

DLG Expert report 2/2020

# Practice guide for sensory panel training

Part 3: Measurement of the performance of sensory assessors and panels



## 1. Introduction and background

The requirements in accordance with DIN EN ISO 8586:2014-05 and training plans for sensory assessor and panel qualification were presented in Parts 1 and 2 of the practice guide for sensory panel training. Part 3 will now focus on measuring the performance of individual sensory assessors and entire sensory panels with different methodological approaches. The intention of the guide is to draw attention to the essential aspects to be observed when measuring the performance of sensory assessors and panels. The backgrounds and criteria for determining performance will first be described, and selected methods for measuring and optimising it will be presented; these can be carried out even without an advanced knowledge of statistics. Case examples will be provided for a practical illustration of various applications so that specialists and managers working in the field of sensory analysis and evaluation in the food and beverage sector can obtain valuable suggestions for practical implementation. However, these must always be adapted to the respective issue in the specific company and to the sensory evaluation projects undertaken there.

Besides the availability of the technical infrastructure that is crucial for sensory analyses and adherence to 'good sensory analysis and evaluation practice', one of the most essential prerequisites for reliable sensory analysis and panel results as well as for valid data in human sensory analysis is the qualification of the sensory staff. By no means least, the high professional importance of a high-performance sensory panel that also ideally supports product development and quality assurance in the sense of food fraud and food safety (see IFS, BRC standard, etc.) in the food industry is also made clear by the speedy amortisation of capital invested in food sensory analysis (see DLG Trendmonitor 2019).

Proof of professionally qualified sensory assessors and panels whose performance is monitored additionally concerns all organisations that operate as test institutions or test laboratories for third parties, use sensory methods in their food analysis and quality monitoring and have been accredited in accordance with ISO/IEC 17025:2018-03.

## 2. Importance of the topic

Being able to trust in the reliability of the measurement equipment that is used is necessary in laboratory operations and daily work at companies. Prior to use, for example, a check has to be performed to determine whether the measuring instruments are suitable for the respective purpose, cover the measurement range and offer the necessary measurement accuracy. During practical operation, the performance of the measuring instruments has to be checked at specified or agreed intervals in order to detect possible deviations and be able to implement appropriate measures. Just like technical measurement equipment, a sensory assessor group or panel can also be regarded as an analytical measuring instrument that has to meet similar specifications.

Essentially, a measuring instrument is expected to deliver correct and precise results. **Correct results** match the true value or only deviate from it slightly (see DIN ISO 5725-2:2012). In chemical analysis, a true value can be assumed comparatively easily if a sample is analysed in which, for example, a certain quantity of table salt has been dissolved in demineralised water. Since human perception and the categorisation of a stimulus have to be taken into consideration in sensory evaluation of food, however, the mean value of several measurement results delivered by a sensory assessor group or panel is used as a substitute for the true value of an attribute, particularly in the case of common foods, which are usually complex mixtures. In sensory analysis of food, the true value therefore corresponds to a relative measurement depending on the food matrix and interaction with other sensory modalities.

The ability to achieve matching measurement results when the same sample is analysed several times is described as **repeatability or precision**. Precision must not be confused with trueness in this case (see DIN ISO 5725-2:2012). A measuring instrument is certainly able to repeatedly deliver matching results and therefore demonstrate high precision. If these precise repetitions are far removed from the true value, however, the trueness of the result is not sufficient. To clarify this, an overview of the respective technical requirements for sensory assessors will first be provided in the following before the criteria for checking and monitoring performance are described.

The detailed technical requirements for sensory assessors and panels usually vary with regard to their respective, specific area of work, with the result that training and also monitoring should be focussed on this.

While the screening used to recruit assessors for sensory analysis and evaluation is used more to fundamentally test the general sensory capabilities of the recruited sensory assessors and to determine their colour vision (or colour blindness) or also anosmia, to register odour and taste thresholds, to check their acuity in recognising differences in odour, taste or texture and to deal with their general sensory language skills and expressiveness when describing sensory sensations, the more in-depth training and qualification phase that follows on from sensory assessor selection is more specialised and aimed at the sensory assessors' subsequent areas of work.

As described in DLG Expert report Part 1 (7/2017) and Part 2 (12/2018), corresponding requirement profiles for the sensory assessors and their sensory performance must be defined based on their respective areas of work. Based on these, specific training plans and sensory evaluation projects have to be developed and minimum requirements to be met have to be formulated to enable the specific selection of sensory assessors due to their sensory perceptions and measurement skills. In this context, it should once again be pointed out that at least (if possible) two to three times the number of possible sensory assessors should undergo screening in order to ultimately obtain the sensory panel size required to statistically validate sensory analysis results.

Figure 1 shows the most common areas in which sensory assessors and panels are used. As shown there, a distinction is generally made between the two method groups of discrimination tests and descriptive tests in terms of the respective problem definition in the case of sensory evaluation methods. There are also methods that are assigned to the discrimination tests, but use elements from the area of the descriptive methods in the form of variants:

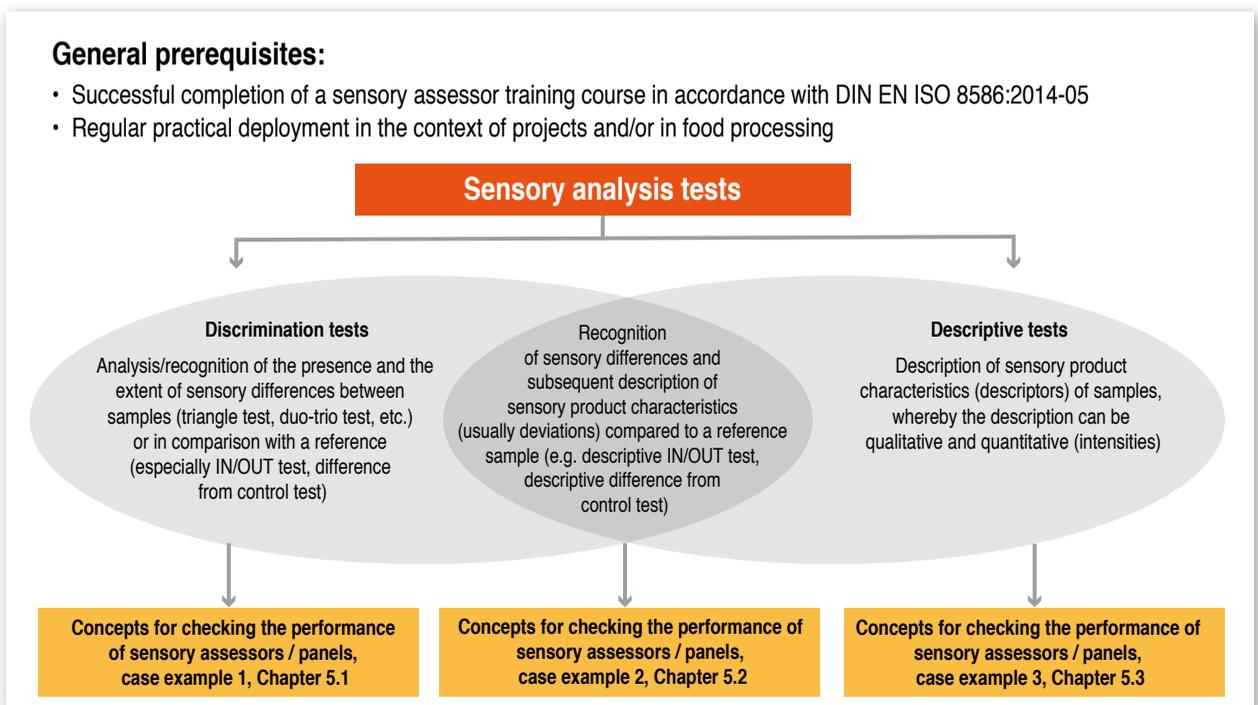


Figure 1: Areas in which sensory analysis assessors and panels work

- **Discrimination tests**

- Product comparison to determine (slight) differences that can be recognised by means of sensory analysis (e.g. Triangle Test, Duo-Trio Test, Paired Comparison Test, Ranking Test)
- Product comparison, particularly in the context of sensory quality control (e.g. IN/OUT Test, Difference from Control Test with the involvement of reference samples)

