

DLG Expert report 4/2012

Efficient planning, preparation,  
performance and evaluation of  
sensory tests (Part 2)



Part 1 of the DLG Expert report on the subject of “Efficient planning, preparation, performance and evaluation of sensory tests”, issue 3/2012, discussed definition of the target and various aspects to be taken into account in planning the test. Part 2 of this DLG Expert report continues with those elements of test design not yet examined and in addition focuses on performance and evaluation of the test.

## 2.6 Test protocol

The design of the test protocol is also important in connection with the planning of a sensory test. It should contain information such as the name of the tester, date and time of the test, designation of the material to be tested and the code number of the sample/s. Statements concerning the health status of the tester are also helpful so that if appropriate, it may be possible to reconstruct why a tester came to different results compared with all the others.

Name of the test person:		Date:
		Health status:
Test samples: Attribute/attribute property		
<u>Test instruction:</u> Test the samples on the tray from left to right in the manner stated by the test manager. Re-tasting is allowed. <b>In which of the two samples is the attribute stated or the attribute property most pronounced?</b> If you are not certain, state in which sample you suppose/suspect that it is most likely to be more strongly pronounced. If necessary you will have to guess.		
Test samples (Enter code numbers)	More strongly pronounced sample (Enter code number)	Correct/wrong
.../...	...	
Remarks:		
<input type="checkbox"/> reliably recognised	<input type="checkbox"/> supposed/suspected	<input type="checkbox"/> guessed
Other remarks:		

The heart of the protocol is the precise test instruction with the questions to be answered. Care should be taken here to ensure that these are described precisely and understandably for every member of the panel. Faulty or misunderstandable descriptions and explanations can completely ruin a sensory test.

Figure 4: Test protocol for duo tests

Finally, the protocol must also offer the possibility of entering or ticking the test result. Here too, care should be taken to ensure that it is unmistakably clear how this is to be done. For example, in the case of a ranking test, the possibility that the testers enter the ranking of the samples in reverse order must be ruled out. It is also helpful to add an additional field “Remarks”. Here the testers should note, for example, if they have noted any discrepancies or inconsistencies during the test, such as inhomogeneities in a sample that have led to problems with the assessment, different sample temperatures, or an insufficient quantity of the sample. In the case of a Difference Test they can also note whether they are certain of the result or whether they simply suspect or have even guessed it. This may represent important additional information. Figure 4 shows an example of a test protocol.

## 2.7 Test time and procedure

The timing of the test also deserves consideration. It is not advantageous to conduct sensory tests directly after a meal, because then the testers are not hungry and not very inclined to ingest further foods. However, the testers should not be hungry either. That is why test times in the middle of the morning or afternoon are most suitable. When testing alcoholic beverages, afternoon times should be avoided by way of precaution as otherwise it cannot be guaranteed that the blood alcohol level is back within the admissible range at the end of the working day. This could represent a problem for drivers.

A specific procedure must be observed for some food tests. For example products with an intensive taste, beverages with a stronger alcohol content or samples that leave a sustained aftertaste should generally be tested at the end of a session.

## 2.8 Number of tests per session

The number of tests per session should not be too high. A large number of samples or sets of samples to be tested promotes adaptation and thus a reduction in the sensitivity of the senses. It also leads to fatigue among the testers. Furthermore, it then becomes ever more difficult to distinguish between various samples. Their sensory images increasingly overlap in the testers' memory. These effects are all the more pronounced, the more similar the samples are.

The number of tests to be conducted per session also depends on the test method selected. For example, it is possible to carry out distinctly more sample comparisons per session in Duo Tests or "A"- "not-A"-Tests than in a Triangle test, as in the latter case it is necessary to taste backwards and forwards much more in order to reach a result.

The nature of the samples must also be taken into account when determining the number of tests per session. The more intensive the taste of the samples to be tested, the higher their alcohol content is, or the longer-lasting the aftertaste, the fewer tests per session are possible.

