestringant et al. (2019) calculating 2.7. There were no differences associated with number of replicates and type of food analyzed ($p > 0.05$; χ^2 test).

Besides discussing the optimal number of replicates, one must take into account also total number of products presented to each panelists recommending five (Neilson, Ferguson, & Kendall, 1988) to maximum of 15 (Leight & Warren, 1988), considering various other criteria such as evaluation methods employed, complexity of food matrices or motivation of panelists (Pagès & Périnel, 2004) and sensory fatigue (Sauvageot, 1990).

3.5 Practical implication of the study

Results of this study provide valuable insight for various stakeholders in the life cycle of sensory study publications. First it may assist researchers in designing experiments as well as providing necessary data (in the "materials and methods" section) associated with their sensory panels regarding demographic characteristics of the panel, selection of panelists, their training and assessment performance. In parallel, it can provide help to journal editors and reviewers when evaluating sections linked with sensory analyses since it is often used in majority of food-related studies. Unfortunately, similar trend of failing to provide information needed to replicate and/or interpret instrumental color data has just been reported by (Tomasevic, Djekic, Font-i-Furnols, Terjung, & Lorenzo, 2021).

4.0 Conclusion

This review revealed the importance of providing data associated with sensory panels engaged in analyzing various characteristics of food. No pattern has been observed in terms of type of food analyzed and quality of data provided regarding panels, training, sensory methods and replications. This leads to a conclusion that lack of some important data became a generic trend in food science when sensory methods are employed as part of a much wider food research.

For explaining sensory panels, three types of data should be provided: (i) demographic characteristics of the panel, (ii) recruitment and training characteristics, (iii) performance analysis results.

Regarding panel size, optimal number should be between eight and 12. It is required to provide data on gender and age while preferable data are the health status of the panelists and their nationality. Since some papers confirm potential differences in sensory perception between different nationalities and/or gender, need for this type of information is even pronounced.

Mandatory data that in most studies were missing are associated with selection and training of the panel. When it comes to the recruitment phase, it is common to use experienced panelists. However, regardless of the years of experience of the panelists, performance prescreening is appreciated. Training phase should provide data on training duration and number of sessions. At least one session should be dedicated to developing sensory vocabulary, including definition of descriptors and references used. Duration of this phase including number of sessions and length of each session is recommendation to be minimum 10h. When selecting panelists, it is advisable to verify detection of target tastes/odors, along with training for the methods used. As a final step, it is important to validate the panel, and if necessary, exclude panelists with low sensory performance. Type of validation method employed should also be clearly stated.

Finally, besides validation of the panel (performed prior to sensory analysis), it is important to assess the panel performance after testing to verify its capability once more in terms of results obtained. This concurs with the need to clearly state number of replicates as the minimal number should be kept at two. Finally, some scholars used hedonic tests typical for consumer studies, instead of descriptive methods, without any rationale for such a choice.

Ethical statements and data sharing

Conflict of Interest - All authors of this manuscript declare that they do not have any conflict of interest.

Data Availability Statement - The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Berget, I. (2020). Comment to the paper: To replicate or not to replicate, or when did we start to ignore the concept of statistical power? *Food Quality and Preference, 79*, 103632. doi:https://doi.org/10.1016/j.foodqual.2019.01.004

Bi, J. (2003). Agreement and reliability assessments for performance of sensory descriptive panel. *J. Sens. Stud., 18*(1), 61-76. doi:https://doi.org/10.1111/j.1745-459X.2003.tb00373.x

Bratcher, C. L. (2013). Trained Sensory Panels *The Science of Meat Quality* (pp. 207-213).

Cairncross, S., & Sjöström, L. (1997). Flavor profiles: a new approach to flavor problems. *Descriptive Sensory Analysis in Practice, 1522*.

Chambers, D. H., Allison, A. M. A., & Chambers, I. V. E. (2004). Training effects on performance of descriptive panelists. *J. Sens. Stud., 19*(6), 486-499.

Chen, J. (2009). Food oral processing—A review. *Food Hydrocolloids, 23*(1), 1-25.