While these are primarily discrimination tests in methodological terms, they can be extended with descriptive elements from the descriptive test methods in the respective descriptive variant (descriptive IN/OUT Test or Descriptive Difference from Control Test), however.

- **Descriptive tests**

Determination and description of sensory product characteristics; the intensity attribute can additionally be tested in individual methods.

Sensory analysis methods that are focussed on objectively ascertaining sensory product differences are discrimination tests. In this process, slight sensory differences between products usually have to be recognised and named.

There are also the descriptive tests, which are used to describe sensory product characteristics (attributes, descriptors) in detail and to measure their intensity. Such sensory product profiles (often shown in the form of spider web charts) are required in order, e.g. within the context of product management, to compare competitors' products with own products, to define sensory specifications for a new product and therefore also the basis (e.g. reference samples, also referred to as 'gold standards') for quality control or to identify sensory product changes on use of alternative raw materials, in the event of modified ingredients, after long periods of product storage or as a result of interactions with the packaging.

Sensory methods that are also used, in particular, in sensory quality assurance in order to ensure the consistent quality level of the manufactured products include IN/OUT Tests (Inside/Outside Tests) or Difference from Control Tests, for example. These are primarily assigned to the discrimination tests. However, descriptive variants also exist for both of these test methods. The objective of these sensory tests is to analyse presented products in comparison with a defined reference sample (e.g. product specification, 'gold standard') and to identify possible deviations. Based on the results achieved in this process, a product rating is then also carried out in the sense of product release, rework or product blocking.

Pursuant to the DIN 10973 standard, as distinction is made between three forms of IN/OUT Test, the general [categorical], scaled and descriptive test, each of which is used to test whether the presented sample lies inside or outside of the specifications in comparison with a reference. In the descriptive variant of the IN/OUT Test, the method can 'be extended by simple profiling of the key attributes/descriptors in addition to the inside/outside assessment'. The descriptive part then follows on immediately after the discrimination part of the test. In terms of the discrimination and descriptive test elements, the same also applies to the Descriptive Difference from Control Test (DIN 10976).

While sensory assessors and panel members are trained specifically in terms of their sensory expressiveness and powers of description for their work in descriptive sensory panels, i.e. they have to internalise the identification and designation of characteristic sensory attributes/descriptors and their quantification on the basis of defined scales, focus in the qualification of discriminatory sensory panels is placed on the selective perception of what are usually slight sensory differences. High acuity of the senses and knowledge of the relevant product quality parameters of specific product categories are the essential goals of qualification in this case. Irrespective of the area of food sensory analysis and evaluation for which the sensory assessors and panels are qualified and used, regularly checking and monitoring their performance are crucial to the quality and meaningfulness of the sensory evaluation results to be delivered.

### 3. Criteria for measuring performance

Regular participation in sensory analyses as well as continuous training of the senses are required on the part of sensory assessors in order to maintain their sensory performance. To achieve this, the panel leader not only has to ensure the training, but must first and foremost continuously monitor the criteria and performance requirements that he/she has defined with regard to the necessary sensory analysis and evaluation competence in order to guarantee that the analytical test results delivered by the sensory panel are also valid. The panel leader must implement specific countermeasures (e.g. focussed training, sensory exercises) in the event of deviations from the required sensory performance. If these measures are not sufficient, sensory assessors must be excluded from certain sensory analyses and/or the composition of sensory assessor panels must be modified.

The necessity of checking the performance of sensory assessors was pointed out even in the early sensory analysis literature (see e.g. Schweizerisches Lebensmittelbuch (Swiss Food Book) since 1965, Stone and Sidel, 1985). Attention is also urgently drawn to these requirements in current works on food sensory analysis and sensory panel qualification (Kemp et al., 2009; Lawless and Heymann, 2010; Raithatha, 2018; Raithatha and Rogers, 2018).

The various authors define the following three basic requirements for sensory assessor and panel performance, to which sensory assessor training as well as the performance review and result monitoring must be oriented. The following criteria are also listed in DIN EN ISO 8586:2014-05:

- **Discrimination ability:** The sensory assessors should also be able to perceive different samples as different (i.e. recognise, identify and name differences) and scale them (i.e. measure or show the extent of the difference on scales). The underlying question is: are product differences always recognised consistently well?
- **Repeatability, comparability (precision):** In repeated measurements of the same sample under identical conditions, the sensory assessors should deliver identical or similar results and scale values (= repeatability: same sample, same sensory assessor, same place and same time, i.e. assessment is carried out in one session). Reproducibility, in which two or more assessments of the same sample under different conditions are compared with one another, can also be assigned to this requirement area (= same sample, same sensory assessor, possibly the same place, but different time, i.e. several sessions).  
The question in this case is: are the test results repeatable within defined ranges?  
(Attention: foods are subject to natural quality fluctuations.)
- **Consistency (homogeneity):** The sensory assessors should deliver data that is comparable with the data of the other sensory panel members or that lies close to the mean value of the sensory panel.  
The question in this case is: do the sensory panel members' results extensively match (within a defined range) or are there serious outliers?  
(Note: zero scatter is not possible in practice.)

These criteria for monitoring sensory performance generally apply to all sensory assessors who take part in objective sensory analyses and to the sensory panel as a whole. Irrespective of this, the criteria and factors can also be used during the sensory panel training phase in order to document the current training status.

In practical monitoring, however, the parameters underlying these criteria and their respective limit values as well as the weighting vary depending on the respective purpose for which the sensory assessors or panel are used. The focus in product development and optimisation lies more in the descriptive area, with the result that sensory assessors and panels not only have to have the ability to measure differences and intensities, but must, in particular, also be articulate and confident in their use of the product-specific sensory language (terminology/descriptors).

In methodological terms, discrimination approaches are more readily required for quality assurance (QA), i.e. sensory assessors and panels require pronounced acuity and discrimination ability. In turn, the requirements made on pure QA panels, which are used above all for IN/OUT Tests, are even more extensive and also include food technology and product knowledge competence when they are called on to identify deviations from reference samples.

