9. Methodological aspects of acceptance and preference tests

Adriana Gámbaro
Sección Evaluación Sensorial Departamento de Ciencia y Tecnología de Alimentos.
Facultad de Química. Universidad de la República, Uruguay

Abstract. Affective studies are one of the most important research fields within the Sensory Evaluation area and their main objective is to determine the consumers’ acceptance of certain food products. Acceptance and preference tests are widely used at the development and launching stages of new products and during market studies. Usually, studies involving consumers have used 9-point hedonic scales in order to determine how much consumers like certain food products. However, this methodology has some drawbacks, which have led some researchers to question its validity. The same happens with traditional paired comparison tests. In this Chapter, we will gather the questions raised by various authors regarding the validity of both tests and the suggestions made to improve them.

Introduction

Sensory Evaluation as a scientific methodology is integrated by several research areas, which use human beings as tools for analysis. Even though there are various classifications in sensory analysis as a field, most authors agree
to differentiate two main areas with different methodologies and which are not supposed to overlap:

- **Analytical Tests:** These tests use the human senses as analytical tools to measure products’ sensory properties under controlled conditions. Human beings used as analytical tools for these tests are called sensory judges and are generally individuals who have higher sensory skills than the rest of the population and who have been trained to further develop these innate skills.

- **Affective Tests:** These tests use consumers who have not been trained to participate in research studies and who normally consume or use food products. The main purpose of affective studies is to assess the response to a product, a product idea or feature from real or potential consumers. These type of studies are essential for the industry during the product development stage to be able to determine the potential market for a certain product, so as to be able to optimize processes, assess new ingredients and technologies as well as to decide whether or not to keep a product in the market.

As mentioned above, affective tests use consumers as tools for research studies; however, how can ‘consumers’ be defined? We could define consumers as individuals who have not received sensory evaluation training and who represent the population in question in terms of age, gender, social or economic level, etc. We could even state that consumers are ‘ignorant wise individuals’, since they know what they want, but mostly cannot express what they want. Every time we carry out research studies with consumers we can notice that they do not know (or do not know how to express) the reasons for their purchasing choices to the researcher and therefore generally feel intimidated.

Several authors have mentioned the need to integrate the consumers’ needs to the product development process as an alternative for companies to offer more successful proposals and improve the efficiency of the product development process and therefore, its profitability [1, 2].

Affective tests are generally used as methodologies for acceptance tests using hedonic scales and for preference tests. Consumers are recruited according to certain gender, age, social and economic level and consumption frequency variables and they were asked how much they liked a certain product and asked to use hedonic scales or to choose a certain sample through the use of preference tests.

Nowadays we have realized that consumers’ behavior cannot be easily explained merely by selecting a few variables, so we have tried to get to know consumers’ behavior better by analyzing also, for instance, their
academic level, marital status, number of people in their household, nutritional knowledge, physical activity, lifestyle, mood, interest in health issues, neophilia or neophobia, etc.

Similarly, we have noticed that in addition to the information that classical acceptance and preference tests provide, we are in a position to obtain more information about consumers which can help explain how they perceive a certain product. That is why different scales have been developed, such as opinion scales, action scales and purchasing intention scales, JAR (Just-about-right) scales, CATA (Check All That Apply) scales and different methodologies, such as Sorting, Projective Mapping, Conjoint Analysis, Free Listing, Word Association, Laddering, etc.

In the present chapter we will look back at the initial methodologies used for consumer studies: acceptance tests with hedonic scales and preference tests, which are still widely used for new products launches or market studies.

**Paired preference tests**

Classic Paired Preference Tests are simple to carry out and friendly to consumers [3, 4, 5]. Consumers are presented with two samples (A and B) and requested to point out the sample of their preference. Samples should be presented according to a balanced design, i.e., half of the consumers try first sample A and then sample B and the second half of consumers tries first sample B and then, sample A. They can indicate their general preference or their preference for a certain attribute, e.g., color preference.

This is a forced response; thus, the consumer has only two choices, sample A or sample B. The test could be extended by asking consumers to indicate the reason for their preference through the question ‘Why did you prefer Sample A?’

