# SENSORY EVALUATION TECHNIQUES DO'S & DON'TS

*Sitaram Dixit*
*Corporate Manager – Fragrances*
*S. H. Kelkar & Company P. Ltd.*

## INTRODUCTION

Sensory evaluation done to differentiate the fragrance character between two samples is different from knowing which sample is preferred over the other. Defining the test objective is critical and is the primary criteria in any sensory evaluation as it informs exactly what needs are to be determined, the type of testing that is to be carried out, the panellist to be selected, the methodology to be followed, and the level of statistical risk factors that is to be used for interpreting the observations and arrive at the result to provide recommendation.

Sensory testing is of two types:
1. Preference testing and
2. Analytical testing.

## PREFERENCE TESTING

1. Which of the samples do you like?
2. How much do you like it?
3. What in this sample, you do not like?

If you need answers to questions like these, then we need consumer responses. Generally consumers for preference testing are pre-screened to select actual users of the products so that a correct response is obtained. Consumer testing although is the most suitable way to evaluate a product it certainly is not the best method to follow always. If we want to simply identify whether the two samples are different then consumer testing will become expensive. However in case consumer response to new improved product is required then it probably is the only correct way.

Preference testing can also be performed in-house by a panel consisting of employees who are also consumer's frequently using similar types of products. They should not be employees who happen to be present at the right time and place to collect free samples. In-house panellists may not be the exact target group but they prove useful for routine testing done to identify fragrance directions. It is advisable to use as many panellist as possible, preferably over 50 but not less than 30 in number and if possible repeat the testing a couple of times and impart credibility to the results. The standard hedonic scale of 0 to 5 points is used for acceptance testing.

### Employee preference

About 50 employees are screened and selected based on their frequent use of the product under study. The major risk in employee assessment is the bias an employee may have towards the product. The bias can be either positive or negative and this difference is difficult to measure and correct. Employee's preference rating specifically so is used only for initial screening and general direction. After the product tested passes the in-house employee acceptance test it may be necessary to subject the selected samples to actual in-use consumer acceptance testing either in a consumer central location test (CLT) or home placement test (HUT) to verify the results obtained in the in-house employee preference testing.

### Central Location Test (CLT)

CLT is comparatively more expensive than employee preference test. In a CLT subjects after adequate screening are selected based on socioeconomic demographics and product usage profile. Normally 100 consumer's panellists per location are recruited and tests carried out in at least 2-3 target locations. CLT normally takes a couple of months to plan, execute and analyze and so is more time consuming when compared to in-house panel testing. CLT's however are recommended for new products, improvements and competitive evaluations. The standard 6 point hedonic scale is normally used for this purpose.

### *Focus Group qualitative study (FGS)*

Focus group studies are generally carried out to seek insights and directions rather than precise measurements. Typically 10 – 12 screened respondents based on their specific consumer profile requirements are recruited and put together in a room. A qualified focus group moderator in an open round table forum moderates and discusses their feelings and motivations. These open discussions uncover consumer feelings that are not easily brought to the surface by quantitative methods like CLT, HUT, Mail or Telephone surveys. Focus groups can be formed rather quickly and at relatively lower costs. Although they a little difficult to design, execute and analyse it is no doubt a very flexible research technique. Focus group study being simple can also sometimes lead to sloppy invalid research and for this reason, is used only to help clarify issues and not for independent quantitative researches.

### *Home Placement Test (HUT)*

In a home placement test pre-screened normal households that actually use the product under test is selected and respondents recruited. The recruitment may be either by telephone, door-to-door selection or by targeting consumers in a super market. In an HUT the product to be tested, depending upon the product category and consumption patterns are left with the users for a set time period, say from 2 weeks to 2 months. Consumer feedback or responses are collected by an interview, personal or written questionnaire, to quantify reaction and opinions of consumers about the product tested. Questions about price, utility, packaging, etc., can also be assessed in addition to the basic question on product performance and acceptance. Home placement tests are expensive and time consuming as it takes about 3 – 6 months to plan and execute. HUT is absolutely essential when factors like, ease of preparation, formulation flexibility, sensory fatigue test responses are involved or are required to be assessed.

### *Mail panels*

Mail panels are considered as low cost type home use test. It permits testing by consumers spread or scattered over a large geographical area for example a state or country under normal usage conditions. Mailing list is generated and indexed by required consumer profile, socioeconomic demographics, etc., and product samples mailed with necessary instruction of product usage, along with the questionnaire for collecting consumer responses. Generally the consumer contact list generated is maintained as a data base for use in future testing.