Chollet, S., & Valentin, D. (2001). Impact of training on beer flavor perception and description: are trained and untrained subjects really different? *J. Sens. Stud., 16*(6), 601-618.

Christensen, R. H. B. (2020). Commentary on "To replicate or not to replicate, or when did we start to ignore the concept of statistical power?" by Meyners, Carr and Hasted. *Food Quality and Preference, 79*, 103650. doi:https://doi.org/10.1016/j.foodqual.2019.01.022

Dairou, V., & Sieffermann, J. M. (2002). A comparison of 14 jams characterized by conventional profile and a quick original method, the flash profile. *J. Food Sci., 67*(2), 826-834.

Djekic, I., Ilic, J., Lorenzo, J. M., & Tomasevic, I. (2021). How do culinary methods affect quality and oral processing characteristics of pork ham? *Journal of Texture Studies, n/a*(n/a). doi:https://doi.org/10.1111/jtxs.12557

Drewnowski, A. (1997). Taste preferences and food intake. *Annu Rev Nutr, 17*(1), 237-253.

Ellis, B. H. (1969). Acceptance and consumer preference testing. *J. Dairy Sci., 52*(6), 823-831.
Feria-Morales, A. M. (2002). Examining the case of green coffee to illustrate the limitations of grading systems/expert tasters in sensory evaluation for quality control. *Food Quality and Preference, 13*(6), 355-367. doi:https://doi.org/10.1016/S0950-3293(02)00028-9
Forde, C., Van Kuijk, N., Thaler, T., De Graaf, C., & Martin, N. (2013). Oral processing characteristics of solid savoury meal components, and relationship with food composition, sensory attributes and expected satiation. *Appetite, 60*, 208-219.
Gagaoua, M., Micol, D., Picard, B., Terlouw, C. E. M., Moloney, A. P., Juin, H., Meteau, K., Scollan, N., Richardson, I., & Hocquette, J.-F. (2016). Inter-laboratory assessment by trained panelists from France and the United Kingdom of beef cooked at two different end-point temperatures. *Meat Science, 122*, 90-96. doi:https://doi.org/10.1016/j.meatsci.2016.07.026
Gagaoua, M., Terlouw, C., Richardson, I., Hocquette, J.-F., & Picard, B. (2019). The associations between proteomic biomarkers and beef tenderness depend on the end-point cooking temperature, the country origin of the panelists and breed. *Meat Science, 157*, 107871. doi:https://doi.org/10.1016/j.meatsci.2019.06.007
Garcia-Bailo, B., Toguri, C., Eny, K. M., & El-Sohemy, A. (2009). Genetic variation in taste and its influence on food selection. *OMICS A Journal of Integrative Biology, 13*(1), 69-80.
Giovannelli, F., Giganti, F., Saviozzi, A., Rebai, M., Marzi, T., Righi, S., Tramacere, L., Borgheresi, A., Cincotta, M., & Viggiano, M. P. (2016). Gender Differences in Time Perception During Olfactory Stimulation. *J. Sens. Stud., 31*(1), 61-69. doi:https://doi.org/10.1111/joss.12191
Gómez-Corona, C., Pohlenz, A., Cayeux, I., & Valentin, D. (2020). Panel performance and memory in visually impaired versus sighted panels. *Food Quality and Preference, 80*, 103807. doi:https://doi.org/10.1016/j.foodqual.2019.103807
Hirokawa, K., Yamazawa, K., & Shimizu, H. (2006). An examination of sex and masculinity/femininity as related to the taste sensitivity of Japanese students. *Sex roles, 55*(5-6), 429-433.
ISO. (2004). ISO 4120:2004 Sensory analysis — Methodology — Triangle test. Geneva, Switzerland: International Organization for Standardization.
ISO. (2005). ISO 5495:2005 Sensory analysis — Methodology — Paired comparison test. Geneva, Switzerland: International Organization for Standardization.
ISO. (2006). ISO 8587:2006 Sensory analysis — Methodology — Ranking. Geneva, Switzerland: International Organization for Standardization.
ISO. (2012a). ISO 8586:2012 Sensory analysis — General guidelines for the selection, training and monitoring of selected assessors and expert sensory assessors. Geneva, Switzerland: International Organization for Standardization.
ISO. (2012b). ISO 11132:2012 Sensory analysis — Methodology — Guidelines for monitoring the performance of a quantitative sensory panel. Geneva, Switzerland: International Organization for Standardization.
ISO. (2016). ISO 13299:2016 Sensory analysis — Methodology — General guidance for establishing a sensory profile. Geneva, Switzerland: International Organization for Standardization.
ISO. (2017). ISO 10399:2017 Sensory analysis — Methodology — Duo-trio test. Geneva, Switzerland: International Organization for Standardization.
ISO. (2019). ISO 16820:2019 Sensory analysis — Methodology — Sequential analysis. Geneva, Switzerland: International Organization for Standardization.
Labbe, D., Rytz, A., & Hugi, A. (2004). Training is a critical step to obtain reliable product profiles in a real food industry context. *Food Quality and Preference, 15*(4), 341-348. doi:https://doi.org/10.1016/S0950-3293(03)00081-8
Lawless, H. T., & Heymann, H. (2010). *Sensory Evaluation of Food - principles and practices*. New York, USA: Springer Science+Business Media, LLC
Leight, R. S., & Warren, C. B. (1988). Standing panels using magnitude estimation for research and product development. *Applied sensory analysis of foods, 1*, 225-249.
Lestringant, P., Delarue, J., & Heymann, H. (2019). 2010–2015: How have conventional descriptive analysis methods really been used? A systematic review of publications. *Food Quality and Preference, 71*, 1-7. doi:https://doi.org/10.1016/j.foodqual.2018.05.011

