Selecting and training a panel to evaluate the rancid defect in soybean oil and fish hamburgers

C. Marques, A.S. dos Reis, L.D. da Silva, S.T. Carpes and M.L. Mitterer-Daltoé[✉]

Graduate Program in Chemical and Biochemical Technology Process, Chemistry Department,
Federal University of Technology, Km 01, 85503-390 - Paraná, Pato Branco, PR, Brazil.

[✉]Corresponding author: marinadaltoe@utfpr.edu.br

SUMMARY: Due to its composition of unsaturated and polyunsaturated fatty acids, oils and fats are very susceptible to oxidation, with rancidity being one of the main defects. Among the several existing methodologies to monitor oxidation in foods, sensory analysis stands out because of the sensitivity of responses. Accordingly, this study aimed to select and train a panel of expert assessors to identify the rancid flavor, showing the statistical steps in the process. Assessors were selected according to their individual performance, statistically analyzed by ANOVA and Tukey's mean comparison, Wald Sequential Analysis and chi-square test. The validation of the trained panel was carried out with the sensory analysis of fish burgers and soybean oil. F Value and box-plot graphic methods were effective for better visualization of results when used along with the mean and standard deviation tables. The final trained panel consisted of seven assessors, who have been able to identify and differentiate rancid taste in both samples used for validation.

KEYWORDS: Bioactive Phytochemicals; Fish; Flavors; Food Quality; Sensory Analysis

RESUMEN: *Selección y entrenamiento de un panel para evaluar el defecto rancio en aceite de soja y hamburguesas de pescado.* Debido a su composición en ácidos grasos insaturados y poliinsaturados, los aceites y grasas son muy susceptibles a la oxidación, siendo la rancidez uno de los principales defectos. Entre las diversas metodologías existentes para seguir la oxidación en los alimentos, el análisis sensorial destaca por la sensibilidad de las respuestas. Este estudio tuvo como objetivo seleccionar y capacitar a un panel de catadores entrenados para identificar el sabor rancio y se muestra estadisticamente los pasos del proceso. Los catadores fueron seleccionados de acuerdo a su capacidad individual y estadísticamente analizados por ANOVA comparando las medias mediante Tukey, análisis secuencial de Wald y prueba de chi-cuadrado. La validación del panel entrenado se realizó con el análisis sensorial de hamburguesas de pescado y aceite de soja. El valor F y la representación gráfica boxplot fueron eficaces para una mejor visualización de los resultados cuando se utilizan junto con las tablas de desviación media y estándar. El panel final estuvo formado por siete catadores que fueron capaces de identificar y diferenciar el sabor rancio en ambas muestras utilizadas para su validación.

PALABRAS CLAVE: Análisis sensorial; Calidad de los alimentos; Fitoquímicos Bioactivos; Pescado; Sabores

ORCID: Marques C http://orcid.org/0000-0002-0215-2679, dos Reis AS http://orcid.org/0000-0002-5508-8412, da Silva LD http://orcid.org/0000-0002-7538-4198, Carpes ST http://orcid.org/0000-0001-8625-7795, Mitterer-Daltoé ML http://orcid.org/0000-0002-6761-9321

1. INTRODUCTION

The sensory perception of food is placed first on the in-mouth transformation and that is the reason why it is so dynamic (Ares *et al.*, 2015) depending on the response of each individual. However, consumer preferences, acceptance and feedback are very important to the market (Ares *et al.*, 2015), which demonstrates the importance of sensory analysis in many areas, from the development of products to quality control (Ares *et al.*, 2015; Latreille *et al.*, 2006).

Therefore, selection and training are required to assess reliable measurements from individual reactions (Latreille *et al.*, 2006; Etaio *et al.*, 2010). To accredit a trained panel, the assessors must present repeatability (consensus), reproducibility, discrimination ability; and they must be able to notice differences that might seem small to consumers, with all technical competence acquired remaining over time (Etaio *et al.*, 2010; González *et al.*, 2007; López-Aguilar *et al.*, 2007). A number of five to eight trained assessors are sufficient for a reliable evaluation (Dutcosky, 2007).

The steps to be followed to achieve a reliable sensory trained panel are basically: assessor selection, basic, and specific training, assessor qualification and method validation (Etaio *et al.*, 2010). However, this entire process is known to be time consuming, expensive and not practical due to several aspects (Kamruzzaman *et al.*, 2013). It is common practice by many researchers that require precise measurements to hire an accredited trained panel (Sinesio *et al.*, 1990; Campo *et al.*, 2006; Kamruzzaman *et al.*, 2013; Borràs *et al.*, 2015) or to simply use instrumental methods for flavor testing (Lee and Choe, 2012) instead of training their own assessors.

Several studies, from the oldest (Banfield and Harries, 1975; Sinesio *et al.*, 1990; Lea *et al.*, 1995) to the newest (Latreille *et al.*, 2006) approach the statistic performance of trained assessors. However, references for selecting and training demonstrate the details of the process, including the temperature and time of rancidity, and are very limited in the literature (Latreille *et al.*, 2006; Elortondo *et al.*, 2007; Etaio *et al.*, 2010).

To our knowledge, there are no published studies reporting the selection and training of assessors for the rancid defect in oils and fats specifically, including the time and temperature of rancidity with results demonstrated statistically. The closest to it is the sunflower oil shelf-life estimation detailed by Houhg and Fiszman, (2005) where the focus is a demonstration of the cut-off point methodology.

The undesirable compounds known as *off-flavors* in oil and fat are products of oxidation reactions which also destroy essential fatty acids resulting in a loss in nutritional value in addition to the sensory rejection (Lee and Choe, 2012). The most common off-flavor is rancidity. Two of many contributors to the rancid flavor are the hexanal and nonanal compounds (Campo *et al.*, 2006; Ibrahim, 2001; López-Aguilar *et al.*, 2007) which lend a distinct odor to this defect. Human perception of oxidized flavors in food with high fat content is more accurate than chemical methods, and aids in the extent of deterioration evaluation when a well-trained panel is available (Sinesio *et al.*, 1990).