These tests’ results are analyzed using tables made through binomial distribution. The corresponding table shows the necessary number of statements to indicate, with a maximum significance level of 5% (p < 0.05), that there is a difference among the samples. The results are obtained by comparing the number of favorable statements for one given sample (experimental value) with the value obtained from the Table (Table Value).

Let us imagine a very common situation where this test is used: a company requests us to compare, through a paired preference test, their product against the competitor’s. After the test is carried out, the result shows that half of the consumers chose the company sample and the other half chose the competitor’s sample. In that case, we inform the company that ‘There is no significant difference (p>0.05) in the proportion of consumers who prefer one sample over the other’. This is interpreted by the company as
‘consumers like my product as much as the competitor’s, so both products can equally compete in the market’. Is that interpretation really accurate?

Let us now compare this case with another hypothetical situation: on a hot day I come into an ice cream shop to find only two ice cream flavors: strawberry and chocolate and I only have money to buy a one-flavor ice cream. If I choose the strawberry ice cream, what does that mean? Let us analyze some possibilities:

a) That I really like strawberry ice cream and that is the flavor I prefer.
b) That I equally like strawberry or chocolate ice cream, but given the choice, this time I preferred strawberry ice cream, even though I may have chosen chocolate ice cream on another occasion.
c) That I do not like either strawberry or chocolate ice cream, but I am hot and feel like having an ice cream, I have no other choices, so I arbitrarily choose one flavor.
d) I am allergic to chocolate and I can only choose strawberry ice cream, even though I do not really like it.

These are some of the possible situations in the consumers’ minds when given the choice between two samples.

Let us return to the situation pertaining the sample sent by a company and its comparison with the competitor’s whose result was that there was no significant difference between both samples. How can this result be interpreted?

a) Half of the consumers prefers product ‘A’ and the second half prefers product ‘B’. It would have been relevant to have obtained more data regarding the consumers who took the test, in order to spot different variables that could explain the preferences, such as age, social or economic level, or other data which could provide guidance to the company about the market segment to target the product to.
b) Consumers did not show clear preference for either product and randomly choose one or the other. However, this does not suggest that when confronted with a real purchasing decision, consumers will choose the company product to consume.
c) Consumers did not like any of the two products, but, given the need to pick one, choose the ‘least bad’.

Several authors have questioned the forced response required by the classical test when consumers prefer one sample over another. They have proposed to include other options in the preference test, allowing consumers
to express the same preference for two products or to express ‘No Preference’ for any of them.

**Examples of preference tests which offer other options**

Example 1:
451 □  364 □  Both □

Example 2:
451 □  364 □  None □

Example 3:
451 □  364 □  Both □  None □

Even though this modification to the classical preference test enables access to more information a binomial statistical analysis cannot be used to analyze the data obtained.

Several researchers have proposed the following possibilities to continue analyzing the data using binomial statistics, even if the ‘No Preference’ option is included [3, 4, 6, 7, 8, 9, 10]:

- To ignore the ‘No Preference’ responses.
- To divide them proportionally according to the corresponding frequencies.
- To divide both preference choices equitably.
- To analyze them at random, assigning the “No Preference” responses to one of the two preference options, by simply tossing a coin.

Any of these options would not alter the test conclusions as long as the number of consumers who choose the ‘No Preference’ option is small, which is generally the case [6, 7].

Some authors, such as Marchisano et al. (2003) and Alfaro-Rodriguez et al. (2007) have proposed to analyze the data using a Chi-square methodology and comparing the expected frequency pattern under the No Preference assumption (null hypothesis) against the observed frequency pattern with the test stimuli [8,11]. Therefore, the question now would be: what is the expected frequency pattern?

Traditionally, a null hypothesis would set equal frequencies in the three response options (33% of consumers prefer sample A, 33% prefer sample B and 33% express ‘No Preference’). Equal frequencies would also indicate that consumers have randomly chosen one of the options regardless of the products.

This situation led several researchers to question this expected frequency pattern and to ask: what would happen if consumers were presented with two
equal samples (placebo condition) and were asked to choose one and the ‘No Preference’ option was included? In theory, the expected frequency pattern in this case should be a total ‘No Preference’ (0% prefers A, 0% prefers B, 100% does not prefer any sample).