Maheeka, W. N. R. P., Godfrey, A. J. R., Ellis, A., & Hort, J. (2021). Comparing temporal sensory product profile data obtained from expert and consumer panels and evaluating the value of a multiple sip TCATA approach. *Food Quality and Preference, 89*, 104141. doi:https://doi.org/10.1016/j.foodqual.2020.104141

Meilgaard, M. C., Carr, B. T., & Civille, G. V. (2006). *Sensory evaluation techniques*: CRC press.

Meyners, M., Carr, B. T., & Hasted, A. (2020). To replicate or not to replicate, or when did we start to ignore the concept of statistical power? *Food Quality and Preference, 79*, 103633. doi:https://doi.org/10.1016/j.foodqual.2019.01.005

Moser, M., Lepage, M., Pineau, N., Fillion, L., & Rytz, A. (2018). Replicates in sensory profiling: Quantification of the impact of moving from two to one assessments. *Food Quality and Preference, 65*, 185-190. doi:https://doi.org/10.1016/j.foodqual.2017.12.002

Murray, J., Delahunty, C., & Baxter, I. (2001). Descriptive sensory analysis: past, present and future. *Food Res. Int., 34*(6), 461-471.

Neilson, A. J., Ferguson, V. B., & Kendall, D. A. (1988). Profile methods: flavor profile and profile attribute analysis. *Applied sensory analysis of foods, 1*, 21-41.

Pagès, J., & Périnel, E. (2004). Panel performance and number of evaluations in a descriptive sensory study. *J. Sens. Stud., 19*(4), 273-291. doi:https://doi.org/10.1111/j.1745-459X.2004.tb00148.x

Peltier, C., Mammasse, N., Visalli, M., Cordelle, S., & Schlich, P. (2018). Do we need to replicate in sensory profiling studies? *Food Quality and Preference, 63*, 129-134. doi:https://doi.org/10.1016/j.foodqual.2017.09.001

Poisson, L., & Schieberle, P. (2008). Characterization of the Key Aroma Compounds in an American Bourbon Whisky by Quantitative Measurements, Aroma Recombination, and Omission Studies. *J. Agric. Food. Chem., 56*(14), 5820-5826. doi:10.1021/jf800383v

Proserpio, C., Laureati, M., Invitti, C., Cattaneo, C., & Pagliarini, E. (2017). BMI and gender related differences in cross-modal interaction and liking of sensory stimuli. *Food Quality and Preference, 56*, 49-54. doi:https://doi.org/10.1016/j.foodqual.2016.09.011

Sauvageot, F. (1990). Fatigue in sensory evaluation: Myth or reality? *Developments in food science*.