## 2.9 Neutralisation

Within the framework of planning tests, it is also important to think about neutralisation between the individual samples. The purpose of neutralisation is to rinse out residues of the preceding sample and reverse any adaptation of the taste sense cells. On the other hand, neutralisation between the samples promotes forgetting of how the preceding sample tasted. Especially in the case of difference tests, where very similar samples are presented in a set of samples, it is therefore sometimes better to take such adaptation into account and refrain from neutralising within the sample set. However, neutralising must always be carried out in each case prior to repeating the test to secure the result or before moving on to a further set of samples.

It should also be considered what means of neutralisation are to be used. For many foods, largely taste-neutral still mineral water or tap water with a low mineral content and white bread are suitable. On the other hand, for fat-rich foods such as margarine or chocolate, (cold) water is not really to be recommended. However, as many people prefer not to drink warm water, instead very weakly and briefly brewed tea without any added flavour can be used instead. It may not under any circumstances have an astringent affect (causing contraction of the oral mucosa) and must not be too hot. In the case of pungent foods, milk and possibly white bread always followed by water are suitable for neutralisation.

## 2.10 "Warm-up"

It is advantageous to conduct a "warm-up" before starting a sensory test. For this the testers are presented either with samples that are subsequently to be tested or a similar sample in a form which has been rendered anonymous, but without any coding. This is to enable the testers to become accustomed already to the sample/s. In the case of evaluation tests (e.g. DLG Test), the quality standard of the testers can be compared again with the aid of such a warm-up sample. These warm-up samples are collected again before the actual test starts and testers must then first neutralise their senses.

## 2.11 Contact with the testers

In the case of consumer tests it must be decided whether to allow the testers to test independently and leave them alone with the samples and test protocols, or whether an individual should be commissioned to conduct the test in contact with the testers, handing out the samples to them, formulating the questions and noting the results. The first method requires less time and personnel. However, it is possible that the results may be falsified through misunderstandings regarding the test question, confusion of the samples, or mistakes when filling out the test sheet.

The second method requires more time and personnel, but offers the opportunity of avoiding such faults. Furthermore, it supplies more information, as verbal remarks by the testers and non-verbal signals can also be registered. This is significant with regard to the test purpose, especially in the case of popularity/acceptance tests or preferences of food samples. On the other hand, when this method is used there is a risk that the testers might be unintentionally influenced. One way out of this dilemma is a “double-blind” procedure in which neither the tester nor the persons directly involved in performance of the test know the purpose of the test or the question interesting the client.

### 3. Performance of tests

Problems in the performance of sensory tests that could possibly involve falsification of the test results are connected with

- phenomena of the physiology and psychology of the senses
- the behaviour or reactions of the testers
- changes in the samples during the test
- changes in the test environment during the test.

#### 3.1 Physiology and psychology of the senses

Processes in the physiology or psychology of the senses that can influence the test results include “carry-over”. On the one hand if neutralisation is not or only insufficiently carried out between the samples, it can happen that samples become mixed in the mouth. On the other hand, the sample just tested leads to adaptation of sensory cells so that these react less sensitively in the following tests. In both cases sufficient neutralisation helps, but this promotes forgetting of sensory impressions.

However, “carry-over” can also occur in that a sample just tested influences the tester’s evaluation standard. For instance a very intensive or extremely high quality sample just tasted can make a following mediocre sample appear less intensive or poorer in quality than would be the case if the latter were tasted alone. Following a sample of very weak intensity or very poor quality, the same mediocre sample would, on the other hand, be classified as more intensive or of better quality. Such contrast effects can be countered in the case of serial monadic tests by pre-sorting of the samples through a ranking test, or by applying a neighbour-balanced test plan.

Attention has already been drawn to the risk of forgetting sensory impressions during a test, and to overlapping of the sensory impressions in the tester’s brain.