Ennis & Collins (1980) carried out a test with 2 identical cigarettes assessed by consumers in their homes [12]. The assessment included various attributes and preference tests and the option ‘No Preference’ was offered. 40% of consumers stated they preferred one of the products, 40% stated that they preferred the other product and only 20% expressed ‘No Preference’. Ennis has reported to have obtained the same frequency using other type of products [13]. However, other authors have reported different frequencies according to the type of product, response options, consumers or experimental conditions [8, 11, 14, 15].

This leads us to conclude that the same preference test leads consumers to report preference for one of the products when in fact they do not prefer one product in particular, and the expressed preferences have to do with other factors unrelated to the products’ sensory features.

Presently, it is recommended to determine the expected frequency for each specific condition. Some researchers obtain the expected and observed frequencies from the same consumer group [11], while others use different consumer groups and assume they are equivalent [8, 16].

**Example of a test carried out using a placebo condition**

Consumers are presented with two pairs of samples (a placebo pair of identical samples and a problem pair) in the following balanced presentation order:

```
  AA  AA  BB  BB  AB  BA  AB  BA
 AA  BA  AB  BA  AB  BA  AA  AA
```

According to Alfaro-Rodriguez et al. (2007) the placebo pair can be the same, so it would be sufficient to use a single sample for the placebo pair, and counterbalancing the presentation order of both sample pairs, in the following manner [11]:

```
  AA  AA  AB  BA
 AA  BA  AA  AA
```

For data analysis, two possibilities are proposed:
a) To use preference for the placebo samples as the expected frequency for a Chi-square analysis
b) To use the placebo pair to select consumers and analyze just the data from consumers who stated ‘No Preference’ for the placebo pair (not biased). This last option has the disadvantage of considerably reducing the sample size.

When analyzing the way to increase the number of ‘No Preference’ responses, Marchisano et al. (2003), Chapman et al. (2006) and Kim (2007) experimented with various options, and they discovered that the more preference options, including the preference degrees, the more ‘No Preference’ responses they obtained for the placebo condition [8, 15, 17]. However, Villegas-Ruiz et al. (2008a) could not confirm this hypothesis when they studied the preference for liquid strawberry yoghurt under three conditions [18]:

**Condition 1** (with three response options): preference for A, no preference and preference for B.

**Condition 2** (with five response options): preference for A, very slight preference for A, no preference, very slight preference for B and preference for B.

**Condition 3** (with seven response options): preference for A, slight preference for A, very slight preference for A, no preference, very slight preference for B, slight preference for B and preference for B.

**Explanation of the consumers’ false preferences**

The work reported to date indicates that even though the compared stimuli are practically identical, most consumers report preferences for one or for another product [6, 7, 8, 11, 14, 15]. These ‘false’ preferences indicate: a) a very significant effect of the experimental conditions or b) that consumers really perceived slight differences between the products due to the product variations or c) that consumers perceive slight differences due to variations in the signals that their senses sent to their brains during the tasting process, this being an argument used in the Thurstonian signal detection theory [19, 20].

Despite this, the Thursonian effect is not enough to completely explain the high frequency of the ‘false preferences’, and there could be other explanations to this false preference frequency [8, 15]:

- Consumers might have strong preconceptions about the fact that the samples should be different.
- Because the response seems obvious, consumers try to make the test more complex and show how sensitive they are.
- Consumers think that the researcher would not feel comfortable if the ‘No Preference’ option is selected.
Consumers do not want to look weak by not preferring either product.

Regardless of our insufficient knowledge of the causes for the consumers’ reporting false preferences, the use of the placebo samples with identical stimuli seems a good strategy to use during a preference test.

Another way of estimating false preferences is by repeating the paired comparison test; if consumers change their responses, then the response should be considered as not stating preference. For example Wilke et al. (2006) applied the paired preference test to the same group of consumers four times at intervals of seven minutes using different pairs of food products [21]. They observed that 49.8% of consumers changed their preferences at least once when choosing cereals and raisins and 71% when assessing cola sodas. This same lack of consistency has been reported by other authors [11, 14, 15].