#### 4. Practical implementation of the performance review for sensory assessors and panels

Various aspects have to be taken into consideration when getting the practical implementation of a performance review off to a start. Many companies almost certainly have company-specific requirements, e.g. with regard to data protection on the part of the works council or concerning the use of software tools on the part of the IT department, which can lead to restrictions in data evaluation. If the sensory assessors' individual test results therefore can not be used for purposes of performance evaluation and monitoring or also for motivating the sensory assessors (i.e. to illustrate their own sensory skills) due to data protection reasons, not even in anonymised form, and the conclusion of confidentiality agreements or similar forms of 'secure data usage' is not possible either, compressed data in the form of the evaluation of group results provides indications for identifying performance-related weaknesses and improvement potentials, albeit with limited meaningfulness. Equally, this can be used as the basis for planning training measures and implementing training courses to further improve and optimise the sensory assessor groups' sensory performance. The respective panel leader must determine which options are appropriate and permissible in this regard based on the company-specific circumstances, and must then ultimately evaluate and decide on them on an individual basis.

From a general professional point of view, a distinction can generally be made between various approaches to the practical implementation of a performance review and assessment for sensory assessors and panels (Rogers, 2018); these must be taken into consideration in the respective planning in advance:

##### 1. Analysis object:

It must be defined whether the monitoring of an individual sensory assessor's performance is to be examined or the review of a sensory panel's 'bundled' performance is to be analysed instead. These individual results can also be placed into relation with one another, meaning that a sensory assessor can be compared with a panel or a sensory panel can be compared with panels from other companies or company locations.

##### 2. Procedure:

###### a) Ongoing/new projects:

Sensory assessor and panel data from ongoing projects can be used for further evaluations pertaining to performance or results from new projects developed specifically for monitoring can be used for assessment.

###### b) Existing/modified product samples:

The product samples to be analysed can be either market samples or products from internal production. In addition, products can also be specifically modified ('spiked') for practice purposes, e.g. by reinforcing the impression of vanilla in desserts through the addition of a higher quantity of vanilla aroma or similar. The advantage of this is a detailed knowledge of the product, and it clearly ensures that there are differences in specific attributes. These are then used as key measurement variables for categorising the performance that has to be achieved by the sensory assessor or panel.

##### 3. Time period:

As shown in Table 1, the performance of individual sensory assessors or panels can be reviewed either selectively (at the end of a training unit) or also over a longer period of time within the framework of monitoring.

	Performance evaluation at a defined point in time	Performance monitoring as trend evaluations over a longer period of time
Performance evaluation within the framework of <b>regular sensory assessor and panel activity</b>	The data for selectively evaluating the performance of sensory assessors and panels can be collected from ongoing projects and specifically evaluated.	The data for selectively evaluating the performance of sensory assessors and panels can be collected from ongoing projects and specifically evaluated.
Performance evaluation within the framework of <b>specially defined performance tests</b>	Defined requirements/performance criteria are selectively checked by means of specially defined projects.	Defined requirements/performance criteria are checked in a continuous process by means of specially defined projects.

Table 1: Approaches for implementing a performance evaluation (based on Rogers, 2018)

Sensory performance has to be trained on a continuous basis. Ideally, sensory assessors are deployed regularly and only with short interruptions. In such a case, the results from ongoing projects can be used for the performance review. Developing a corresponding concept and appropriate documentation that shows the separate evaluation of data from the actual sensory analysis project is then also recommended in this regard. In addition to evaluations from ongoing projects, determining performance with additional tests, possibly within the framework of regular training courses, or also with short tests prior to the actual project sessions is also recommended.

While the performance review in the area of descriptive tests is also extensively described and standardised in DIN EN ISO 11132:2017-10, with the result that company-specific options that are adapted to the respective product category can be derived on this basis for practical implementation, the design of measures in the area of ‘discrimination tests’ is the responsibility of the respective panel leader, who has to compile corresponding information from various literature sources if necessary. Regardless of whether the sensory assessors are used in the context of discrimination tests or descriptive tests, the three criteria of discrimination ability, repeatability/reproducibility and homogeneity, as described above, always apply as the basis for reviewing performance.

If the **ability to discriminate** is to be reviewed, the sensory assessors must be provided with several similar samples to test. When selecting the samples, it must always be ensured that the differences are neither too small nor too large, as the risk of subsequently interpreting the result as (excessively) good or (excessively) poor discrimination ability is otherwise run. A sensory assessor’s or the panel’s discrimination is poor if a) no sensory difference is ever found or b) what are actually different products are evaluated as very similar in terms of their sensory attributes and/or intensities. Training courses that train both the acuity of the senses (i.e. Ranking Tests, Triangle Tests) and also the use of scales (i.e. the representation of the extent of the difference) help to improve the ability to discriminate.

The quality of the **repeatability** is clearly indicated by comparing the test results of both individual sensory assessors with themselves and also those of the sensory panel as a whole with itself. In this process, identical samples should also be rated identically when they are assessed again, i.e. identical or similar results and scale values should be delivered. ‘**Concealed double samples**’ that are integrated into a single session and presented to the sensory assessors as an equivalent product (same batch) (i.e. same sensory assessors/panels, same product, same place, same time) are suitable for reviewing this capability, for instance. ‘**Repetition sessions**’ are an additional option in this regard. In this process, the sensory assessors are required to complete several similar sessions, i.e. the defined sensory tests with the same samples (same batch, otherwise comparable samples) are carried out again at certain intervals over a long period of time (i.e. same sensory assessors/panels, same product, same place, different time).

**Good homogeneity** is ensured if the degree to which the sensory assessments of the assessors match compared to those of other assessors and the result achieved by the sensory panel as a whole is high. This can be clearly determined by means of a data evaluation.

There is also the option of data comparison with other sensory panels in this context. Such **round robin tests** (proficiency tests) are conducted to review the respectively used analysis method and to compare the results of one's own sensory panels with those of other panels. Round robin tests, which have already been established in chemical analytics for some time now in the context of good laboratory practice (GLP), can be organised on a product-specific and intracompany (e.g. comparison of sensory panels at various company locations) or industry-specific (e.g. cross-company comparison of sensory panels) basis in food sensory analysis. To do this, the organising body sends uniform samples for all sensory assessors; these have to be analysed according to the same previously defined test methods within a specified period of time by the participating sensory panels (i.e. different sensory assessors/panels, same products, different places, different times). The submitted results can be evaluated according to the previously described criteria to measure the performance of a sensory panel.

## 5. Case examples for the assessment of sensory assessor and panel performance

The use of specially developed, professional software programmes in sensory analysis offers the advantage that all data for the evaluation is available in digital form and can therefore be immediately further processed. Software programmes such as, e.g. FIZZ or Compusense offer the option of storing the analysis process digitally, depending on methodology, and help to avoid input and transfer errors at the same time, e.g. by compelling the input of values (forced choice) and only enabling further processing once a response has been forthcoming. Digital data collected in this way forms the basis for statistically evaluating sensory analysis methods on the one hand and, on the other hand, is also the basis for the statistical evaluations used to assess sensory assessor and panel performance. Many different statistical methods exist for evaluating sensory data. By themselves, no single one of these methods and no single manner of graphical representation is able to comprehensively process all of the available information contained in 'sensory analysis data'. A combination of various points of view and perspectives, which should be generated using various statistical evaluations and graphical representations, is always required.

The following case examples from the fields of 'discrimination tests', 'discrimination and descriptive tests in the sense of Descriptive Difference from Control or Descriptive IN/OUT Tests' and 'descriptive tests' are intended, with the integration of various evaluation tools as examples, to provide an insight into the options available for monitoring sensory assessor and panel performance. The company or the responsible panel leader must always check which options should be used for the specific company. Insofar as the use of corresponding sensory analysis software programmes that are available on the market is not (yet) possible within the company, simple evaluation options are also available to panel leaders by using the standard version of MS Excel or XLSTAT, a more detailed and chargeable add-in module in MS Excel.