To encourage the consumption of fish, due to the high nutritional value of this type of meat, one of the strategies is to turn the fish into a practical product, such as the hamburger (Corbo *et al.*, 2008; Del Nobile *et al.*, 2009). One of the most questionable parameters of the stability of a fish product is lipid oxidation, due to its composition with high levels of polyunsaturated fatty acids that are more susceptible to oxidation due to the double bonds in the chain, a reaction that occurs even in low temperatures (Soares and Gonçalves, 2012; Wu and Mao, 2008).

Due to proven importance, the aim of this study is to select and train a panel of assessors specialized in recognizing the rancid defect taste in fish hamburgers and oils and to demonstrate the entire process, emphasizing on the statistical treatment of data.

2. MATERIALS AND METHODS

Approved by the Ethics Committee – CAAE number 48687815.0.0000.5547 – UTFPR, Pato Branco/PR, the study was performed with professors (6), undergraduates (12) and graduate students (8) of UTFPR, 26 subjects in total. The ones involved already had prior contact with the Sensory Analysis discipline facilitating the understanding of the analysis and the terms involved, but none had previously participated of a rancid flavor defect training session. Each assessor performed the analysis in a sensory cabin, properly lit and isolated from the others and from the sample preparation area, with access to a sink for sample disposal and water at will.

2.1. Global performance at selection

The procedure for selection included a previous interview before the difference test addressed to the product (Dutcosky, 2007). Questions about allergies and availability for training were made, along with filling in the form required by the Ethics Committee.

The selection of assessors was performed through the triangle test, a modality of sensory analysis called discriminative, which differentiates two samples that received different treatments (ASTM, 2010). The probability of accuracy is one-third. It is recommended to use 20 to 40 subjects for a solid result (NÆS *et al.*, 2010).

The test consisted of two samples: rancid and regular sunflower oil. The rancid oil was produced in an oven at 60 °C for 14 days (Houhg and Fiszman, 2005) with air circulation, within an open amber glass recipient, with 10% head space. Oils of the same brand were purchased at Pato Branco – PR local market. Three samples in random order were presented to the assessors, ten different times, where two samples were equal and they had to identify the different one by circling it. They had to taste it, advised to not smell it or try to differentiate by color somehow. The color was masked by black cups.

The main conditions were kept constant; 15 mL of oil (Borràs *et al.*, 2015) at 50 °C ± 2 °C (Houhg and Fiszman, 2005) in a plastic cup of 50 mL, coded with three random digits. Each replicate contained two equal samples oil (regular oil, no rancidity) and a different (rancid) alternating with two equal (rancid) and a different (no rancidity). Warm distilled water kept at 40 °C and plain crackers were provided to clean the palate between samples (Houhg and Fiszman, 2005; Borràs *et al.*, 2015). Each session lasted from 10 to 15 minutes, from 9 am to 12 pm.

The number of correct answers from the assessors so that there was a significant difference between the samples was found in a table based on the chi-square test; if the assessor reached the minimum of correct answers he was selected– 10 replicates requires 7 right answers ($p < 0.05$). Another statistical analysis applied to the selection was Wald Sequential Analysis, according to the graphical method (ISO, 2004) to further evaluate the assessors approved or rejected, and those who required training (Santana *et al.*, 2006).

The decision system was obtained through hypothesis testing (ISO, 2004) H_o: $p_1 \leq p_0$, and using the values $p_0 = 0.33$ (probability of a correct response when no perceptible difference exists), $p_1 = 0.67$ (probability of a correct response when a perceptible difference does exist), for α risk = 0:05 (probability of concluding that a perceptible difference exists when one does not) and β risk = 0.05 (probability of concluding that no perceptible difference exists when one does).

2.2. Training process

An unstructured scale of 10 cm was used for training, presented with the numbers 0 and 10 at the extremes (Houhg and Fiszman, 2005), where the assessors had a choice of where to place the intensity of the rancid defect of the sample on any point.

The training procedure consisted of three different days/stages of analysis, to calculate the accuracy of the answers and consistency of the team. On each day, four dilutions with rancid oil (0%, 10%, 50% and 100%) were provided to the assessors selected in a sufficient amount of 15mL (Borràs *et al.*, 2015), using plastic cups coded with random digits with three numbers.

Dilutions of 0% and 100% were presented as the extremes of the scale, where 100% represented the sample at its maximum rancidity (14 days – 60 °C) and 0% represented the regular oil sample with no rancidity. The remaining, 10% and 50% dilutions were placed between 0 and 10 cm by assessors, corresponding to little–none/much rancid flavor. This procedure was repeated three times within the same day in order to have mean and standard deviations for each day. Each session lasted 10-15 minutes.

2.3. Ability to discriminate between dilutions in training

Assuming that samples were only 10% and 50%, a paired test was applied to check whether there was a difference between them, and those who inverted the order of samples on the scale (placed 50% before 10%) had their responses considered incorrect. To check the difference, the bilateral paired test table was consulted ($p < 0.05$) (ASTM, 2010).

2.4. Individual performance of assessors

The responses were measured in centimeters along the 10 cm scale. ANOVA statistical analysis evaluated individual results, means and standard variations, giving the three days' precision using Tukey's mean comparison test ($p < 0.05$) performed by Statistica® software 12.7.

2.5. Panel performance and homogeneity

Similarly to the individual performance, the mean of each day's responses was calculated, with respective standard deviations, to evaluate panel homogeneity. Assessors that did not differ statistically ($p > 0.05$) from each other, by Tukey's mean comparison test, coinciding in the analysis of both samples, 10% and 50%, were selected for the final trained sensory panel.

2.6. Trained panel validation

Validation is important to test the panel reproducibility, which means that if the test is repeated after some time, or by another sensory panel trained exactly as in the present study, the results would not differ significantly (Lea *et al.*, 1995).