Schiano, A., Harwood, W., & Drake, M. (2017). A 100-year review: Sensory analysis of milk. *J. Dairy Sci., 100*(12), 9966-9986.

Spreafico, C., Russo, D., & Rizzi, C. (2017). A state-of-the-art review of FMEA/FMECA including patents. *Computer Science Review, 25*(Supplement C), 19-28. doi:https://doi.org/10.1016/j.cosrev.2017.05.002

Stone, H., Sidel, J., Oliver, S., Woolsey, A., & Singleton, R. C. (2008). Sensory evaluation by quantitative descriptive analysis. *Descriptive sensory analysis in practice, 28*, 23-34.

Tabary, C., Miège, M., Brémaud, D., Carvalho, L., & Vincenzi, F. (2021). A tool to help the panel leader to best monitor a sensory panel performance. *J. Sens. Stud., 36*(1), e12613. doi:https://doi.org/10.1111/joss.12613

Talsma, P. (2016). Assessing sensory panel performance using generalizability theory. *Food Quality and Preference, 47*, 3-9. doi:https://doi.org/10.1016/j.foodqual.2015.02.019

Talsma, P. (2018). How much sensory panel data do we need? *Food Quality and Preference, 67*, 3-9. doi:https://doi.org/10.1016/j.foodqual.2016.12.005

Teo, P. S., van Langeveld, A. W. B., Pol, K., Siebelink, E., de Graaf, C., Martin, C., Issanchou, S., Yan, S. W., & Mars, M. (2018). Training of a Dutch and Malaysian sensory panel to assess intensities of basic tastes and fat sensation of commonly consumed foods. *Food Quality and Preference, 65*, 49-59. doi:https://doi.org/10.1016/j.foodqual.2017.11.011

Tomasevic, I., Djekic, I., Font-i-Furnols, M., Terjung, N., & Lorenzo, M. J. (2021). Recent advances in meat color research. *Current Opinion in Food Science, 41*, 81-87.

Tomasevic, I., Tomovic, V., Milovanovic, B., Lorenzo, J., Đorđević, V., Karabasil, N., & Djekic, I. (2019). Comparison of a computer vision system vs. traditional colorimeter for color evaluation of meat products with various physical properties. *Meat Science, 148*, 5-12. doi:https://doi.org/10.1016/j.meatsci.2018.09.015

Visalli, M., Mahieu, B., Thomas, A., & Schlich, P. (2020). Concurrent vs. retrospective temporal data collection: Attack-evolution-finish as a simplification of Temporal Dominance of Sensations? *Food Quality and Preference, 85*, 103956. doi:https://doi.org/10.1016/j.foodqual.2020.103956

Wichchukit, S., & O'Mahony, M. (2015). The 9-point hedonic scale and hedonic ranking in food science: some reappraisals and alternatives. *J. Sci. Food Agric., 95*(11), 2167-2178. doi:https://doi.org/10.1002/jsfa.6993

Wolters, C., & Allchurch, E. (1994). Effect of training procedure on the performance of descriptive panels. *Food Quality and Preference, 5*(3), 203-214.

Xia, Y., De Mingo, N., Mendez Martín, J., Bodeau, J., Perret, M., Zhong, F., & O'Mahony, M. (2021). Is the absolute scaling model the basis for the 9-point hedonic scale? Evidence from Poulson’s Stimulus Range Equalizing Bias. *Food Quality and Preference, 89*, 104153. doi:https://doi.org/10.1016/j.foodqual.2020.104153

Yang, Q., Williamson, A.-M., Hasted, A., & Hort, J. (2020). Exploring the relationships between taste phenotypes, genotypes, ethnicity, gender and taste perception using Chi-square and regression tree analysis. *Food Quality and Preference, 83*, 103928. doi:https://doi.org/10.1016/j.foodqual.2020.103928