### *Telephone interview*

Today with communications on the rise, use of telephone and cellular services is also a very popular method to conduct market research. However as the consumer responses received are purely verbal it is not very much suitable for obtaining responses for sensory perceptions. Moreover the consumers recruited for telephonic responses should be genuine; otherwise the responses obtained may be incorrect leading to wrong analysis and incorrect final judgements.

### ANALYTICAL TESTING

In an analytical testing procedure, untrained subject or trained judges use objective sensory methodologies to help them define characteristic properties of end product application samples. However, it does not define any acceptance or rejection measures. Analytical testing answers some typical questions as mentioned below.
1. Which sample is stronger in fragrance impact?
2. Which of the samples are different?
3. Which one of the sample has a stronger sandal note?

Analytical test methods help in defining the sensory properties but will not give any indication as to how much it is liked or disliked. Analytical tests include Differential testing tests like, triangle, duo – trio, paired, multiple comparison test, signal detection, etc.

### *DIFFERENCE TEST*

This test is used to find out whether there is a difference between two products with respect to any specific attribute or characteristic. Difference test is extremely sensitive and can often indicate minute differences between two similar products. Various versions of difference test are used in the consumer product industry. The most common ones are the Triangle test, Duo-trio test, Dual standard test, Paired difference test, A not A test, and Multiple standard test. Triangle test is the most popular one and very widely used.

## 1. Triangle test

### Example:

Due to the rising cost of edible grade roasted coconut oil a hair oil manufacturer want to change to solvent extracted grade of coconut oil for making their product. Solvent extracted coconut oil performs very similar to edible grade coconut oil. However, the hair oil base made by use of solvent extracted coconut oil has a slightly different base odour. The manufacturer wants to know whether this base odour is masked by fragrance or is still detectable in the final hair oil blend made.

### Objective:

To determine whether the hair oil made with solvent extracted coconut oil can be detected easily because of its base odour.

### Method to use:

- The company intends to replace edible grade coconut oil one of the main ingredients in the hair oil blend. A small but sensitive test is required before the company undertakes a larger market research study.
- Time is not a constraint; however the test should be reliable.
- Hair oils made with both old and new oil blends are to be tested.
- A triangle test is the most suitable for the purpose. The results can be validated by $\chi 2$ (Chi square) test.

### Protocol and Standard operating procedure:

- Forty volunteers having an odour discriminating nose considered as good odour evaluators are selected as panellist.
- Two hair oil samples made with edible grade coconut oil (reference standard / control) are labelled "P2" and "Q8" and one sample made out of solvent extracted coconut oil (experimental) is marked "R7".
- The hair oil samples are all made having identical formulation, fragrance and colour, except for the coconut oil type.
- The panellists are asked to smell the hair oil sample and select the odd one, out of the three samples given.

### Test results obtained:

|  | Correct selection | Incorrect selection | Total |
|---|---|---|---|
| **Experimental results** | 23 | 17 | 40 |
| **Expected results if no difference found (Hypothesis)** | 13 | 27 | 40 |

### Validity:

The function $\chi 2$ is given by
$$\chi 2 = \frac{\Sigma \ (O\text{-}E)^2}{E}$$
where O = observed value and E = expected value.
$$\chi 2 = \frac{(23\text{-}13)^2}{13} + \frac{(17\text{-}27)^2}{27}$$
$\chi 2 = 10^2 / 13 + (\text{-}10)^2 / 27$
$\chi 2 = 100 / 13 + 100 / 27$
$\chi 2 = (100 \ X \ 2) + 100 / 27$
$\chi 2 = (200 + 100) / 27$
$\chi 2 = 300 / 27$
$\chi 2 = 11.11$

Now, the calculated value of $\chi 2$ is 11.11. This value is compared with the table[I] value of $\chi 2$ for given degrees of freedom at a certain specified level of significance. (Generally 5% level is selected and used). If the calculated value of $\chi 2$ is more than the table values of $\chi 2$ the difference between theory and observation is considered to be significant, i.e., it could not have arisen due to fluctuations of simple sampling. If on the other hand the

calculated value of $\chi 2$ is less than the table value the difference between theory and observation is not considered as significant, i.e., it is regarded as due to fluctuations of simple sampling and hence ignored. The results though not statistically significant in the experimental observations may become statistically significant when carried out with larger number of sample observations. $\chi 2$ are always positive with its upper limit as infinity. As $\chi 2$ are derived from observations it is a statistics and not a parameter. The $\chi 2$ test is also termed non-parametric where the test involves no assumption about the form of the original distribution from which the observation come. The test conducted has one degree of freedom, as there is only one constraint governing the data namely the total number of selection must be 40.