Presently, some authors [13, 22] are proposing a possible cause for the low number of ‘No Preference’ responses for the placebo condition and state that the ‘No Preference’ concept is ambiguous (for instance, a consumer might choose ‘A’ during the test but ‘B’ on another occasion) and that there are different types of preferences. During the classical paired comparison test consumers are asked to indicate what product they like best, which these authors call ‘liking’ preference. Another less ambiguous possibility that also increases the number of ‘No Preference’ responses for the placebo condition, would be the performance of a buying preference test, where consumers are asked to indicate which product they prefer to buy, since consumers do not always buy what they like best, due either to health reasons or because they shop for their families and not just for themselves.

Some ‘buying’ preference options are

Example 1:
- I would only buy the product on the left and would never buy the product on the right.
- Depending on price, availability, and my mood, etc., I might buy either product.
- I would only buy the product on the right and would never buy the product on the left.

Example 2:
- Other things being equal, I would buy the product on the left, but would not say I would never buy the product on the right
- Depending on price, availability, my mood, etc., I might buy either product
- Other things being equal, I would buy the product on the right, but would not say I would never buy the product on the left

Example 3:
- Other things being equal, I would tend to buy the product on the left more than the product on the right
- Depending on price, availability, my mood, etc., I might buy either product
- Other things being equal, I would tend to buy the product on the right more than the product on the left

Wichchukit & O’Mahony (2010) also suggest the performance of a ‘take away’ preference test, where the consumer is asked to pick one of the previously tried products and take it home [22]. The most important conclusion regarding the comparison of this test with the other preference tests is that the ‘take away’ preference test seemed not to be consistent regarding what consumers had stated they had liked or would buy and what they actually took home with them.

Wichchukit & O’Mahony (2011) also proposed another type of test called ‘choosing’ preference test, where consumers indicate which product they would be likely to choose when offered two or choose both or none if they can [23]. These authors concluded that ‘liking’, ‘buying’, ‘choosing’ or ‘taking away' preferences are all measures of different aspects of the consumers’ behavior and that there is no direct measure of the operational preference.

Acceptance tests

One method for studying the consumers’ acceptance of a certain product is to determine how much they say they like it when they are consuming it. In 1940 and with this objective, the hedonic scale was developed at Quartermaster Food and Container Institute of the US Armed Forces in order to be able to predict the acceptance of certain canteen food products by soldiers [24, 25, 26].

The scale consists of nine labeled categories which range from ‘Extremely Dislike’ to ‘Extremely Like’, and have a neutral category in the middle:

9 – Extremely Like
8 – Like Very Much
7 – Moderately Like
6 – Slightly Like
5 – Neither Like nor Dislike
4 – Slightly Dislike
3 – Moderately Dislike
2 – Dislike Very Much
1 – Extremely Dislike
The use of the hedonic scale has extended worldwide and has become a standard tool to determine consumers’ acceptance of food products [27]. However, this methodology poses various problems that have led some authors to question its validity.

**Problem 1:** *Due to the fact that the original scale was devised in English, there has been some difficulty regarding the translation of the scale categories.*

The original hedonic scale is integrated only by categories that indicate how much consumers like the product. These categories were selected taking for granted that they would be equally understood by the population, and ambiguous terms were not considered [28]. However, every time the hedonic scale is translated into other languages, categories can be misunderstood or not understood as having the same intensity that they were designed to convey [29].

Even the translation of the same original hedonic scale is not the same for countries that speak the same language. For example, in Spanish, the following translations can be found:

<table>
<thead>
<tr>
<th>Argentina (literal translation)</th>
<th>Argentina (free translation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Me gusta extremadamente</td>
<td>Me gusta muchísimo</td>
</tr>
<tr>
<td>Me gusta mucho</td>
<td>Me gusta mucho</td>
</tr>
<tr>
<td>Me gusta moderadamente</td>
<td>Me gusta bastante</td>
</tr>
<tr>
<td>Me gusta apenas</td>
<td>Me gusta un poco</td>
</tr>
<tr>
<td>Ni me gusta ni me disgusta</td>
<td>Ni me gusta ni me disgusta</td>
</tr>
<tr>
<td>Me disgusta apenas</td>
<td>Me disgusta un poco</td>
</tr>
<tr>
<td>Me disgusta moderadamente</td>
<td>Me disgusta bastante</td>
</tr>
<tr>
<td>Me disgusta mucho</td>
<td>Me disgusta mucho</td>
</tr>
<tr>
<td>Me disgusta extremadamente</td>
<td>Me disgusta muchísimo</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Spain</th>
<th>Mexico</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extremadamente agradable</td>
<td>Me gusta muchísimo</td>
</tr>
<tr>
<td>Muy agradable</td>
<td>Me gusta mucho</td>
</tr>
<tr>
<td>Agradable</td>
<td>Me gusta moderadamente</td>
</tr>
<tr>
<td>Ligeramente agradable</td>
<td>Me gusta un poco</td>
</tr>
<tr>
<td>Ni agradable ni desgradable</td>
<td>Ni me gusta ni me disgusta</td>
</tr>
<tr>
<td>(ni fu ni fa)</td>
<td>Me disgusta un poco</td>
</tr>
<tr>
<td>Ligeramente desgradable</td>
<td>Me disgusta moderadamente</td>
</tr>
<tr>
<td>Desagradable</td>
<td>Me disgusta mucho</td>
</tr>
<tr>
<td>Muy desagradable</td>
<td>Me disgusta muchísimo</td>
</tr>
<tr>
<td>Extremadamente desagradable</td>
<td></td>
</tr>
</tbody>
</table>

**Problem 2:** *The distance between categories is not homogeneous*
Some authors [3, 30] have reported that the acceptance categories are not equally spaced for consumers, which means that results cannot be interpreted directly. Psychologically, the distance from 8 (‘Like Very Much’) to 9 (‘Extremely Like’) is greater than the distance from 6 (‘Slightly Like’) to 7 (‘Moderately Like’). A possible solution to this problem would be to try to quantify the consumers’ acceptance of a product and to avoid using verbal categories [29].

**Problem 3: Is it better to have a verbal or a numeric scale?**

The original 9-point scale consisted of a series of verbal categories that represented various degrees of acceptance of a product that ranged from ‘Extremely Dislike’ to ‘Extremely Like’. Later, and in order to be able to statistically analyze the data, verbal categories became numeric values: ‘Extremely Like’ = ‘9’ and ‘Extremely Dislike’ = ‘1’. A verbal scale requires that consumers categorize food products according to whether they like them or not; the numeric scale requires consumers to numerically differentiate between food products in terms of how much they like each product. Food products that receive the same verbal category with the first scale could be evaluated with different numeric values in the second scale. Nicolas et al. (2010) compared the original non-structured hedonic scale version [24, 25] using only verbal categories that they called the ‘words only’ scale with a non-structured scale version using only numbers that they called ‘numbers only’ scale [31]. Most consumers responded differently to both scales. Authors suggest that consumers use different cognitive strategies for verbal or for number categories. They introduced the hypothesis that the cognitive strategy used with numbers is relative, whereas the one used with words is absolute.

**Problem 4: There are different versions of the hedonic scale.**

In order to try to avoid the problem mentioned above, several authors have used different versions of the hedonic scale. Yeu et al. (2008) use the numeric 9-point scale only labeled at both ends as ‘Extremely Like’ and ‘Extremely Dislike’ [32]. Sabbe et al. (2009) apparently use the same scale with an additional label in the middle: ‘Neither Like nor Dislike’ [33]. Other authors used a 9-point hedonic scale that consisted of nine empty category boxes (not numbers, not labels) labeled at the appropriate ends with ‘Extremely Dislike’ and ‘Extremely Like’ [34].

**Problem 5: Consumers tend not to use the ends of the scale.**

Another one of the problems of the traditional hedonic scale arising from the problems mentioned above is that consumers tend not to use the scale ends [35]. In order to overcome this problem Schutz & Cardello (2001) and Cardello & Schutz (2004) developed a scale called ‘Labeled Affective Magnitude’ [36, 37]. These
authors added the phrases ‘Greatest Imaginable Liking’ and ‘Greatest Imaginable Disliking’ to the ends of the scale, in order to better discriminate the scales. However, this type of scale has been applied mainly to highly accepted products [38], and research is still needed for averagely accepted or lowly accepted products.

Problem 6: The scale is unable to predict preference.