Table 1. Demographic characteristics of sensory panels

Demographic characteristic		**N (%)**
Size of trained sensory panelists	Up to 7 panelists	23 (12.8%)
	8 – 12 panelists	114 (63.7%)
	Over 12 panelists (up to 32)	36 (20.1%)
	No data	6 (3.4%)
Age of panelists	Provided average age	6 (3.3%)
	Provided range (18 – 70)	68 (38.0%)
	No data	105 (58.7%)
Gender data	Not provided	91 (51.8%)
	Provided	88 (49.2%)
	Average number of males	5 ± 3
	Average number of females	7 ± 3

Table 2. Geographic distribution of affiliations

Continent	Number of countries	Number of co-authorships
Europe	23	116
Asia	12	66
North America	2	27
South and central America	5	15
Australia*	2	15
Africa	3	6

(*) including New Zealand.

Table 3. Types and duration of training

Training of the sensory panel		**N (%)**
Overall duration of the training	1 week – 3 months	21 (11.7%)
	4 – 32 h	13 (7.3%)
	No data	145 (81.0%)
Number of sessions	Up to 10 sessions	41 (22.9%)
	Over 10 sessions	9 (5.0%)
	No data	129 (72.1%)
Duration of training session	Up to 2 hours (0.33h – 2h)	29 (16.2%) | 1.45h ± 0.49h
	No data	150 (83.8%)
Type of training	Training for the method	61 (34.1%)
	Training for the method + ISO 8586	10 (5.6%)
	ISO 8586	13 (7.3%)
	Previous experience and competence	79 (44.1%)
	Previous experience and competence + ISO 8586	4 (2.2%)
	No data	12 (6.7%)

Table 4. Number of replications

Number of replications	**N (%)**
Duplicate	49 (27.4%)
Triplicate	74 (41.3%)
Over 3 replications and/or block-design	13 (7.3%)
No data	43 (24.0%)
Total	**179 (100%)**

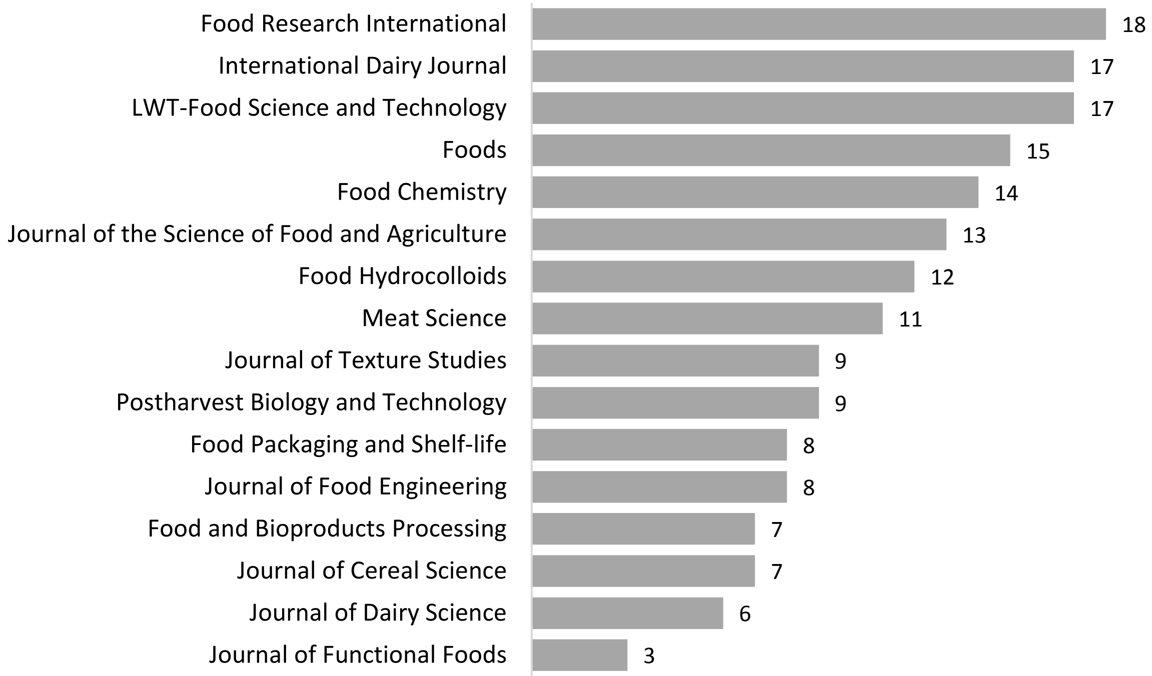

Figure 1. Selected international scientific journals and number of papers per journal.