The validation was performed eight times with the products under study, fish burgers which had been stored for 30 days, and soybean oil with two distinct antioxidants. The burgers were made with grass carp fish meat (79.00%), where 33% of the total fatty acids were polyunsaturated (Wu and Mao, 2008); ice (10.00%), vegetable fat (5.00%), textured soy protein - TSP (3.00%), spices (2.99%) and BHT (0.01%). The water to hydrate the TSP was discounted from the ice. They were vacuum-packed,

and stored under refrigeration until the days of sampling, then frozen each day (initial – 0 days; 7, 14, 17, 21, 23, 25 and 30 days).

The burgers were thawed and grilled to serve to the assessors. The samples were cut into uniform sizes of about 1.5 cm³, and maintained at 75 °C (internal center) to the time of delivery (Mitterer-Daltoé *et al.* 2012) using plastic cups coded with three-digit random numbers. Water at room temperature was provided to clean the mouth between samples.

Samples of soybean oil with the tertiary butyl-hydroquinone (TBHQ) antioxidant 200 mg/Kg, 100 and 200 mg/Kg of *Quassia amara (Q.a)* extract were tested after 96 hours of rancidity (60 °C – oven) to detect any difference among them, regarding the rancid flavor.

An unstructured scale of 10 cm was applied again, for the distribution of burger and oil samples within range (different sheets), anchored in little–none/much rancid flavor. ANOVA was applied to the trained team's results to check for differences between samples (p < 0.05). The recognition of the difference between samples were compared for equivalence with the training rancid oil to validate the trained panel.

3. RESULTS

3.1. Global performance at selection

Twenty-six people attended the selection (9 males;17 females; ages ranging from 20–50), all of which were assessors (A), 15 of which, got seven right responses or more of the ten replicates provided, based on the chi-square table for the triangle discriminatory test, and they were considered suitable for training. By means of Wald Sequential Analysis, eight more assessors were between the acceptance lines ($a_x = 2.0789 + 0.5n$) and rejection lines ($r_x = 2.0789 + 0.5n$). These assessors obtained results that made them eligible for training within the applied statistics. At this point, three people were excluded, as they were found below or at the rejection line (Fig. 1).

3.2. Training process, ability to discriminate between dilutions in training

From the 23 assessors selected, 18 agreed to continue the training. According to the unilateral paired test table (p < 0.05) 13 assessors should set the right order of sample concentrations, 10% before 50%, in the unstructured 10 cm scale, so that, according to the paired test, the standard dilutions would present significant difference (Table 1) and become standards for the rest of the training process. Fourteen assessors got all the correct orders, verifying significant differences between dilutions.

To be approved, the assessors should have shown a total of nine correct responses, which means no change in the order of sample concentrations inside the triplicate, for every day of training.

3.3. Individual performance of assessors

According to the results of each assessor, the mean and the standard deviations of the responses were calculated in triplicate for each day of training through ANOVA, with the mean comparison analysis of p-values (Tukey), the mean of the tested three days which assessors presented homogeneity among the days (Table 2).
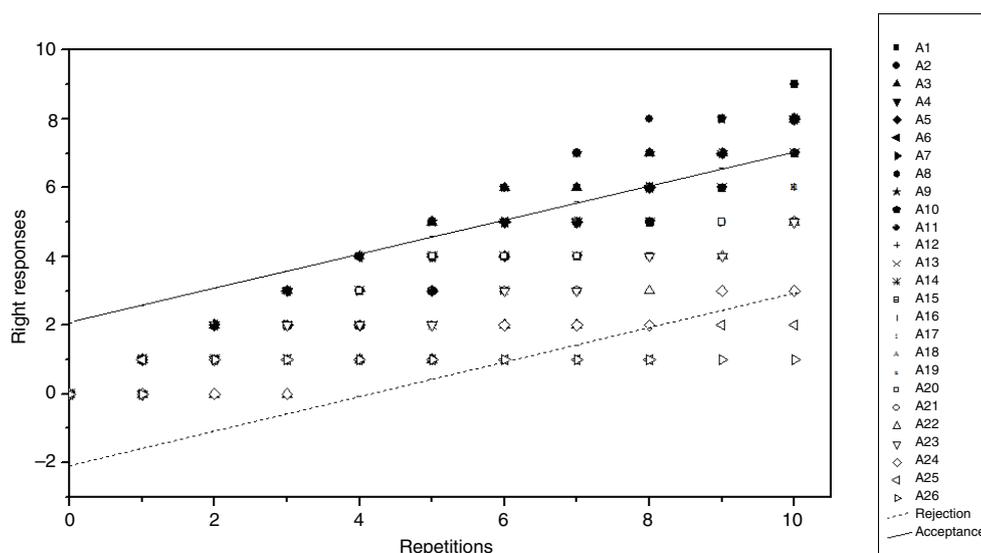

FIGURE 1. Wald Sequential Analysis for selection of assessors; α=β=0.05; p0 = 0.33; p1 = 0.67.

TABLE 1.   Number of correct responses regarding the order of samples in each triplicate, per day of training

Assessor	Day 1	Day 2	Day 3	Total
A1, A3, A4, A5, A6, A7, A8, A9, A10, A12, A13, A15, A17, A18	0 – Zero. No incorrect answers regarding the order of samples			9
A2*, A16*	3	2	2	7
A11*	3	2	3	8
A14*	2	3	3	8

* Eliminated

TABLE 2.   Inter-day precision of the rancid flavor (cm) in 10% and 50% oil dilutçions