### 5.1 Case example 1: 'discrimination tests' using the example of the Triangle Test

The fundamentally high acuity of the individual sensory assessors and their ability to identify differences between test samples is particularly important for sensory panels, especially those that use discrimination tests. A high ability to discriminate on the part of individual sensory assessors results in a high ability to discriminate on the part of the entire sensory panel.

In order to better assess the performance of sensory assessors and panels, it is appropriate to carry out discrimination tests, e.g. Triangle Tests, Paired Comparison Tests, Duo-Trio Tests, etc., with different degrees of difficulty. The ability of the sensory assessors to correctly identify differences can be recognised when the individual results are analysed. Data sets of entire sensory panels are usually distributed binomially and can be evaluated with the aid of the tables shown in the respective standards or by using statistical programmes such as the add-in software XLSTAT for MS Excel, for example.

Summary of selected options:	
Test	Triangle Test
Number of assessors	15
Prop. of correct answers	0.533
Guessing probability	0.333
Sensory discrimination tests:	
Prop. of discrim.	0.300
z	1.590
p-value	0.056
alpha	0.050
Power	0.401
Test interpretation:	
H0: The two products are similar.	
Ha: The two products are different.	
As the computed p-value is greater than the significance level $\alpha = 0.05$ , one cannot reject the null hypothesis H0.	
The risk to reject the null hypothesis H0 while it is true is 5.60 %.	

Table 2: Example of triangle test evaluation using XLSTAT (guessing model), 15 sensory assessors, 8 correct responses

Summary of selected options:	
Test	Triangle Test
Number of assessors	21
Prop. of correct answers	0.524
Guessing probability	0.333
Sensory discrimination tests:	
Prop. of discrim.	0.286
z	1.793
p-value	0.037
alpha	0.050
Power	0.415
Test interpretation:	
H0: The two products are similar.	
Ha: The two products are different.	
As the computed p-value is lower than the significance level $\alpha = 0.05$ , one should reject the null hypothesis H0, and accept the alternative hypothesis Ha.	
The risk to reject the null hypothesis H0 while it is true is lower than 3.65 %.	

Table 3: Example of triangle test evaluation using XLSTAT (guessing model), 21 sensory assessors, 11 correct responses

Two examples of the evaluation of discriminatory data sets and the interpretation of the results that are obtained are described in the following. Specifically, this involves the results of a Triangle Test (DIN EN ISO 4120). The number of correctly solved tests is calculated as a percentage per test for the sensory assessors and panel, assuming a probability of error of  $\alpha = 0.05$ .

Table 2 shows that the p value is greater than 0.05 (5% probability of error) and is 0.056. This means that the difference between the test samples was not recognised significantly. In this case, the risk of identifying a difference although none exists is 5.6% and the proportion of discriminators  $p_d$  is 30%.

Table 3 shows that the p value is less than 0.05 (5% probability of error) and is 0.037. This means that the difference between the test samples was recognised significantly. In this case, the risk of identifying a difference although none exists is 3.7% and the proportion of discriminators  $p_d$  is 28.6%.

If, for example, triangle tests are then carried out in the same form in several sessions over a long period of time, the monitoring diagram shown in Figure 2 could be produced; it shows the continued professional development of a sensory panel's discrimination ability based on the percentage of correct responses. In this example,  $n = 21$  sensory assessors completed a Triangle Test in five identical sessions over a period of five weeks. On average throughout all of the tests, 69% of the responses were correct (red line). This means that the discrimination performance of the sensory assessors was successively increased by means of systematic training.

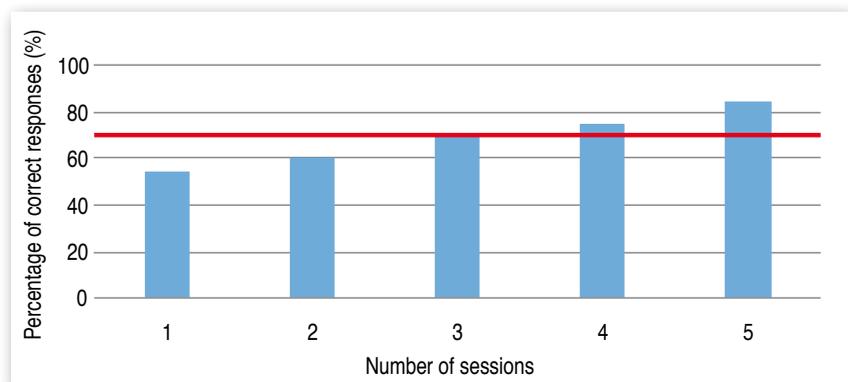


Figure 2: Monitoring the discrimination ability of a sensory panel

## 5.2 Case example 2: ‚discrimination tests and descriptive tests‘ using the example of the ‚Descriptive Difference from Control Test‘

In the context of a Descriptive Difference from Control Test for beverages (see DLG Expert report 2-2013 and 2-2013 as well as DIN 10976:2016-08), in which a sensory panel consisting of four sensory assessors each analysed ten presented product samples based on a 6-point difference scale (ratings from 1 = no deviation, full match to 6 = very large deviation, not tolerable), the results shown in Figures 3 and 4 were achieved. While the assessments of the four trained and experienced sensory assessors in example 1 (Figure 3) were relatively homogeneous, i.e. the deviations of sensory assessor 1's results from the group result are around  $\pm 0.5$  points, the situation in example 2 (Figure 4) is different. In part, the group results deviate significantly from those of sensory assessor 2 here.

The latter is a new sensory assessor and may possibly not yet be very skilled in either professionally assessing the products or in using the scale. There is a need for training in this case either in terms of using the scale or in determining sensory differences.

### 5.3 Case example 3: ‚descriptive test‘

Explanations regarding the monitoring of sensory assessors and panels used in descriptive tests are published in DIN EN ISO 11132:2017:10. The procedure described here necessitates detailed knowledge of statistics and the use of software that simplifies or automates data processing and result determination. The examples from the PanelCheck software that are shown in 5.3.2 are part of the explanations in this standard. Sensory panel performance evaluations should generally be oriented to these specifications. The examples described in the following using MS Excel are also suitable for newcomers to this topic or for panel leaders who would like to obtain a quick overview of their descriptive panel's performance or that of the individual sensory assessors, whereby the sensory assessor and panel performance criteria from DIN EN ISO 8586:2014-05, i.e. the ability to discriminate, repeatability/reproducibility (precision) and consistency (homogeneity), are regarded as the basis.