#### 3.2 Test persons

The test result can be influenced by the test persons, e.g. through interactions between the testers during the test. Action to prevent this can be taken through test cabins and their appropriate arrangement (Figure 5). Possible influences on the testers by the test manager and other context effects have already



Figure 5: Test cabins for sensory tests

Source: SAM Sensory and Marketing International

been mentioned in connection with the planning of tests, as have changes in the performance capability of the testers during the test (training effect, fatigue, dwindling motivation). Furthermore, there may be a change in the cognitive strategy in the course of a test, i.e. in the way in which the brain approaches a solution to the test tasks.

For instance, in a triangle test testers concentrate first on identifying the sample that deviates – in whatever properties. However, as soon as testers believe they have noticed a difference in a specific property, they concentrate only on this. Such modifications cannot be ascertained from the exterior and can only be established by questioning the testers.

### 3.3 Changes in the samples and the test environment

Changes in the samples during the test session can occur for instance in the sample temperature and texture. It is also possible for the aroma substances to volatilise.

In addition, the test environment (temperature, brightness, odour stress) can change during the session and influence the test results.

Such modifications can be prevented by swift testing, sessions that do not last too long and via appropriate technical measures (temperature control and covering of samples, lighting, ventilation and climate control of the test room).

The potential influences that might possibly falsify the test result during a sensory test should be considered thoroughly in advance and taken into account in the design of the test room, in the test planning and during performance of the test in order to obtain reliable results.

## 4. Test evaluation

Statistical methods play a major role in evaluating sensory tests. The method applied depends for instance on the nature of the data and the purpose of the test conducted. Furthermore, it should be noted whether an individual sample is to be characterised or two or more samples are to be compared with each other and whether the samples are interdependent or independent (Figure 6).