Assessor	Dilution (%)	Day 1	Day 2	Day 3
A1*	10	$3.47^a \pm 0.46$	$3.53^a \pm 0,06$	$1.67^b \pm 0.06$
	50	$8.07^a \pm 0.21$	$8.37^a \pm 0.15$	$6.87^b \pm 0.11$
A3	10	$1.70^a \pm 0.36$	$1.07^a \pm 0.74$	$0.37^a \pm 0.23$
	50	$4.33^a \pm 2.40$	$6.30^a \pm 1.15$	$5.83^a \pm 4.37$
A4	10	$1.50^a \pm 0.30$	$1.53^a \pm 1.10$	$1.27^a \pm 0.68$
	50	$5.77^a \pm 1.42$	$5.33^a \pm 1.15$	$5.90^a \pm 0.66$
A5	10	$2.50^a \pm 1.28$	$3.63^a \pm 1.91$	$4.67^a \pm 1.53$
	50	$5.93^a \pm 2.50$	$6.83^a \pm 0.76$	$6.80^a \pm 0.72$
A6	10	$2.23^a \pm 0.25$	$2.03^a \pm 0.84$	$2.13^a \pm 0.60$
	50	$7.30^a \pm 0.75$	$7.47^a \pm 0.50$	$6.93^a \pm 0.93$
A7	10	$3.53^a \pm 2.15$	$4.43^a \pm 2.18$	$4.93^a \pm 7.20$
	50	$4.93^a \pm 2.41$	$7.70^a \pm 0.62$	$7.24^a \pm 0.80$
A8	10	$3.43^a \pm 1.85$	$3.60^a \pm 0.17$	$1.63^a \pm 0.23$
	50	$7.10^a \pm 2.13$	$7.33^a \pm 0.47$	$7.13^a \pm 0.32$
A9	10	$1.50^a \pm 0.44$	$0.93^a \pm 0.06$	$1.03^a \pm 0.15$
	50	$4.73^a \pm 0.67$	$4.00^a \pm 0.62$	$4.73^a \pm 0.46$
A10	10	$1.43^a \pm 0.31$	$1.77^a \pm 0.67$	$3.40^a \pm 1.40$
	50	$8.10^a \pm 0.78$	$5.67^a \pm 2.75$	$8.50^a \pm 0.87$
A12	10	$2.47^a \pm 0.49$	$2.50^a \pm 1.04$	$2.33^a \pm 0.29$
	50	$5.57^a \pm 1.20$	$5.80^a \pm 1.21$	$5.73^a \pm 0.68$
A13	10	$2.93^a \pm 0.59$	$1.63^a \pm 0.38$	$1.77^a \pm 1.29$
	50	$7.10^a \pm 0.79$	$5.40^a \pm 2.33$	$5.87^a \pm 1.27$
A15	10	$1.60^a \pm 0.53$	$1.50^a \pm 1.21$	$2.47^a \pm 1.27$
	50	$6.50^a \pm 1.50$	$6.90^a \pm 0.10$	$8.37^a \pm 0.85$
A17	10	$2.37^a \pm 0,15$	$2.23^a \pm 0.40$	$2.40^a \pm 0.53$
	50	$6.23^a \pm 0.64$	$6.00^a \pm 1.81$	$5.00^a \pm 0.62$
A18	10	$1.17^a \pm 0.42$	$1.57^a \pm 0.35$	$1.43^a \pm 0.21$
	50	$5.57^a \pm 0.51$	$5.60^a \pm 0.10$	$6.10^a \pm 0.30$

*A1: Eliminated; Same letters in the same line: means do not differ (Tukey p > 0.05).

Assessor A1, whose day 3 differed significantly from the others ($p < 0.01$), was eliminated at this stage of the statistical analysis. The remaining assessors exhibited homogeneity among days, with no significant differences among means.

For better visualization of this outcome, the F value of ANOVA (one-way) was calculated along with the p-value to test the individual performance of each assessor. F values higher than $F_{critical}$ (5.1432) demonstrate significant differences among days of training.

The inability of A1 ($F_{10} = 45.87$ and $F_{50} = 70.87$) was also computed by the F value (Figure 2) which tested the ability and the homogeneity among the days of the other assessors in training.

3.4. Panel performance and homogeneity

The individual performance, in terms of homogeneity among the days and differentiation of samples, removed five assessors thus far. The thirteen remaining were evaluated according to panel homogeneity. Those who did not differ from each other for both samples (rancid oil; 10% and 50% standard solution) were considered part of the final trained panel (Tukey p > 0.05; n = 9).

By analyzing the p-values, only seven (A3, A4, A9, A12, A13, A17 and A18) of the thirteen remaining assessors showed homogeneity in their responses (Table 3).

Figures 3 and 4 exhibit a graphical representation through box-plots of the variation in assessors' behavior, regarding the given dilutions. The trained panel (7 female assessors; ages ranging from 20 to 40) demonstrated proximity to 1 cm of the scale (dilution 10%) and 5cm (50% dilution), 1.70 ± 0.58 for 10% standard and 5.57 ± 0.51 for 50%. Through box-plot charts, it can be seen that the trained panel presented a combination of factors to be chosen: low variability of data, means (□) in the center of responses and proximity between means and medians (——).

Although A3 showed a large standard deviation for the 50% dilution (Fig. 4), he presented the same consistency with the panel. The assessor with the best results (box-plot and Table 3) was A9, revealing very low data variability, high precision and accuracy. The box-plot not only allows to visualize the behavior of each assessor, but also to behold the influence of each standard used. A better data distribution was found with less variation and values closer to 1 cm for the 10% standard solution of rancid oil (Fig. 3), which was not observed in the 50% standard.
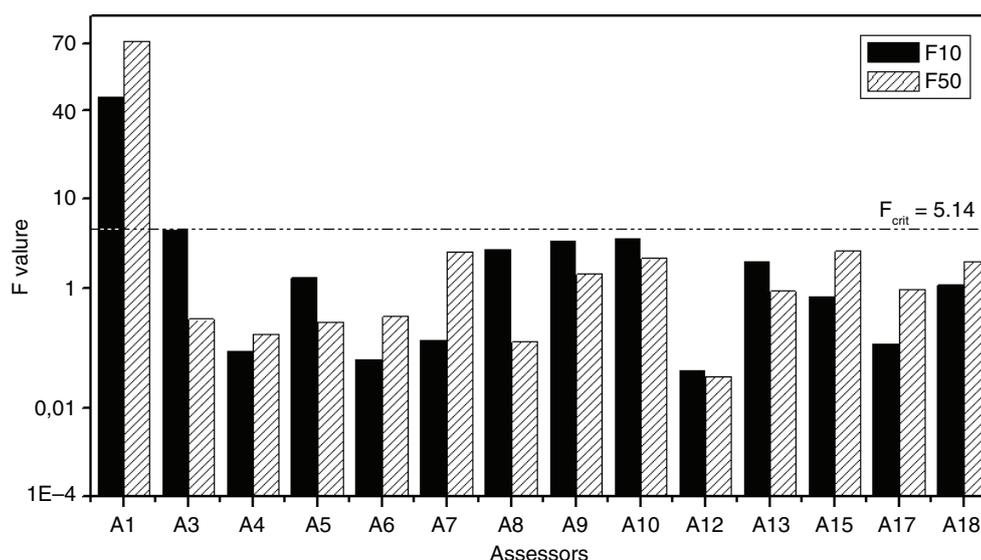
FIGURE 2.   Column chart for F values with logarithmic scale (significance level: 5%). The F values were divided by assessors and samples with concentrations of 10% (F10) and 50% (F50).