*Conclusion:*

From the $\chi 2$ distribution table the value of $\chi 2_{0.05}$ with one degree of freedom i.e., $v = 1$ is 3.841. The calculated value of $\chi 2$ is 11.11. Therefore the calculated value $\chi 2_{0.05}$ is higher than the table value. Therefore the difference between theory and experiment is significant and it could not have arisen due to fluctuations of simple sampling. The base odour in the hair oil made out of solvent extracted coconut oil is detectable.

**As discussed earlier the alternative methods of sensory testing are as follows.**

*2.   Duo – trio test :*

This test is similar to the triangle test. However in this test the control is identified and presented first. The volunteers are then asked to choose from the remaining sample the one that is different to the control given.

The expected frequencies and so the contingency tables are as given below.

| | Correct selection | Incorrect selection | Total |
|---|---|---|---|
| **Experimental results** | Ox | Oy | N |
| **Expected results if no difference found (Hypothesis)** | N / 2 | N / 2 | N |

The function $\chi 2$ is given by $\chi 2 = \Sigma (O - E)^2 / E = (Ox - N/2)^2 / N/2 + (Oy - N/2)^2 / N/2$
where O = observed value and E = expected value.

*3.   Dual – standard test :*

In this case, four samples are used. Both one control and one experimental or different sample are identified to the subjects. Then the subjects are given two unknown samples and asked to pair them. The calculations are similar to the Duo-trio test.
Now supposing the control and the experimental samples are not identified in the beginning, but presented together for pairing with two more samples that are similar the expected frequencies are altered to the standard triangle test. For example, If $P_1$ and $P_2$ are the first pair of samples and $Q_1$ and $Q_2$ the second pair and if there are no differences between them then the following pairs are possible.

**$P_1P_2 : Q_1 Q_2$ ; $P_1Q_1 : P_2Q_2$ ; $P_1Q_2 : P_2Q_1$** of which **$P_1P_2 : Q_1 Q_2$** is the only correct one.
The contingency table in this case can be given as below

| | Correct selection | Incorrect selection | Total |
|---|---|---|---|
| **Experimental results** | Ox | Oy | N |
| **Expected results if no difference found (Hypothesis)** | N / 3 | 2N / 3 | N |

The function $\chi 2$ is given by $\chi 2 = \Sigma (O - E)^2 / E = (Ox - N/3)^2 / N/3 + (Oy - 2N/3)^2 / 2N/3$
where O = observed value and E = expected value.

*4.   Paired difference test :*

In this case, the control standard is not a part of the test. It is identical to the Duo-trio test.

### 5. "A" not "A" test :

In this test the subject is given a control and a test sample and asked to study them, in detail so that they can easily identify the samples. The subjects are then given a series of samples containing both test and control samples. The subjects are then asked to identify and segregate the samples given. The calculation is identical to the Dual standard test method discussed above.

### 6. Multiple standard test :

Sometimes the control sample cannot be represented by a single product. In such a situation, multiple standard tests  are taken recourse to. In this case a number of samples are used to cover the general product type and presented to the respondent along with the test and experimental sample. The respondent is then asked to pick out the sample that is different from the entire lot of product samples given.

The contingency table in this case is as below.

|  | Correct selection | Incorrect selection | Total |
|---|---|---|---|
| **Experimental results** | Ox | Oy | N |
| **Expected results if no difference found (Hypothesis)** | N / E | (E – 1) N / 3 | N |

Where E= number of control sample + 1 test sample.
The function $\chi2$ is given by $\chi2 = \Sigma (Ox – N / E)^2 / (N / E) + (Oy – (E – 1) N / E)^2 / (E – 1)(N / E)$

### The "R – test"

This method based on signal detection theory is infrequently used in the F&F industry. The advantage of this method is that it can be used to do multiple comparisons. The R – Index can also measure the degree of difference between two products. Firstly the respondent is given a reference control standard. The reformulated product are then individually compared to the reference control and the respondent is questioned "Whether the samples are same or different to the reference control standard". The reference standard is also given to the respondent in blind or coded form as a test sample in order to measure  whether the respondent is able to distinguish the incoming sensory signal (test sample) from background signal (reference standard).