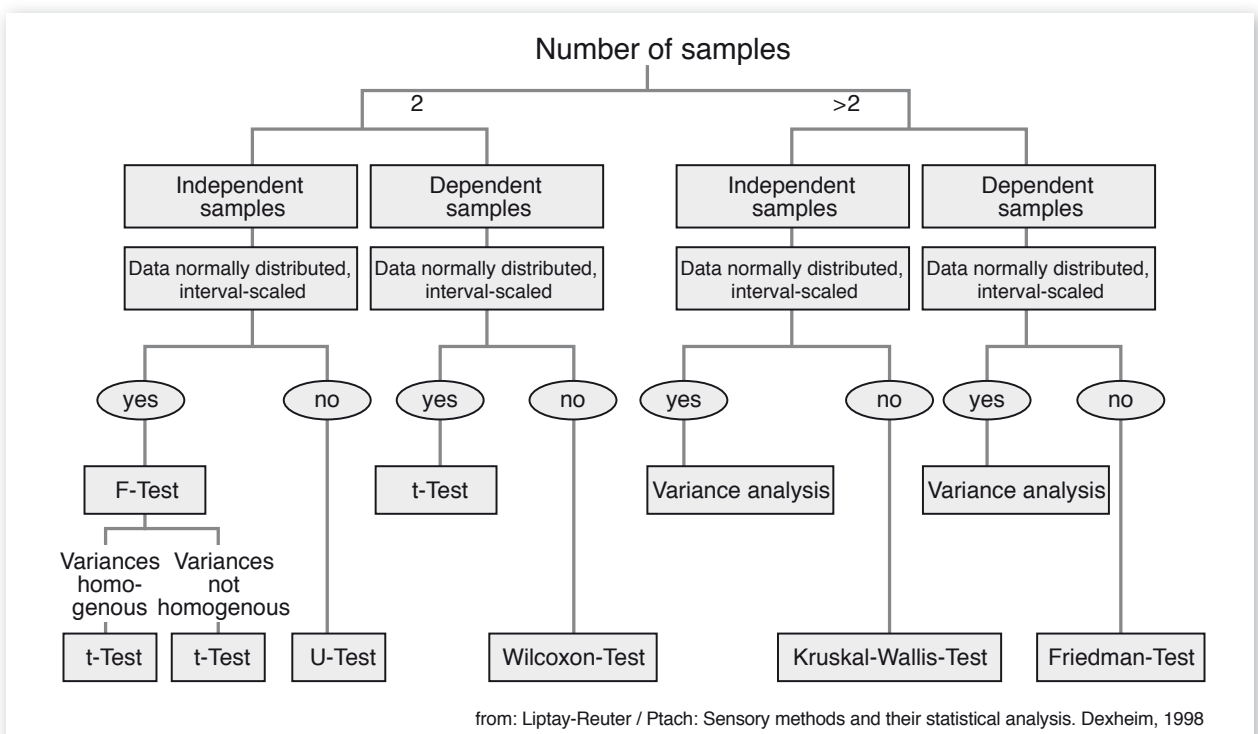


Figure 6: Analysis of intensity tests

## 4.1 Nature of the data

As regards the nature of the data, a distinction is made between nominal and category data such as are generated in difference tests (number of correct or incorrect answers), ordinal data (ranking), interval data and relational data. Only the last two data groups that result from the use of scales are sufficiently numerical in the mathematical sense, so that it is possible to calculate with them (e.g. calculating mean values, standard deviations, correlations). Both come from a scale in which the scale values display equal intervals. Interval data differ from relational data in that the scale used does not show any genuine zero point.

## 4.2 Samples, sampling

Different statistical methods are used depending on whether an individual sample of a food is to be characterised or two or more such samples are to be compared with each other. In the case of several samples, sensory testing generally involves dependent samples, i.e. the data originate from the same panel or the same tester. The same testers have, for example, examined both a sample of the standard product and a sample with the new, improved formulation. Independent samples mean that samples are tested by different panels, e.g. when a consumer test is conducted in parallel in different towns or at different times.

## 4.3 Test target

Depending on the target of the sensory test, the statistical evaluation will also differ. In the case of a Difference Test for example, the purpose is to find out whether a difference between two samples is significant. This is the case when the probability that the test results received reflect a difference purely by chance, even though this does not exist, is sufficiently low. In this way it is possible to check, for instance, whether a change undertaken in the formulation has in fact produced the desired sensory effect.

It is also possible to examine statistically whether the tested samples are significantly similar to each other, for example whether the replacement of a more expensive ingredient by a lower cost ingredient leads to approximately the same sensory result in the end product. In this case the probability of the test results leading to the conclusion that no difference exists, even though there is such a difference, must be sufficiently low. In Ranking Tests the purpose is to identify the significance of a ranking found.

In the case of Intensity Tests using a scale, such as are conducted for a profile analysis, it is of interest to find out whether there are significant differences between samples in one or more sensory attributes. In addition, it is possible to determine from the mean intensity values how great the intensity of the respective attribute is. The standard deviation is a measure for the scatter of the individual values around the mean value. Mean values and standard deviations are calculated in order to characterise one or more samples (descriptive statistics) or in order to draw conclusions regarding the character of the population (e.g. production batch) with the help of these samples (conclusive statistics). Such calculations (known as "parametric statistics") are however only admissible when the data display a normal distribution. This must therefore always be checked in advance. Even intensity data that appear normally distributed need not necessarily be normally distributed, as in the test for normal distribution it is assumed that the scale intervals are the same, while in the heads of the testers the intervals between the individual scale points may in reality be quite different. For example, some testers shy away from using very high or very low scale values. This problem also remains when unstructured line scales are used. Mean values close to the ends of the scale cannot be based on any normal distribution of the individual values, as the scale does not offer "any space" on one side for a normal distribution. Parametric evaluation methods may not be used in the case of evidently not normally distributed data. In such cases it is necessary to change to non-parametric statistical methods (e.g. Wilcoxon-Test, Friedman-Test), even if ultimately these only supply the information as to whether differences are significant or not and what sample/s are more intensive or less intensive in the respective sensory attribute. It is not possible to

make any statement about the size of the sensory differences between the samples. In the case of approximate or apparent normal distribution, it is expedient to apply the non-parametric evaluation at least for control purposes, as it is not known how strongly the conditions for the parametric statistics are breached and what faults result from this.