TABLE 3.   Panel consistency for 10% and 50% standard solutions of sunflower rancid oil

	10%	50%
A3	$1.04^c \pm 0.67$	$5.49^{ab} \pm 1.03$
A4	$1.43^c \pm 0.14$	$5.67^{ab} \pm 0.30$
A5*	$3.60^{ab} \pm 1.08$	$6.52^{ab} \pm 0.52$
A6*	$2.13^{bc} \pm 0.10$	$7.23^a \pm 0.27$
A7*	$4.30^a \pm 0.71$	$6.62^{ab} \pm 1.48$
A8*	$2.88^{abc} \pm 1.09$	$7.19^a \pm 0.13$
A9	$1.15^c \pm 0.30$	$4.49^b \pm 0.42$
A10*	$2.20^{bc} \pm 1.05$	$7.42^a \pm 1.53$
A12	$2.43^{abc} \pm 0.09$	$5.70^{ab} \pm 0.12$
A13	$2.11^{bc} \pm 0.71$	$6.12^{ab} \pm 0.88$
A15*	$1.85^{bc} \pm 0.53$	$7.25^a \pm 0.98$
A17	$2.33^{bc} \pm 0.09$	$5.74^{ab} \pm 0.65$
A18	$1.39^c \pm 0.20$	$5.75^{ab} \pm 0.30$

*Eliminated. Same letters in the same column: means do not differ (Tukey p > 0.05).

3.5. Trained panel validation

The validation of the trained panel was carried out in the analysis of samples: fish hamburgers and soybean oil with *Quassia amara* extract and synthetic antioxidant TBHQ.

Table 4 shows the equivalence of the rancid taste among samples and standards where burgers H1 and H2 did not differ from each other in rancid taste, nor when compared to H3, H4, H5. However, they showed a significant difference when compared to H6, H7 and H8, the last one considered more pronounced in this parameter. The sample H8, with 30 days of storage under refrigeration, manifested a significant difference with the other days except from the 25th and equivalent to the 50% standard. H3, H4 and H5 have a rancid flavor deffect corren-ponding to a 10% rancid oil standard.

The assessors' perception showed an equivalent degree of difference in standards and hamburger samples. At the first phase of training it became clear that 10% and 50% standards are different, proving that hamburguers are too. The trained panel was able to find differences equivalent to the training of samples, which have a more complex matrix, and therefore validates the assessors and the method.

For oil samples, after 96 hours in an oven at 60 °C, there was no significant difference among the samples tested (Soybean oil with TBHQ, soybean oil with 100 mg/Kg of *Quassia amara and* soybean oil with 200 mg/Kg of *Quassia amara)* regarding the rancid flavor (Table3). These results demonstrate that the concentration of *Quassia amara* extract used was effective in preventing rancidity, like the synthetic antioxidant. Even the lowest concentration tested, revealing that is possible to use 100 mg/Kg for economic purposes, with a similar effect and no significant difference. Also, it is evident that the rancidity found in all three samples corresponds to the 10% rancid oil standard.