In 1952 Peryam & Girardot reported that the hedonic scale was accurate to detect differences in the preference for food products, assuming that a consumer would give a higher hedonic score to a product when this is preferred [24]. We could also assume, that equal hedonic punctuation suggests that there is no preference for either product. However, in 1957, Simone & Pangborn measured consumer preferences for canned cling peaches using various paired preference and hedonic tests and they reported that when the peaches were presented in pairs and consumers responded using the 9-point structured hedonic scale, …’participants were frustrated when they ‘liked’ the two samples equally, but had a preference for one, and could not indicate this on the score sheet’ [39].

Villegas-Ruiz et al. (2008b) researched whether, despite the assessors assigned liquid strawberry yoghurt the same score using the 9-point structured hedonic scale, they still preferred one product over the other one [40]. They observed that 29% of consumers assigned identical score to two or more types of yoghurt and all of them reported that, even though they had provided identical hedonic responses, they preferred one product over another one.

Problem 7: The scale is unable to predict the consumers’ purchasing intention.

Hedonic scales have been widely used due to the assumption that they enable us to predict the consumers’ consumption and purchasing decisions [41]. However, the hedonic scale scores are difficult to translate into purchasing or consumption intentions. Several research studies have shown that acceptance scores may not reflect the consumers’ behavior at the purchasing moment [42]. Rosas-Nexiticapa et al. (2005) requested 101 consumers to assess 3 types of strawberry yoghurt using the 9-point hedonic scale and using the 5-point purchasing intention scale, presented through a blind test and in their original packaging [35]. Later, consumers were monitored during a year and their purchasing behavior was followed up. Consumers were visited every two weeks with the excuse of offering them nutritional advice and the monitors collected the yoghurt packages of the yoghurt they had consumed. They found out that the scales had not been good predictors of the purchasing behavior, especially for that yoghurt that had been blindly assessed.
Giménez et al. (2007) and Ares et al. (2008) observed a relationship between the acceptance and the purchasing intention, depending on the product in question [43, 44]. Some products showed high acceptance scores (e.g. ‘dulce de leche’, chocolate), whereas others rarely obtain scores higher than 6 even if they have an excellent sensorial quality (e.g. vegetables, some fruits, liquid milk). This could be explained by considering the consumers’ motivations for consuming different types of food products. Highly accepted products are associated with products consumed due to their sensory features and due to the pleasure they provide. Instead, other products are consumed due to their nutritional value, such as mineral water or some vegetables.

A common way of solving this problem is to present consumers with two scales at the same time: a hedonic and a purchasing intention scale where the consumers indicate their intention to buy a certain product using categories that range from ‘I Would Definitely Buy It’ to ‘I Would NOT Definitely Buy It’. The main drawback of this methodology is that it does not consider the product price, only its sensory features. Therefore, consumers may show a strong intention to buy a certain product but not buy it due to money restrictions [42]. In order to solve this problem purchasing intention scales started to be used in the last few years, where the product price is specified and the consumer is asked to indicate their purchasing intention considering the different previously determined prices. Another alternative is to ask consumers directly how much they would be willing to pay for a certain product [45].

**Problem 8: Influence of social environments in acceptance tests results.**

Central location tests (CLTs) are the most common types of consumer tests [3, 51]. CLTs can be performed at relatively low costs, in short periods of time and under experimental conditions that can be easily controlled, compared to use tests carried out at households (Home Use Tests, HUTs). HUTs consist of the repeated use of a certain product under common daily use circumstances, usually for a period of several days. However, CLTs have been criticized by different authors [52] due to their artificial preparation, the amount of food being consumed, the consumption time and that the available time the consumer has to perform CLTs is short and could affect the validity of the results. The amount of food consumed during CLTs is predetermined and usually small, whereas for HUTs the consumption is free and the amount of food is usually bigger. Similarly, HUTs enable individuals to choose when to eat, so the time of the day for eating is usually

---

1 ‘Dulce de Leche’ is a sweet dairy product similar to caramel or sweetened condensed milk, very popular in some South American countries, such as Argentina or Uruguay.
the right one, which is not always the case for CLTs [53]. These differences can affect the results of hedonic tests. Even though CLTs are more cost effective, HUTs can more accurately predict consumption patterns in the long term. It has been suggested that CLTs can work better for snack consumption than for food consumed as part of a meal [54].