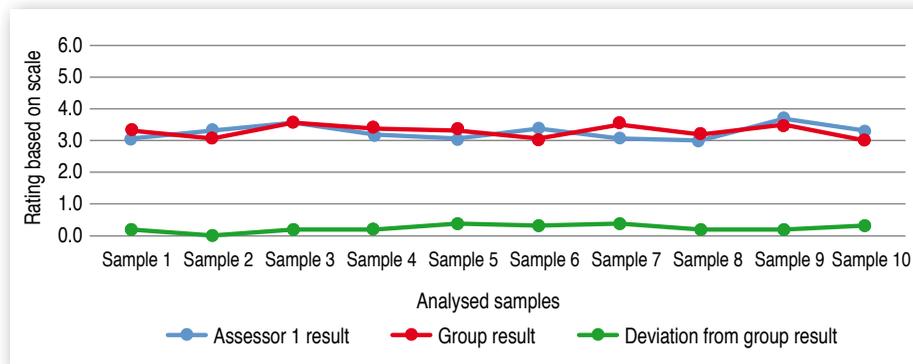


Figure 3: Example 1: sensory assessor 'Difference from Control Test' monitoring with MS Excel;  $n = 4$  experienced sensory assessors; 10 beverage samples were rated

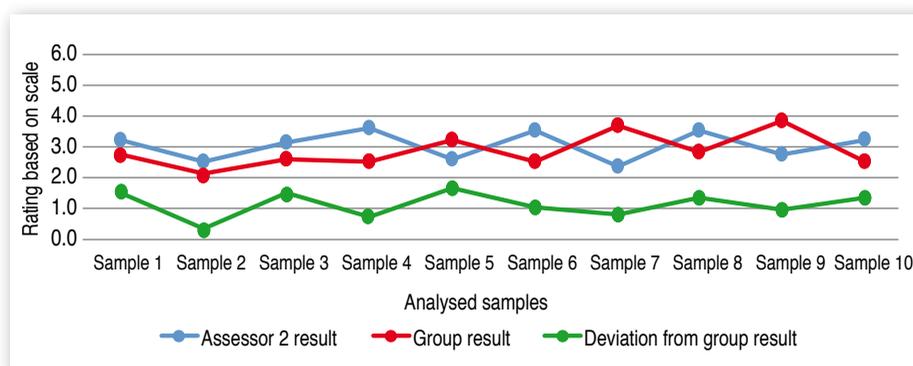


Figure 4: Example 2: sensory assessor 'Difference from Control Test' monitoring with MS Excel;  $n = 4$  sensory assessors including 3 experienced assessors and 1 new assessor; 10 beverage samples were rated

### 5.3.1 Evaluations and visualisations using MS Excel

#### 5.3.1.1 Ability to discriminate

The use of familiar reference samples is suitable for checking whether the sensory assessors are also able to perceive different samples as different (i.e. recognise, identify and name differences) and to scale them (i.e. measure or show the extent of the difference on scales). In accordance with the procedure described in DIN EN ISO 13299:2016-09 for creating a 'profile of the deviation from the reference standard', the samples to be analysed are provided in pairs. One is referred to as the reference, the other as sample 1. In this example, the taste of sample 1 was modified by the addition of salt. The test director is familiar with the product profiles of both samples, and can therefore assess the sensory assessor's ratings and the ultimate test result. The sensory assessor is now required to describe sample 1 both qualitatively and quantitatively in comparison with the reference sample according to a specified list of descriptors and on the basis of a scale that are known to him/her. The underlying question is: are product differences recognised and described appropriately or always recognised equally well in repeat tests?

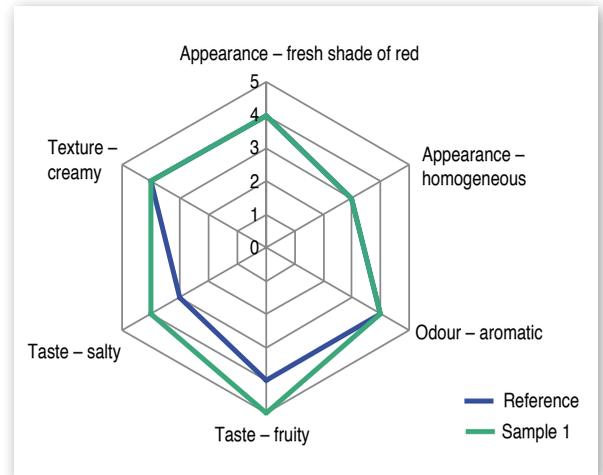


Figure 5: Sensory assessor monitoring for the profile analysis of a tomato juice; good discrimination ability of sensory assessor no. 5 (session with reference sample); evaluation with MS Excel;  $n = 1$  sensory assessor; 2 samples were rated, including 1 reference

Figure 5 shows an example of the good discrimination ability of the sensory assessor. Sensory assessor no. 5 has recognised all of the descriptors and their ratings well and has described the addition of salt in sample 1. He/she has documented the fact that this also possibly affects the fruitiness in the taste on the basis of the intensity. To check the repeatability of the results, several sessions of the same type would have to be performed and the results analysed accordingly, as described in 5.3.1.2.

#### 5.3.1.2 Repeatability/reproducibility (precision) of the sensory panel and the sensory assessors

As explained in the preceding chapters, the use of double samples or the performance of repeat tests offer various options for checking the **repeatability of the performance of a sensory panel as a group**. In this case example, eight sensory panel members test two tomato juices in accordance with the procedure described in DIN EN ISO 13299:2016-09 for creating a 'quantitative description profile'. To do this, they have defined the essential descriptors in advance and decided on the use of a scale from 0 to 5. Here, 0 means that the specified attribute is 'not recognisable' and 5 that it is 'very clearly recognisable'. The sensory panel is currently being set up and therefore consists of experienced and less experienced sensory assessors. In the following example, the first session was followed by a repeat session after two weeks (reproducibility), i.e. same product batch, same time, same sensory assessors, same place, but different time (two weeks later). In each case, the

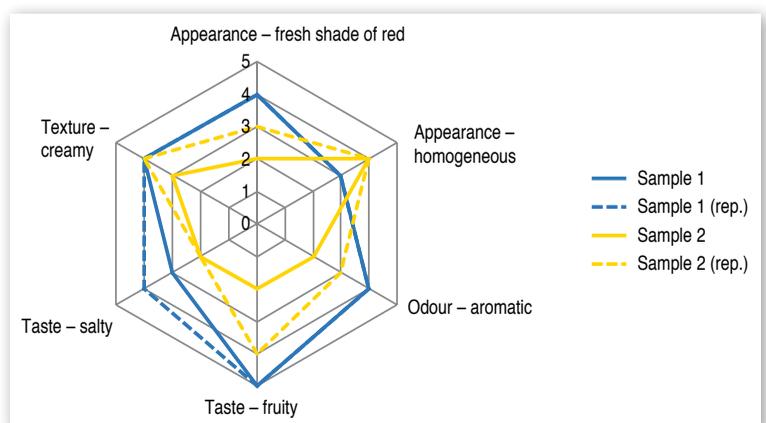


Figure 6: Sensory panel monitoring for the profile analysis of a tomato juice: repeatability of the sensory panel as a group (session 1); evaluation with MS Excel;  $n = 8$  sensory assessors, including 6 experienced assessors and 2 new assessors; 2 different samples were rated at different times

mean sensory panel value was calculated from the individual test results and visualised graphically. The better the two ratings (initial rating and repeat rating) match, the better the sensory panel's repeatability. Such charts can be used well as feedback for the sensory panel.

The results of both sessions are visualised in Figure 6, whereby the continuous lines show the mean sensory panel values for the first session and the dashed lines show the sensory panel results from the repeat session. It can be seen very clearly here that the group results differ. While the measurement result of the sensory assessors for sample 1 (blue lines) shows good repeatability (spider web lines are congruent almost everywhere), the sensory panel ratings for sample 2 (yellow lines) are more divergent. The repeatability during the repeat session is less good in this case; the two spider web lines are barely congruent. When researching the causes, it is necessary to analyse both the sensory assessors' sensory physiology and their use of scales in greater detail and to design appropriate training measures.