4. DISCUSSION

The sequence of analysis applied was efficient for selection, training and panel validation for the rancid taste in oil and fish hamburgers. Wald Sequential Analysis proved to be more effective than the chi-square method for selection, especially since eight important assessors would be automatically
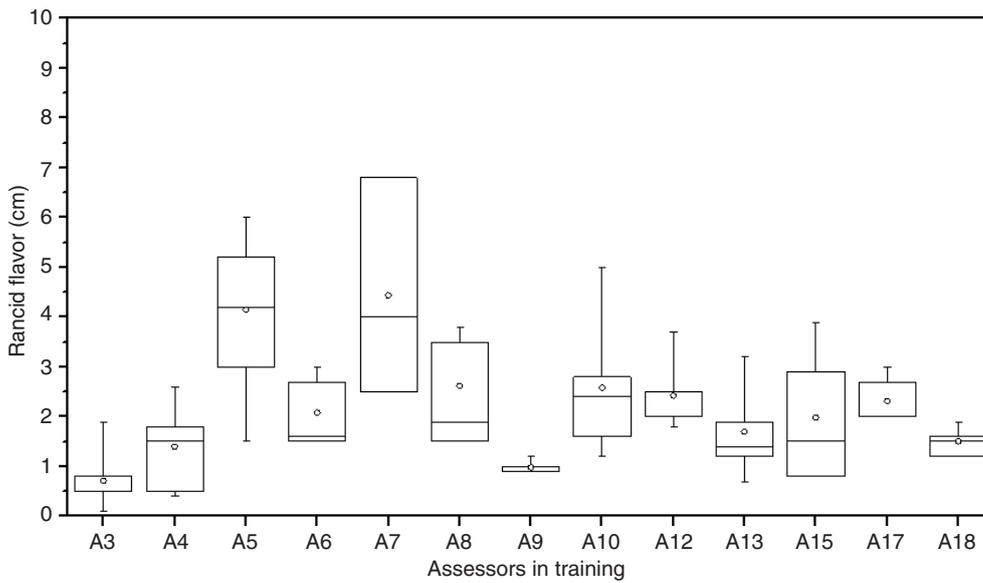
FIGURE 3. Box-plot for rancid oil sample diluted to 10%, from three-day training, separated by assessor; N = 9. Means (□); 1–99% Range ├──┤ ; Medians (——).
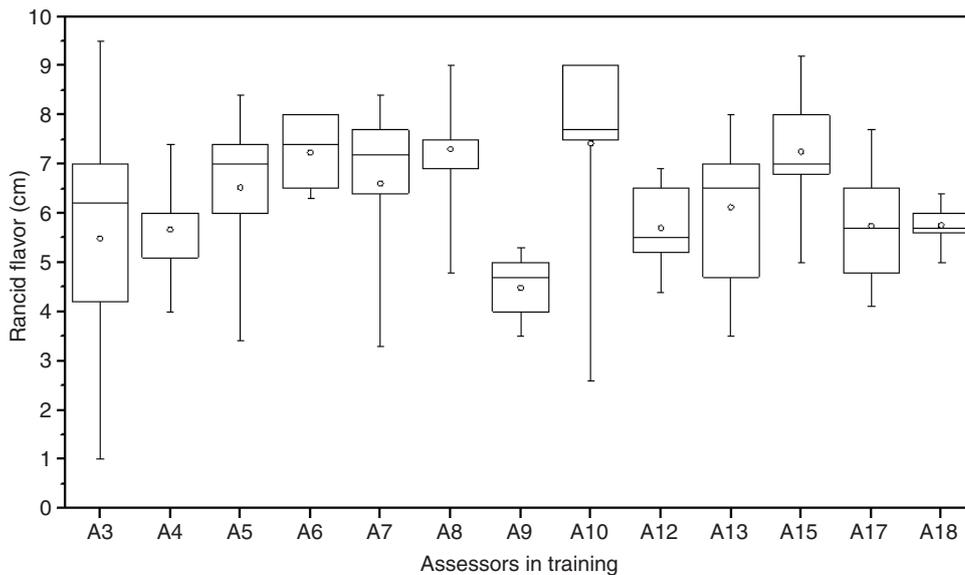

FIGURE 4. Box-plot for rancid oil sample diluted to 50%, from three-day training, separated by assessor; N = 9. Means (□); 1–99% Range ├──┤ ; Medians (——).

excluded already without the Wald test. Each assessor has a different sensitivity for flavors, and the Wald Analysis shows the individual performance graphically, giving a chance to those with potential.

Another point to be emphasized about the Wald Sequencial Analysis is the evaluation of assessors who had acuity at first, but saturated before the end of repetitions, such as A20 who presented saturation on repetitions 5 to 10, performing right only on the 9th and even then was selected for training.

As well as the selection, the training also proved to be an important part of the process as established by A1 and A2, both with the best results, 9 correct answers, in the triangle test, but didn't pass through training. This demonstrated the fact that it is possible to have an optimum performance during selection, but not the same happens during the second phase, which confirms the importance of training in order to select a panel.

ANOVA with Tukey mean comparisons represented the means and standard deviations, F value

TABLE 4. Values (mean ± standard deviation) for the rancid flavor of samples equivalent to standards

Sample	Rancid flavor (cm)
10% standard	*1.70^{deB} ± 0.58*
50% standard	*5.57^{abA} ± 0.51*
H1 – initial	0.00^{e} ± 0.00
H2 – 7 days	0.16^{e} ± 0.30
H3 – 14 days	0.76^{de} ± 0.62
H4 – 17 days	1.54^{de} ± 1.03
H5 – 21 days	2.30^{cde} ± 1.46
H6 – 23 days	3.31^{bcd} ± 2.02
H7 – 25 days	4.77^{abc} ± 2.60
H8 – 30 days	5.97^{a} ± 2.67
Soybean Oil + TBHQ	2.32^{B} ± 1.75
Soybean Oil + 100 mg/Kg*Q. a.*	2.92^{B} ± 2.35
Soybean Oil + 200 mg/Kg*Q. a.*	2.10^{B} ± 2.^{19}

Equal letters in the column show that the means do not differ significantly (Tukey $p < 0.05$; n = 7). Capital letters indicate oil samples; lower case letters for burger samples.

column plot representation, and box-plots completed the statistical analysis and proved effective in the training process. Graphic representations, such as the F value plot and box-plot, provided a better visualization of why only seven were chosen in the end, or why some were eliminated. Tables represented by means and standard deviations seem confusing at times, especially with a larger amount of data.

F values in the table format were used by Braghieri *et al.* (2012) to verify panel agreement for five attributes during a meat evaluation, proving to be an effective method for this purpose, since they found a high level of homogeneity between the panels tested. F values' column plot is called a suitable method to compare assessors' ability to detect differences between products. ANOVA one-way is the most standard way to treat the raw data (Næs *et al.*, 2010) and was proven valuable even before, with studies such as Lea *et al.* (1995), which measured validity in a sensory analysis.

The box-plot in the present work allowed better understanding and visualization in the variation of assessors' responses for each standard, also emphasizing the difference between standards. Regarding box-plot graphics, a study by Williamson *et al.* (1989) and Coli *et al.* (2015) mention the importance of the box-plot graphic method to highlight the visualization of data distribution, means and medians, and extreme values, which give a better idea about the response variation than tables.

By means of the box-plot it is noted that the standard deviations shown by 50% dilution (Figure 4) were higher when compared to the differences shown by 10% dilution (Figure 3), which emphasizes the

perception limit of the judges for rancid taste. There is the possibility that the 50% standard alters or saturates the assessors' sensations, as it is near the saturation threshold of the rancid flavor.