Sosa et al. (2008) assessed the effect of the social environment of HUTs against CLTs for low income populations, which usually have lower educational levels [55]. Evidence showed that HUTs had higher acceptance scores than CLTs for both products assessed (soups and chocolate drinks). This trend was also described by Boutrolle et al. (2007) for fermented milk, salted crackers and sparkling water [53]. Sosa et al. (2008) and Boutrolle et al. (2007) concluded that because CLTs conditions were more controlled and standardized, consumers could be more analytical and critical regarding the samples because they were able to focus more on the assessment itself, which suggests that CLTs could be more appropriate for discriminating the acceptance levels of the samples [53, 55]. This was also confirmed by Sveinsdóttir et al. (2010), who compared consumers’ acceptance of two cod products using CLTs and HUTs [56].

**Future trends**

Future research regarding the methodology for acceptance and preference tests will be based on the development and evaluation of alternative techniques to hedonic scales. An example of this, the Labeled Affective Magnitude Scale, has not been fully optimized due to the lack of information regarding its interpretation, use and reliability. This type of scale has been mostly applied to highly accepted products [38], though it is still necessary to research its applicability to averagely accepted products as well as to develop variations to be used in less accepted products.

Future studies should also deal with the comparison between the results obtained from the different hedonic scales and techniques to measure the consumers’ willingness to pay. One of the methods used in the last few years has been the use of purchasing intention scales where the price of the product is specified and the consumers are asked to indicate their willingness to pay for a certain product. Another alternative for determining purchasing intention or willingness to pay has been the use of auctions [42]. This technique makes consumers participate in an auction that enables them to reveal their preferences and indicate how much they think the products are worth. Even though these methodologies enable researchers to gather relevant information regarding the consumers’ willingness to buy an actual product, they have not yet been thoroughly used, so their applicability needs to be studied further and their results compared with the traditional hedonic scales.
Even when numerous techniques have been developed to identify the attributes expressed by consumers regarding their acceptance for food products, most of these methodologies are based in correlations provided by trained judges. Lately, there has been an increasing worldwide interest for the development of methodologies that provide information regarding the food products’ sensory features, using only data provided by consumers. One of the most common ways of obtaining sensory product information from consumers is to use Just-about-Right or JAR [46] scales. These scales enable us to gather information regarding the intensity of various attributes to consumers and the ideal levels of these attributes. Even if we consider that this methodology has received some degree of interest, some authors [47] have reported that the use of these scales can modify consumers’ perception of some products, since they are forced to focus on some previously selected attributes. This produces unreliable results. Alternatively, Van Trijp et al. (2007) requested consumers to directly assess the intensity of the different attributes [48]. However, this methodology can be difficult to understand for consumers, it can be perceived as quite unnatural and provides highly variable data.

Another alternative proposed by Adams et al. (2007) is the use of Check-All-that-Apply (CATA) question [49]. This question consists of a series of terms related to the products’ sensory features, occasions for use and other characteristics. Consumers try the product and tick all the terms that they consider appropriate to describe the product. This type of question is easy to understand for consumers, quick to complete, it requires minimum instructions and it can be considered as 'more natural’, since consumers tick words to describe the products instead of using scales to quantify the sensory intensity of a certain feature [50]. Nevertheless, so far, there is not enough information regarding the applicability of this methodology, its influence on the consumers’ acceptance scores or its effect on the number, order or type of terms used.

In summary, even if there are various available methodologies to determine the consumers’ perception of the food products sensory characteristics, there is not yet enough information to select the most appropriate methodologies.

Conclusions

Even though affective studies are one of the most important areas in the sensory evaluation field, acceptance as well as preference tests present some challenges that have led researchers to question their validity. These tests are not as simple as they may seem and results should be carefully interpreted.
There are still several issues to be studied that are generating future research studies in the area. Presently, modifications to the classical techniques are being evaluated and alternative techniques are being developed to determine the acceptance of certain products by consumers, but their applicability still has to be analyzed and results compared with the ones obtained using traditional tests.

References

49. Adams, J., Williams, A., Lancaster, B., Foley, M. 2007, Advantages and uses of
check-all-that-apply response compared to traditional scaling of attributes for salty snacks. In 7th Pangborn Sensory Science Symposium. Minneapolis, USA, 12-16 August 2007.