The R – Index ranges from 50 to 100. When R = 100 it means that there is a 100% noticeable difference between test and reference control. When R = 50 it means that a non-distinguishable difference between test and control exists. The higher the R – index the greater the difference between reference standard and sample. The biggest advantage of "R – test" is that it can be used to rapidly compare a reference sample with several other test samples all at the same time using a smaller number of respondents. However the disadvantage of R – Index is that, it is not universally accepted as statistical tool for testing as there are no tables to interpret the statistical significance of the R – Index result unlike other available standard statistical methods. This is highlighted by the following example.

A customer's popular fragrance spray oil contains Geranyl nitrile prohibited by IFRA. In line with the latest IFRA guidelines the fragrance supplier wants to replace the prohibited ingredient with an approved aroma chemical. The spray marketer wants to know whether this ingredient replacement is easily detectable by consumers in the final product.

### Objective:

To determine whether the fragrance spray made by the use of IFRA compliant fragrance oil is markedly different in odour profile to the original fragrance spray made with non IFRA compliant fragrance oil.

### Method to use:

- Four different perfumers have each created three versions of the original fragrance oil using aroma ingredients permitted by IFRA.
- We have to do multiple comparisons with the basic purpose of measuring the degree of difference between various products in comparison to standard.

- There are several test samples that have to be tested at the same time with a smaller number of respondents; we use the R – Index method. (See Table V)

*Protocol and Standard operating procedure:*

- Fifteen respondents are presented with reference sample.
- The respondents are then supplied with coded samples of both reference standard control and experimental test sample.
- The panellists are asked to smell the fragrance spray samples and asked the following questions.

  1. Whether they are sure the coded sample is the same as the control.
  2. Whether they think but are not sure that the coded sample is the same as the control.
  3. Whether they think but are not sure that the coded sample is not the same as the control.
  4. Whether they are sure that the coded samples are different.

| Table V Test Results Obtained. | | | | | |
|---|---|---|---|---|---|
| Total Number of respondents = 15 | Sure Signal | Doubtful Signal ? | Doubtful Noise ? | Sure Noise | R – Index = ( S / N ) X 100 |
| Sample (Sensory Signal) | 9 | 3 | 3 | | Level I ( 9 / 10 ) X 100 = 90 |
| Standard (Background Noise) | | 2 | 3 | 10 | ( 12 / 13 ) X 100 = 92 Level II |

*Conclusions:*

The R – Index 90 & 92 are closer to 100. Thus it can be concluded that the samples are different. However this method does not answer as to how different it is from the standard, although we can say that it is more different than a test sample who's R – Index is say 82. Thus this method has to be calibrated using products with known differences so that the R – Index scale has a practical meaning. Similar comparisons can be done with other fragrance test samples and conclusions drawn using R – index. The tests have to be repeated for confirming precision and accuracy of the sensory measurements.

## USE OF SENSORY EVALUATION TECHNIQUES

Although basic methods of sensory evaluation presently available and widely used ensure validity of results, an experienced professional evaluator can create new methodologies that are unique and customised for specific product application categories. Some typical application areas where sensory testing is used are as detailed below.

*New Product Development*

We are aware that nowadays, most new products developed are only imitations or variations of well established products. Sensory tests in such cases assure that the newly developed product is either a close match of the original standard product or is uniquely different in some respects with the original. Descriptive analysis is normally preferred and if preference tests are also carried out it will also determine acceptability requirements. It is interesting to note that a sandal fragrance and a rose fragrance may smell different to each other, but may exhibit similar acceptance scores.

*Product Matching*

The basic purpose of this test is to verify whether the test sample is matching the standard or the identified benchmark. Analytical tests like triangle or duo – trio are useful in such cases.

*Product Improvement*

Any improvement in the experimental product affected can be measured by preference tests. Similarly, whether an improved experimental product is liked over the original standard can also be confirmed.

*Process Change*

Sensory evaluation tests can also be used as a verifying procedure to either maintain or improve a product during the manufacturing process.

1. Analytical testing can be done to determine the difference between an experimental sample and control product.
2. Preference test can be carried out in case sample and control is different and also confirm whether the experimental sample is preferred or is only at par with the control standard.