Sinesio *et al.* (1990) found a similar intensity of rancid flavor (around 6 cm out of a 9 cm scale) for sausages added with 63.5% rancid fat. In beef steaks, after only 9 days under refrigeration, a nine-assessor trained panel verified 11.4 to 21.4 points, from a 100-point unstructured scale, for rancid flavor (Campo *et al.*, 2006). This data confirms the rancidity levels found (around 6 out of a 10 cm scale for H8) in the analyzed burger.

For soybean oil, López-Aguilar *et al.* (2007) found 5.89 to 6.58 cm (in a 10-cm scale) for rancid flavor in commercial oils, with 12 weeks of analysis. In this study, trained assessors assigned lower values to the oils with the synthetic antioxidant TBQH and *Quassia amara* extract (100 and 200 mg/Kg) in 96 hours of analysis, confirming the efficiency of the tested antioxidants. Considering the results, because of their composition, the soybean oil and the fish burgers were highly susceptible to oxidation. A trained sensory panel for the rancid flavor would be an appropriate method for quality control, despite the time, effort and money applied.

The validation of the trained panel with a product is essential, because the products' matrix is often more complex than the solutions used in the selection and training, and this may confuse the panel. Elortondo *et al.* (2007) and Mitterer-Daltoé *et al.* (2012) validated the sensory panel with a product that recalled the sensation by which the panel had been trained. Elortondo *et al.* (2007) didn't use ANOVA for this purpose, only the percentage, where a sensation mentioned by 66.6% of assessors was considered a parameter, in which they were trained afterwards.

Seven proved to be a good number of trained assessors in the rancid flavor defect case. This number was not explored in depth in the literature, appearing in studies with other flavors and defects in descriptive analyses (Sinesio *et al.*, 1990; Campo *et al.*, 2006; López-Aguilar *et al.*, 2007). The trained panel proved accurate in differentiating samples with the rancid flavor defect according to the combined methods applied. In another study Etaio *et al.* (2010) started a selection for a wine trained panel with 31 assessors, finishing with 13 in the training, ending with seven experts as in the present study, but using only percentages of success; no ANOVA was applied. This proves the difficulty in obtaining a large number of expert assessors.

Because of this difficulty, researchers hire trained panels or use untrained assessors/consumers. Wang *et al.* (2012) used only 4 trained assessors to analyze the tenderness, chewiness and juiciness of pork meat samples. Borrás *et al.* (2015) used 8 trained assessors

to evaluate defects like rancidity and metallic in olive oils. None of the studies trained the assessors; they used a trained panel that already existed. For fish burgers, Corbo *et al.* (2008) and Del Nobile *et al.* (2009) used five and ten untrained assessors, respectively, to determine color, odor, texture, drip loss and general appearance values in the first case and overall quality in the second case.

The International Organization for Standardization usually gives some directions for the selection and training of sensory panels. Companies and universities in some countries follow the methods undergo a lengthy training process to further borrow or rent teams to research, showing that training a panel for a particular research is unusual for the reasons already mentioned. However, according to Resolution CNS/MS 196 (BRASIL, 1996) on the conduct of research involving humans in Brazil, the research subject must be free and voluntary, prohibited of any type of compensation, except for personal damages. However, it is possible to outsource the service, receiving only the outcome.

Therefore, the value and importance of a trained panel previously justified, along with the fact that in some countries such as Brazil there is no possibility of hiring a trained panel without third parties involved, or more financial resources, highlights the relevance of the results presented here which detail and discussed the selection and training of a sensory panel.

5. CONCLUSIONS

It was possible to select and train a panel for the rancid flavor, with seven assessors that were able to distinguish the target products through triangular discriminative analysis, unstructured and hedonic scale, using ANOVA for data statistics, in addition to the Wald Sequential Analysis, box-plot and bar graph for F values. This study is important for other studies that require trained panels, especially if training is necessary because it demonstrated clearly and in a reproducible way, step by step, how to achieve a trained panel.

ACKNOWLEDGMENTS

REFERENCES

Ares G, Jaeger SR, Antúnez L, Vidal L, Giménez A, Coste B, Picallo, A Castura, JC. 2015. Comparison of TCATA and TDS for dynamic sensory characterization of food products. *Food Res. Int*. **78**, 148–158. http://dx.doi.org/10.1016/j.foodres.2015.10.023

ASTM, 2010. *ASTM —Standard Test Method for Sensory Analysis — Triangle Test*. ASTM Int. 1–8. http://dx.doi.org/10.1520/E1885-04R11

Banfield CF, Harries JM. 1975. A technique for comparing judges' performance in sensory tests. *J. Food Technol*. **10**, 1–10. http://dx.doi.org/10.1111/j.1365-2621.1975.tb00001.x

Braghieri A, Piazzolla N, Carlucci A, Monteleone E, Girolami A, Napolitano F, 2012. Development and validation of a quantitative frame of reference for meat sensory evaluation. *Food Qual. Prefer*. **25**, 63–68. http://dx.doi.org/10.1016/j.foodqual.2012.01.007

BRASIL, 1996. *Diretrizes e normas regulamentadoras de pesquisas envolvendo os seres humanos, in: CNS/MS 196/96*. Diário Oficial da República Federativa do Brasil, Brasília, DF: Ministério da Saúde.