In as far as different samples were tested by the same panel, it is generally possible to assume that the intensity data obtained (dependent samples) display roughly the same scatter for all samples (homogeneous variance). However, if the sample intensity data originate from different panels (independent samples), this is not admissible. In this case the data are to be tested for homogeneity of the variances (F-Test). The homogeneity or non-homogeneity of the variances for independent samples is to be taken into account in the calculations to establish significant sample differences (t-Test).

Figure 6 shown an overview of the statistical methods for evaluating intensity tests.

Acceptance Tests with consumers also supply data that apparently result from an interval scale. Here, however, it is even less possible than in the intensity tests to assume that the intervals on the hedonic scale in the heads of the consumers are the same. That is why non-parametric evaluations are more reliable.

#### **4.4 Commercial sensory analysis programs**

The availability of commercial sensory analysis software with the possibility of conducting sound statistical analyses of the data obtained in sensory tests and illustrating these in graphic form makes matters much easier in practice. However, it must not be forgotten that the results of the statistical analysis can always only be as reliable as the sensory data on which they are based. Careful planning and performance of the sensory tests, which also includes thorough training of an analysis panel, thus deserve absolute priority. Otherwise even the most sophisticated statistics are not worth much.

### Reading list:

- Busch-Stockfisch, M. (Hrsg.): Praxishandbuch Sensorik in der Produktentwicklung und Qualitätssicherung. Loseblattsammlung. Behr's Verlag, Hamburg
- O'Mahony, M.: Sensory Evaluation of Food. Marcel Dekker Inc., New York und Basel 1986
- Stone, H., Sidel, J. L.: Sensory Evaluation Practices. Academic Press Inc., San Diego, New York, 2004
- Derndorfer, E.: Lebensmittelsensorik. Facultas-Verlag, Wien 2006
- Quadt, A., Schönberger, S., Schwarz, M.: Statistische Auswertungen in der Sensorik. Behr's Verlag, Hamburg, 2009
- Liptay-Reuter, I., Ptach, C.: Sensorische Methoden und ihre statistische Auswertung. Dexheim, 1998
- DIN and DIN EN ISO standards on sensory analysis

### Author:

Prof. Dr. Dietlind Hanrieder, Anhalt University of Applied Sciences, Bernburg, Dietlind.Hanrieder@hs-anhalt.de

### Contact:

Bianca Schneider-Häder, DLG Competence Center Food, Sensorik@DLG.org

In cooperation with the DLG Sensory Analysis Committee ([www.DLG.org/Sensorikausschuss.html](http://www.DLG.org/Sensorikausschuss.html)).

© 2012

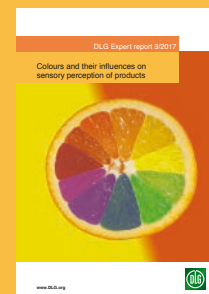
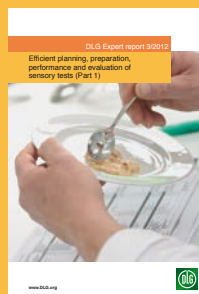
All information and references are provided without any guarantee or liability. Any reproduction and transmission of individual text sections, drawings or pictures – even for educational purposes – requires the prior consent of DLG e.V., Service Department Communication, Eschborner Landstrasse 122, 60489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

## DLG Expert report: concise information on topical issues in the food industry

Expertise, trends and strategies first-hand. The DLG produces numerous publications informing readers about current issues and developments in the fields of food technology, quality management, sensory analysis and food quality. In the "DLG Expert report" series, experts tackle current issues and provide concise information and assistance.

Individual editions of DLG Expert report are available as downloads at [www.DLG.org/publikationen.html](http://www.DLG.org/publikationen.html).

Further information on DLG Expert report: DLG e.V., Marketing, Guido Oppenhäuser, email: [G.Oppenhaeuser@DLG.org](mailto:G.Oppenhaeuser@DLG.org)



**DLG e.V.**

**Competence Center Food**

Eschborner Landstr. 122 · 60489 Frankfurt am Main · Germany

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

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