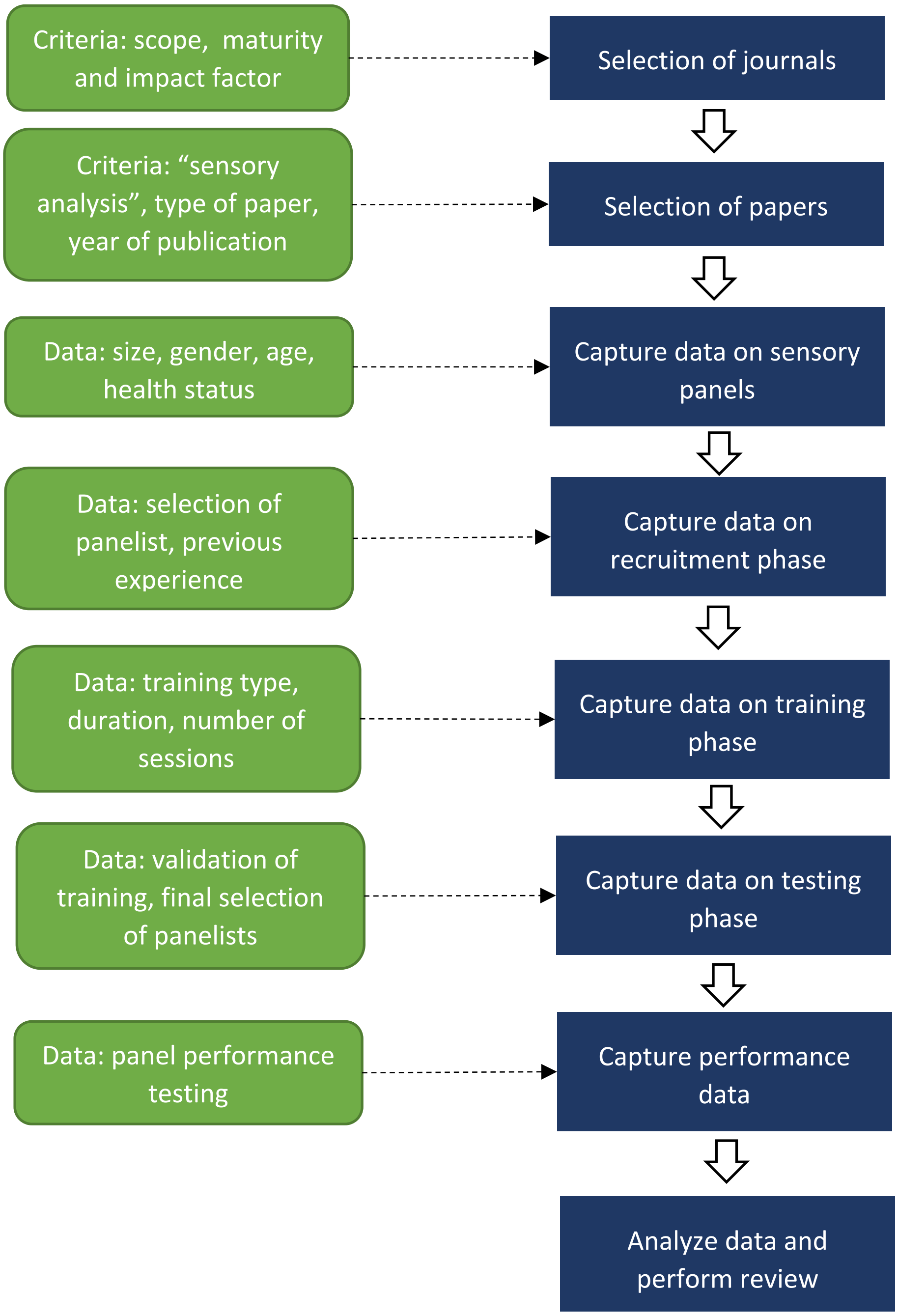

Figure 2. Methodological flowchart

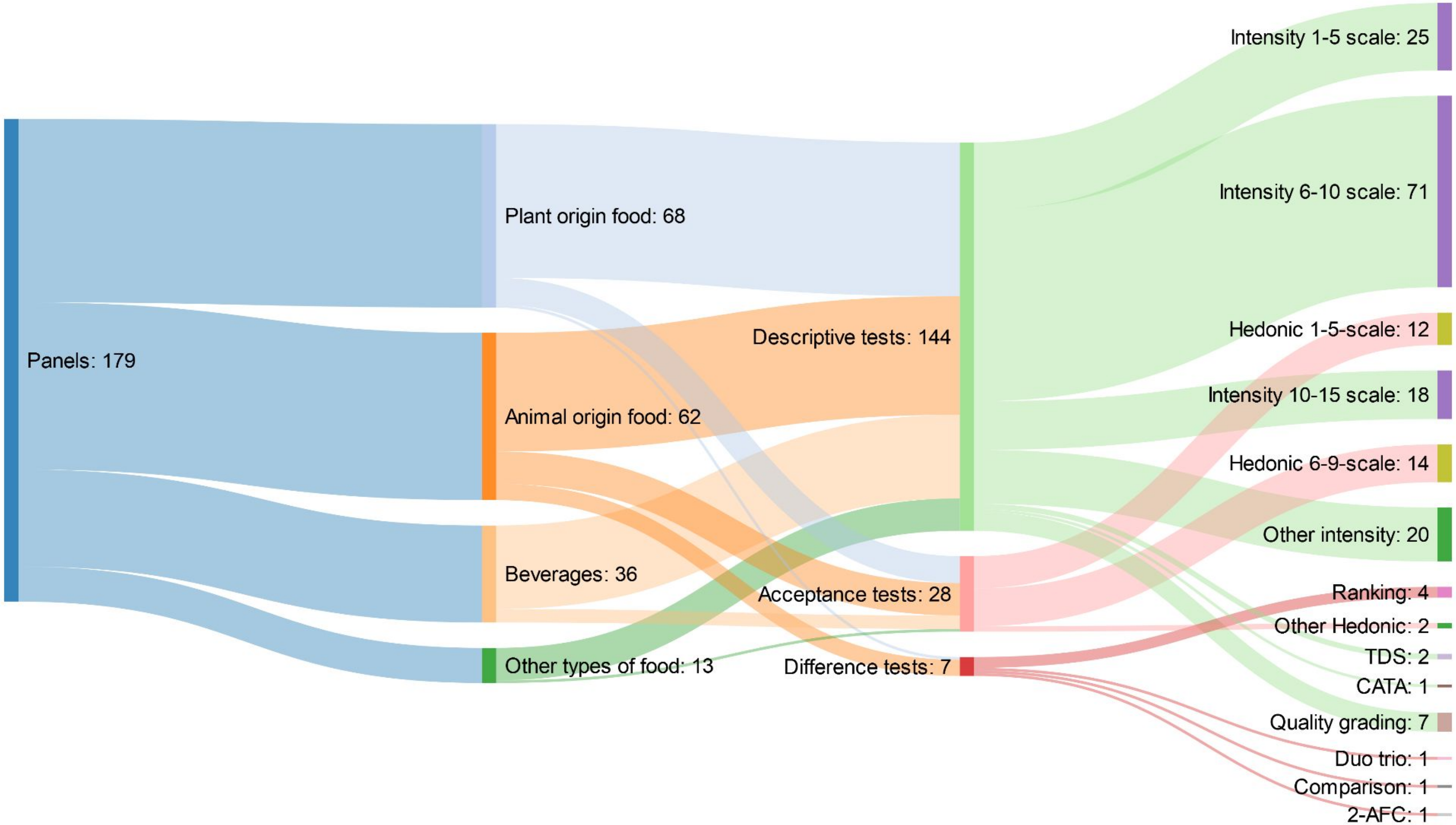

Figure 3. Sankey chart showing distribution of types of sensory methods employed by panelists