Using double samples in a session is appropriate for assessing performance as regards the **repeatability of sensory assessors**. The following examples in Figures 7 and 8 are dedicated to this case. Here, a total of five samples of tomato juice were presented to sensory assessor no. 3, four of which differ in sensory terms; one sample was included twice as a double sample.

Figure 7 shows the case of poor repeatability on the part of sensory assessor no. 3. This sensory assessor rated sample 1 (blue line) and sample 5 (light blue line) very differently.

Figure 8 shows the opposite case, namely the good repeatability of one sensory assessor. Sensory assessor no. 4 has rated sample 1 (blue line) and sample 5 (light blue line) virtually identically.

### 5.3.1.3 Homogeneity (consistency of the sensory panel)

In terms of homogeneity, a check must be performed to determine whether the ratings of the respective sensory assessors are comparable to those of the other sensory panel members and/or lie close to the mean panel value or whether there are any outliers. In this case example, eight sensory panel members test one tomato juice in accordance with the procedure described in DIN EN ISO 13299:2016-09 for creating a 'quantitative description profile'. To do this, they have defined the essential descriptors in advance and decided on the use of a scale from 0 to 5. Here, 0 means that the specified attribute is 'not recognisable' and 5 that it is 'very clearly recognisable'. Six of the eight sensory panel members are experienced sensory assessors, while two are still relatively new.

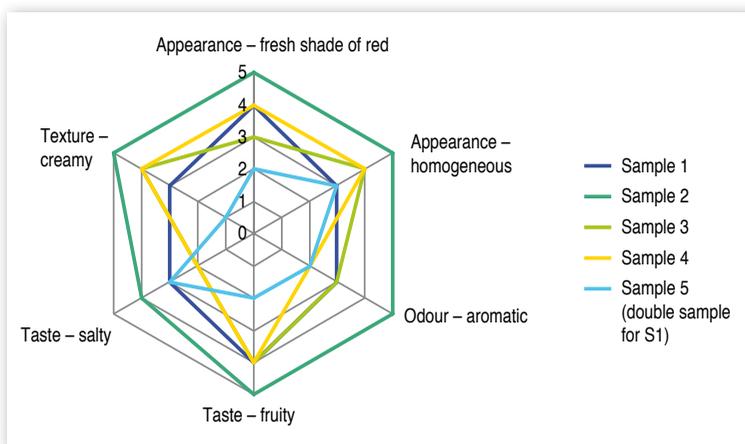


Figure 7: Sensory assessor monitoring for the profile analysis of a tomato juice: poor repeatability on the part of sensory assessor no. 3 (session with double sample); evaluation with MS Excel; n = 1 sensory assessor; 5 samples were rated, including 1 double sample

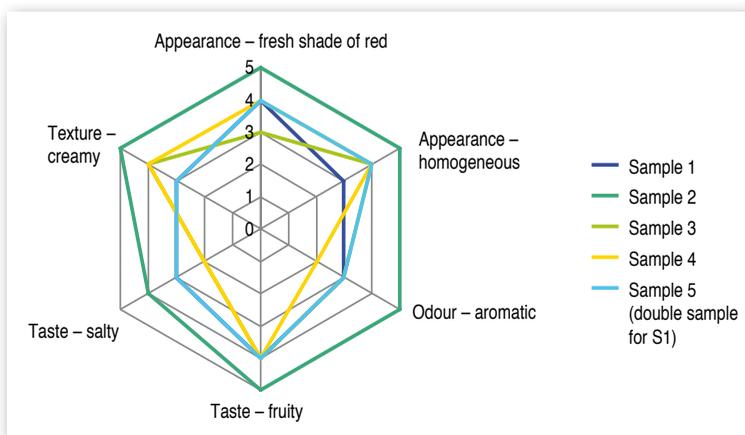


Figure 8: Sensory assessor monitoring for the profile analysis of a tomato juice: good repeatability on the part of sensory assessor no. 4 (session with double sample); evaluation with MS Excel; n = 1 sensory assessor; 5 samples were rated, including 1 double sample

Figure 9 shows the results of the individual sensory assessors and the average sensory panel result (red line) in the form of a spider web. Sensory assessor 2 and sensory assessor 4 are particularly noticeable here because their test results deviate from those of the other sensory assessors and from the sensory panel average. While sensory assessor 2 (violet line) extensively rates the product attributes as 'clearly recognisable' to 'very clearly recognisable', i.e. as very intensive (scale range 4-5), sensory assessor 4 (yellow line) perceives the intensiveness of the tomato juice sample as low to weak, because his/her ratings lie in the 0 to 2 scale range. The sensory panel is therefore not

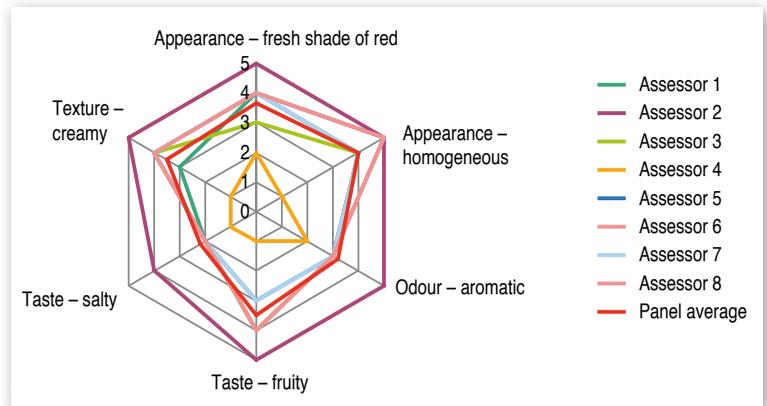


Figure 9: Sensory panel monitoring for the profile analysis of a tomato juice: homogeneity of the sensory panel

Evaluation with MS Excel;  $n = 8$  sensory assessors including 6 experienced assessors and 2 new assessors; 1 sample was rated

working homogeneously and further training courses are needed. The different ratings by sensory assessors 2 and 4 can be due to sensory physiology deficits or also to the fact that sensory assessors are not yet very confident in using the selected scale. The individual sensory assessor performance can be improved and adjusted to the sensory panel performance through corresponding screening and training of the sensory performance capability and by means of subsequent, intensive scale training. Homogeneity can be improved as a result.

The above case examples visualised in the form of spider webs using MS Excel show simple evaluation options, in part because they focus on a few products or a manageable number of descriptors. In practice, however, projects are frequently more complex, making it more difficult to visualise the data using the aforementioned visualisation form and to identify test patterns, particularly when comparing several products and when a high number of attributes and larger sensory assessor groups are involved. Multi-dimensional visualisation in a coordinate system (see PanelCheck) may then be more expedient.