Borrás E, Mestres M, Aceña L, Busto O, Ferré J, Boqué R, Calvo A. 2015. Identification of olive oil sensory defects by multivariate analysis of mid infrared spectra. *Food Chem*. **187**, 197–203. http://dx.doi.org/10.1016/j.foodchem.2015.04.030

Campo MM, Nute GR, Hughes SI, Enser M, Wood JD, Richardson RI. 2006. Flavour perception of oxidation in beef. *Meat Sci*. **72**, 303–311. http://dx.doi.org/10.1016/j.meatsci.2005.07.015

Coli MS, Gil A, Rangel P, Souza ES, Oliveira MF, Cristina A, Chiaradia N, 2015. Chloride concentration in red wines: influence of terroir and grape type. *Food Sci. Technol*. **35**, 95–99. http://dx.doi.org/10.1590/1678-457X.6493

Corbo MR, Speranza B, Filippone A, Granatiero S, Conte A, Sinigaglia M, Del Nobile, 2008. Study on the synergic effect of natural compounds on the microbial quality decay of packed fish hamburger. *Int. J. Food Microbiol*. **127**, 261–7. http://dx.doi.org/10.1016/j.ijfoodmicro.2008.07.014

Del Nobile MA, Corbo MR, Speranza B, Sinigaglia M, Conte A, Caroprese M. 2009. Combined effect of MAP and active compounds on fresh blue fish burger. *Int. J. Food Microbiol*. **135**, 281–287. http://dx.doi.org/10.1016/j.ijfoodmicro.2009.07.024

Dutcosky SD, 2007. *Análise sensorial de alimentos*, 2nd ed. Champagnat, Curitiba.

Elortondo FJP, Ojeda M, Albisu M, Salmerón J, Etayo I, Molina M, 2007. Food quality certification: An approach for the development of accredited sensory evaluation methods. *Food Qual. Prefer*. **18**, 425–439. http://dx.doi.org/10.1016/j.foodqual.2006.05.002

Etaio I, Albisu M, Ojeda M, Gil PF, Salmerón J, Pérez Elortondo FJ. 2010. Sensory quality control for food certification: A case study on wine. Panel training and qualification, method validation and monitoring. *Food Cont*. **21**, 542–548. http://dx.doi.org/10.1016/j.foodcont.2009.08.011

González MM, Navarro T, Gómez G, Pérez RA, de Lorenzo C. 2007. Sensory assessment of table olive: I. Set up of a panel test and use of standarised scales. *Grasas Aceites* **58**. http://dx.doi.org/10.3989/gya.2007.v58.i3.176

Houhg G, Fiszman S. 2005. *Estimación de la vida útil sensoria de los alimentos*, 1st ed. Martín Impresores, S. L., Valencia - Espanha.

Ibrahim HMA. 2001. Acceleration of curing period of pastrami manufactured from buffalo meat: II-Fatty acids, amino acids, nutritional value and sensory evaluation. *Grasas Aceites* **52**. http://dx.doi.org/10.3989/gya.2001.v52.i2.382

ISO, 2004. *Sensory analysis - Methodology - Sequential analysis*, in: ISO 16820:2004. ANSI, New York, p. 10.

Kamruzzaman M, ElMasry G, Sun DW, Allen P, 2013. Non-destructive assessment of instrumental and sensory tenderness of lamb meat using NIR hyperspectral imaging. *Food Chem*. **141**, 389–396. http://dx.doi.org/10.1016/j.foodchem.2013.02.094

Latreille J, Mauger E, Ambroisine L, Tenenhaus M, Vincent M, Navarro S, Guinot C. 2006. Measurement of the reliability of sensory panel performances. *Food Qual. Prefer*. **17**, 369–375. http://dx.doi.org/10.1016/j.foodqual.2005.04.010

Lea P, Ródbotten M, Naes T. 1995. Measuring validity in sensory analysis. *Food Qual. Prefer*. **6**, 321–326. http://dx.doi.org/10.1016/0950-3293(95)00036-4

Lee E, Choe E. 2012. Changes in oxidation-derived off-flavor compounds of roasted sesame oil during accelerated storage in the dark. *Biocatal. Agric. Biotechnol*. **1**, 89–93. http://dx.doi.org/10.1016/j.bcab.2011.08.003

López-Aguilar JR, Valerio-Alfaro G, Monroy-Rivera JA, Medina-Juárez LA, O'Mahony M, Angulo O. 2007. Evaluation of a simple and sensitive sensory method for measuring rancidity in soybean oils. *Grasas Aceites* **57**, 149–154. http://dx.doi.org/10.3989/gya.2006.v57.i2.30

Mitterer-Daltoé ML, Treptow R, De O, Martins E, Martins VM V, Queiroz MI. 2012. Selecting and Training a Panel to Evaluate the Metallic Sensation of Meat. *Food Sci. Technol. Res*. **18**, 279–286. http://dx.doi.org/10.3136/fstr.18.279

Næs T, Brockhoff PB, Tomic O. 2010. *Statistics for Sensory and Consumer Science*, Wiley, Chichester. http://dx.doi.org/10.1002/9780470669181

Santana LRR, Santos LCS, Natalicio MA, Mondragon-Bernals OL, Elias EM, Silva CB, Zepka, LQ, Martins ISL, Vernaza MG, Castillo-Pizarro C, Bolini HMA. 2006. Perfil sensorial de iogurte light, sabor pêssego. *Ciência e Tecnol. Aliment*. **26**, 619–625. http://dx.doi.org/10.1590/S0101-20612006000300021

Sinesio F, Risvik E, Rodbotten M. 1990. Evaluation of panelist performance in descriptive profiling of rancid sausages: a multivariate study. *J. Sens. Stud*. **5**, 33–52. http://dx.doi.org/10.1111/j.1745-459X.1990.tb00480.x

Soares KMP, Gonçalves AA. 2012. Qualidade e segurança do pescado. *Rev. Inst. Adolfo Lutz* **71**, 1–10.

Wang Q, Lonergan SM, Yu C, 2012. Rapid determination of pork sensory quality using Raman spectroscopy. *Meat Sci*. **91**, 232–239. http://dx.doi.org/10.1016/j.meatsci.2012.01.017

Williamson DF, Parker RA, Kendrick JS. 1989. The box plot: A simple visual method to interpret data. *Ann. Intern. Med*. **110**, 916–921. http://dx.doi.org/10.7326/0003-4819-110-11-916

Wu T, Mao L. 2008. Influences of hot air drying and microwave drying on nutritional and odorous properties of grass carp (*Ctenopharyngodon idellus*) fillets. *Food Chem*. **110**, 647–653. http://dx.doi.org/10.1016/j.foodchem.2008.02.058