*Cost Reduction*

Analytical tests can be done to confirm cost reduction programs where a lower priced ingredient is used for manufacture or a different process is followed during manufacturing operations.

*Checking New Supply Source of Ingredients*

Analytical tests are suitable to test products manufactured by use of ingredients procured from different supplier than the regular on. Comparison tests between the products manufactured by use of the new ingredient with that of the old supply can also be carried out.

*Quality Control*

Analytical tests can be very useful during production, distribution, sales and marketing, so that the products manufactured and marketed are of similar standards.

1. Difference tests are used to determine whether the samples are any different from the standard or are at best only at par when compared to standard sample.
2. Descriptive tests are used to confirm how much the sample is different from standard.

Both the above test results can be used to suggest remedial action and guide to effect necessary and relevant changes during processing.

*Storage Stability*

Testing the stability of a consumer product helps one, understand the ability of the formulation, to maintain its physical appearance, aesthetic appeal, its odour profile, and it's functional and chemical characteristic. Stability tests are conducted under controlled conditions designed specifically to give an early indication and forewarn the formulator about the problems that may occur during transportation, warehousing, retailing, storage all through the life of the product. The stability data, recorded during the stability testing procedures guides the formulator develop and design a superior product that will continue to be as aesthetically appealing to the customer and performing its varied functions as envisioned at the beginning stage of product development.

Samples stored at standard conditions provide valuable data and information to determine the stability and compatibility of the product at normal market conditions. The important part of stability testing is examination of samples stored to make observation and thereby evaluate the product. The samples put on stability testing and storage is examined at set time intervals. The duration of the testing time entirely depends on the condition and on the latest findings of the tests. Although the time is set at the beginning of the stability testing, it is modified according to the test progress. Typically, the duration set for testing is 1, 1½, 3, 4½, 6, 9, 12 or 24 months at the maximum.

Analytical tests used during stability studies include:

1. Difference tests in order to find whether the storage sample is different from control. If it is not different then the sample is said to pass the stability test.
2. Descriptive test is carried out either alone or in conjunction with difference tests, to quantify the change in characteristic that may have occurred during storage.
3. Descriptive analysis is very frequently employed when it is difficult or unavailable, to maintain a control under stability study.
4. Difference test may also be employed to determine relative acceptance of storage product samples.

## SOME OTHER IMPORTANT TEST CONSIDERATIONS

For all descriptive and discriminating difference tests we need a reference control or standard sample. Selecting a suitable reference standard is by itself a very important parameter for correctly carrying out the analysis. Some important points to be observed while selecting the reference standards include the following.

1. Normally test samples are laboratory prepared. In line with the test samples the reference sample should also be ideally laboratory prepared. Choosing a sample made in a production plant or using a marketed product as a reference standard as is normally done is incorrect and will lead to inconsistent and erroneous test results.
2. It is important that the product on scale up from a laboratory scale to actual plant production is similar. In case this not practically possible then both standard reference control and experimental test sample should be prepared in the production plant for carrying out descriptive and discriminating difference tests.
3. All analytical test readings should be recorded only after evaluation is carried out in the proper medium and in relevant final intended application product categories and difference in performance and sensory evaluate result vary remarkably with application.
4. All tests carried out should be confirmed for consistency of results. This is very much important especially with production or bulk samples as variation occurring in use of raw materials can have a great impact on test results.
5. Maturity or ageing of the samples before starting the test is also to be taken note of. It is important that the age of the experimental test sample and that of the reference control are same.
6. The respondents, experts, semi-experts, regular consumers, and judges selected should be screened for their ability to smell and give their opinions.
7. Respondents and evaluators should be free from any mental or sensory disturbances and should not be biased in giving their opinions.
8. All sensory testing procedures should be conducted under controlled conditions following good laboratory practices.

**Additional Reading:**

1. Consumer Product Evaluation, Principles and Practice, Sitaram Dixit, STDR Annual 2003 also Indian Perfumer (EOAI Journal), Vol. 49. No. 4, October – December 2005.
2. Is stability testing of consumer products really necessary? Sitaram Dixit, Chemical Weekly, August 7, 2001.
3. Ingredient Substitutions – Testing a match, Sitaram Dixit, Chemical Weekly, May 1, 2007.

---

[I] *see, the standard statistical tables of $\chi 2$*