### 5.3.2 Evaluation and visualisation using PanelCheck

One possible solution is the free (freeware) PanelCheck software package ([www.panelcheck.com](http://www.panelcheck.com)), which is based on the open-source statistics programme 'R' and is tailored to descriptive tests. Once the internal test data has been loaded in an electronic form that meets the programme's quality requirements (e.g. Excel), corresponding evaluations are generated by means of automated processes and on the basis of various statistical methods. The latter show an overall picture of the performance of a sensory panel or individual sensory assessors (panel members) both as numerical values and also in the form of graphical representations, and offer (advanced) panel leaders ideal support in the sensory panel management process. The programme is user-friendly and can also be used by 'non-statisticians' thanks to the automated determination of statistical parameters that is set as default. The publications of Tomic et al. (2007, 2010) offer a deeper insight into the available evaluations and statistical parameters, which are based on recommendations in current DIN EN ISO standards.

The following examples show a selection of evaluation options concerning the sensory performance of descriptive sensory assessors and panels that can be generated using PanelCheck and are recommended above all in DIN EN ISO 11132:2017:10. The figures shown here represent various sessions and therefore different case examples.

### 5.3.2.1 Discrimination ability

The **discrimination ability** of a sensory assessor or the panel can be checked using the F plots in the PanelCheck software; these can be selected accordingly by the panel leader following raw data transfer (Figure 10). The higher the F value or the further the p value lies below the specified significance level, the greater the ability of a sensory assessor to distinguish between the products. If the sensory panel as a whole reveals low F values for one or more attributes, as shown in Figure 10, this means that the sensory panel members still require further training courses for these attributes.

### 5.3.2.2 Repeatability

As has already been described, the **repeatability** shows the 'match of repeated measurements of a sample under identical conditions such as the same sensory assessors, same place and same time of the session'. This can be statistically evaluated through the ANOVA (Analysis of Variance) by means of the product/repetition interaction. In addition, plots of the mean square error (MSE values) of a single-factor ANOVA represent the ability of the individual sensory assessors to deliver reproducible results (Figure 11). The lower the MSE value is, the less the sensory assessor's rating deviates from measurement to measurement. The MSE values in Figure 11 are generally on a low level; attribute F (Hedley, Bart, Jenny) and attribute C (Clarence, Jenny) are exceptions.

The MSE values of the sensory assessors should always be interpreted in combination with the F plots, as low MSE values can also be caused by the same (non-discriminatory) rating of the samples. The standard deviations of the mean sensory panel values for the sessions conducted under comparable conditions are compared in order to assess the **reproducibility** (correspondence of two or more assessments carried out on the same sample under different conditions).

The explanations so far have clearly shown that the performance of a descriptive sensory panel cannot be determined with just one parameter, but that various analyses have to be carried out to achieve this. Besides the plots shown above, PanelCheck also offers a function for visualising the discrimination ability (P plot) and the repeatability (MSE plot) of the sensory panel members in one single plot. The  $p^*$ MSE plot shown in Figure 12 is used for this. In this case, it is desirable that the p values and MSE values are each low and are located as close as possible to the bottom left corner in the plot. According to Figure 12, this means that the sensory panel members Jenny and Clarence reveal sub-optimal discrimination ability and repeatability.

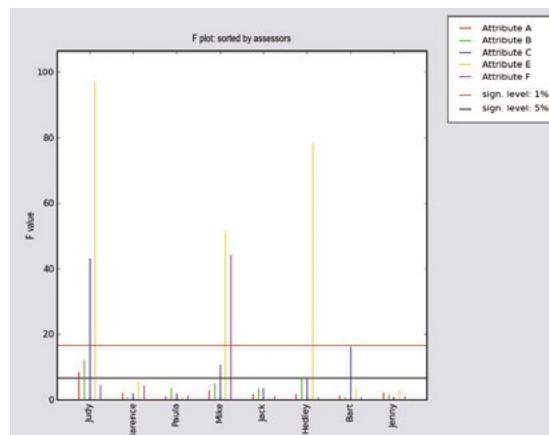


Figure 10: F plot (evaluation from PanelCheck)

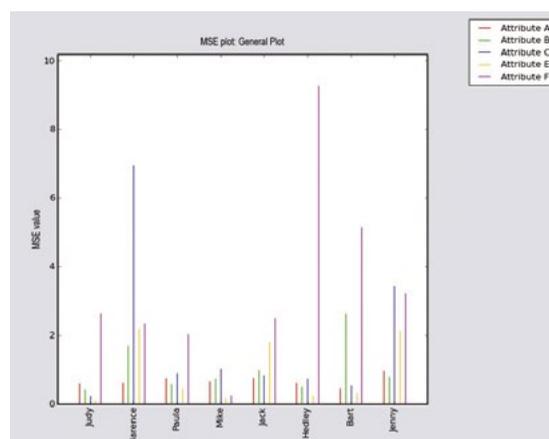


Figure 11: MSE plot (evaluation from PanelCheck)

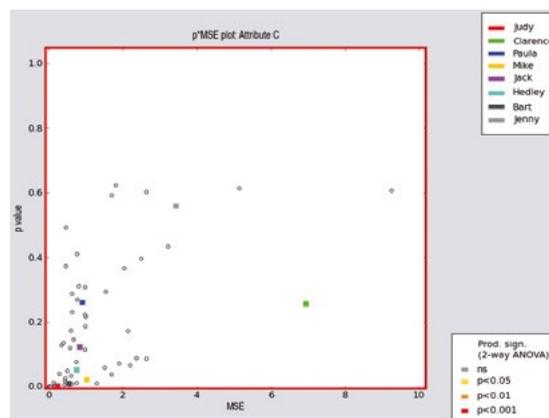


Figure 12:  $p^*$ MSE plot = joint visualisation of the F plot and MSE plot (evaluation from PanelCheck)

### 5.3.2.3 Homogeneity

The **homogeneity** of the ratings is also an essential point. If significant product/sensory assessor interaction occurs, the sensory assessors assess the products significantly differently or use the scales very differently, sensory panel agreement is accordingly not shown in the result. This can also be evaluated relatively simply using the software tool. In various profile plots in which the data of the sensory assessors is shown, interactions and scale usage ranges can be read off for each sensory assessor. The profile plot in Figure 13 provides details on the sensory assessors' results. In this example, the sensory panel members Mike and Paula are particularly noticeable because they deviate from the other sensory panel members.

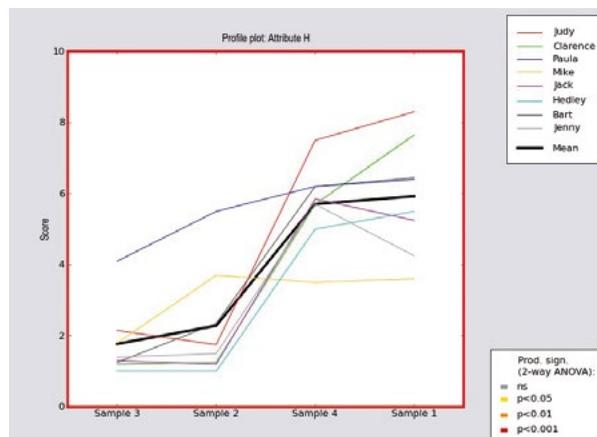


Figure 13: Profile plot (evaluation from PanelCheck)

Further software tools are listed in the Annex to this publication. If corresponding software programmes are to be purchased at your company, contacting the suppliers directly and subsequently conducting a test and training phase for the respective software within the company are recommended.

## 6. Summary and outlook

Part 3 of the 'Practice guide for sensory panel training' series of publications describes the high importance of sensory assessor and panel monitoring. Expert sensory assessors, who function in a manner comparable to chemical-physical measuring instruments, require specific calibration before being used as 'test instruments'. The requirements of 'good laboratory practice' also apply to 'good sensory analysis practice'. Based on a defined sensory performance requirement profile focussed on the area of work, sensory assessors have to be selected by means of screening and specifically trained in a project-related manner. The test results and data obtained both during the training phases and in projects form the basis for continuously monitoring the performance of expert sensory assessors in terms of discrimination ability, repeatability and homogeneity. Panel leaders and project managers in the field of sensory analysis and evaluation should regard the above descriptions as confirmation of the sensory assessor and panel monitoring that they have already implemented or as motivation and suggestions for continuous quality improvements in this process, which is elementary for the reliable use of assessors and panels in the field of sensory analysis and evaluation. Although comprehensive, well thought-out sensory assessor and panel monitoring appears complex at first glance, the conceptual effort undertaken in advance is worthwhile. As the process continues, this input pays off and takes on optimum transparency through the increasing digitalisation of all data, thus facilitating the work even further.

### Literature:

- Busch-Stockfisch, M. (Ed.): Praxishandbuch Sensorik in der Produktentwicklung und Qualitätssicherung. Hamburg: Behr's Verlag, 2002 (44th Supplement 2019)
- Carpenter, R.; Lyon, D.; Hasdell, T.: Guidelines for Sensory Analysis in Food Product Development and Quality Control. 2nd Ed. Gaithersburg: Aspen Publ., 2000
- DLG Sensory Analysis Committee: DLG Sensory Analysis Vocabulary. Frankfurt/M: DLG-Verlag, 2015
- DLG Sensory Analysis Committee: DLG Expert Knowledge Series - Sensory Analysis, free to download from [http://www.dlg.org/expertenwissen\\_sensorik.html](http://www.dlg.org/expertenwissen_sensorik.html)
- Kemp, S.; Hollowood, T.; Hort, J.: Sensory Evaluation: A practical handbook. Chichester: Wiley-Blackwell, 2009
- Lawless, H.; Heymann, H.: Sensory Evaluation of Food. 2nd Ed. New York: Springer, 2010
- Muñoz, A.; Civille, G.; Carr, T.: Sensory evaluation in quality control. New York: Van Nostrand Reinhold, 1992
- Standard DIN 10976:2016-08 – Sensory analysis – Difference from Control-Test (DfC-Test)
- Standard DIN EN ISO 5492:2008-10 – Sensory analysis – Vocabulary
- Standard DIN EN ISO 8586:2014-05 – Sensory analysis – General guidelines for the selection, training and monitoring of selected assessors and expert sensory assessors
- Standard DIN EN ISO 8589:2014-10 – Sensory analysis – General guidance for the design of test rooms
- Standard DIN EN ISO 11132:2017-10 – Sensory analysis – Guidelines for monitoring the performance of a quantitative sensory panel

- Standard DIN EN ISO 13299:2016-09 – Sensory analysis – Methodology – General guidance for establishing a sensory profile
- Standard DIN ISO 5725-2:2012-12 – Accuracy (trueness and precision) of measurement methods and results – Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method
- Standard DIN EN ISO/IEC 17025:2018-03 – General requirements for the competence of testing and calibration laboratories
- PanelCheck: <http://www.panelcheck.com/Home> [Accessed on 14.01.2020]
- Stone, H.; Sidel, J.: Sensory Evaluation Practices. San Diego: Academic Press, 1985
- Tomic, O.; Luciano, G.; Nilsen, A.; Hyldig, G.; Lorensen, K.; Næs T.: Analysing sensory panel performance in a proficiency test using the PanelCheck software. European Food Research Technology 230 (2010), p. 497-511
- Tomic, O.; Nilsen, A.; Martens, M.; Næs, T.: Visualization of sensory profiling data for performance monitoring. LWT - Food Science and Technology 40 (2007), p. 262-269
- Raithatha, C.: Panel performance measures. In: Rogers, L.: Sensory Panel Management. Duxford: Woodhead Publishing, 2018
- Raithatha, C.; Rogers, L.: Panel Quality Management. In: Kemp, S.; Hort, J.; Hollowood, T.: Descriptive Analysis in Sensory Evaluation. Chichester: John Wiley & Sons, 2018

## Selection of software with different functions for evaluating the performance of the sensory assessors

- Compusense: programme for organising, conducting and evaluating sensory analyses. Further information from <https://www.compusense.com/en/>
- FiZZ: programme for organising, conducting and evaluating sensory analyses. Further information from <https://www.biosystemes.com/en/fizz-software.php>
- PanelCheck: programme for evaluating sensory analyses. Further information from <http://www.panelcheck.com/Home/softwarefeatures>
- Red Jade: programme for organising, conducting and evaluating sensory analyses. Further information from <https://redjade.net/sensory-analysis-software/>
- SenPaq: programme for evaluating sensory analyses. Further information from <https://www.qjstatistics.co.uk/product-category/software/>
- SensomineR: programme based on the statistics software R for evaluating sensory analyses. Further information from <http://sensominer.free.fr/>
- XLSTAT: programme that is integrated into the Microsoft Excel spreadsheet programme for evaluating sensory analyses. Further information from <https://www.xlstat.com/de/>

This overview makes no claim to completeness and does not constitute a recommendation.

## Authors:

DLG Sensory Analysis Committee project team; [www.DLG.org/sensorikausschuss.html](http://www.DLG.org/sensorikausschuss.html)

Under the overall charge of:

- Prof. Dr Jörg Meier, Neubrandenburg University of Applied Sciences, Faculty of Agriculture and Food Sciences, Neubrandenburg, Germany, [jmeier@hs-nb.de](mailto:jmeier@hs-nb.de)
- Annette Bongartz, ZHAW, Institute of Food and Beverage Innovation, Food Perception Research Group, Wädenswil, Switzerland, [annette.bongartz@zhaw.ch](mailto:annette.bongartz@zhaw.ch)
- Dr Jeannette Nuessli Guth, ETH Zurich, Department of Health Sciences and Technology, Zurich, Switzerland, [jnuessli@ethz.ch](mailto:jnuessli@ethz.ch)
- Prof. Dr Rainer Jung, Geisenheim University, Institute of Oenology, Geisenheim, Germany, [rainer.jung@hs-gm.de](mailto:rainer.jung@hs-gm.de)
- Dr Eva Derndorfer, Vienna, Austria, [eva@derndorfer.at](mailto:eva@derndorfer.at)
- Bianca Schneider-Häder, DLG Competence Center Food, Frankfurt/M., Germany, [B.Schneider@DLG.org](mailto:B.Schneider@DLG.org)

## Contact:

DLG Competence Center Food, Bianca Schneider-Häder, [sensorik@DLG.org](mailto:sensorik@DLG.org)

© 2021

All information and references are provided without any guarantee or liability. Any reproduction and transmission of individual text sections, drawings or pictures – even for educational purposes – requires the prior consent of DLG e.V., Service Department Communication, Eschborner Landstrasse 122, 60489 Frankfurt am Main.



**DLG e.V.**

**Fachzentrum Lebensmittel**

Eschborner Landstraße 122 · 60489 Frankfurt am Main

Tel. +49 69 24788-311 · Fax +49 69 24788-8311

FachzentrumLM@DLG.org · [www.DLG.org](http://www.DLG.org)