VERY BEYOND SUBJECTIVITY: THE LIMIT OF ACCURACY OF Q-GRADERS

Lucas Louzada Pereira[a*], Rogério Carvalho Guarçoni[b], Taís Rizzo Moreira[c], Luiz Henrique Bozzi Pimenta de Sousa[d], Wilton Soares Cardoso[d], Aldemar Polonini Moreli[d], Samuel Ferreira da Silva[c], Carla Schwengber ten Caten[a].

[a]Universidade Federal do Rio Grande do Sul – UFRGS*. Programa de Pós-Graduação em Engenharia de Produção. Av. Osvaldo Aranha, n 99/5º andar, Bom Fim, CEP: 90.035-190 – Porto Alegre – RS. E-mail: lucaslozada@hotmail.com, tencaten@producao.ufrgs.br.

[b]Instituto Capixaba de Assistência Técnica, Pesquisa e Extensão – INCAPER. Departamento de Estatística. Rua Afonso Sarlo, 160 - Bento Ferreira, CEP:29052-010 – Vitória. ES. E-mail: rogerio.guarconi@incaper.es.gov.br.

[c]Universidade Federal do Espírito Santo. Centro de Ciências Agrárias e Engenharias. Av. Gov. Lindemberg, 316 - Centro, Jerônimo Monteiro - ES, 29550-000. E-mail: taisr.moreira@hotmail.com, samuelfd.silva@yahoo.com.br.

[d]Instituto Federal do Espírito Santo. Departamento de Ciência de Alimentos. Avenida Elizabeth Minete Perim, S/N. Bairro São Rafael, Venda Nova do Imigrante - ES, CEP: 29375-000. E-mail: luizhenriquebozzi@hotmail.com, wilton.cardoso@ifes.edu.br, aldemar.moreli@ifes.edu.br.

*****Corresponding author: Lucas Louzada Pereira. E-mail: lucaslozada@hotmail.com.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/jtxs.12390

Authors' contribution

Conceived and designed the experiments: LLP. Performed the experiments: LLP, RCG and TRM. Analyzed the data: LHBPS, LLP and RCG. Contributed reagents/materials/analysis tools: TRM and WSC. Performed critical revision of the manuscript for important intellectual content: SFS, WSC, APM and CSC. Wrote the paper: LLP, RCG and TRM.

ABSTRACT

When it comes to the sensory analysis of specialty coffees, it is necessary to discuss the protocol of tests of the Specialty Coffee Association (SCA) and the formation of Q-Graders as reliable parameters in the sensory analysis. However, the training of these Q-Graders and the use of the SCA protocol has generated discussions in the scientific community and demonstrated the importance of talking about the level of professional precision as well as the use of the protocol. This study sought to understand the relation of accuracy and efficiency of the Q-graders and protocol in the sensory analysis of coffees. Three experiments were carried out to evaluate and test the hypotheses regarding the level of precision in relation to the shift, the number of samples and the judgment abilities of Q-Graders due to the quality of the specialty coffee and non-specialty coffee. These three experiments were performed by seven Q-Graders in an isolated environment during three consecutive days with 137 specialty coffee samples and 7 non-specialty coffee samples as defined by the SCA. The results indicate that the Q-Graders present high precision when evaluating excellent and outstanding coffees, as defined by the SCA. In addition, the effect of the shift did not exert fatigue on the Q-Graders. However, evaluation errors were made by the Q-Graders on very good coffee (77.00 – 80.00 points), which are considered as non-specialty by the SCA, thus allowing a more in-depth discussion on what would be the boundary between a specialty and non-specialty coffee.

Keywords: Q-Graders, Accuracy, SCA protocol, Sensory Aspects, Foods and *Coffea arabica* L.

Practical Applications

The article deals with the sensory evaluation process of coffees, and it has great importance to teaching and research institutions in Brazil, the largest producer, largest exporter and the country with the highest concentration of coffee scholars in the world. The article focuses on a new approach and a new way of looking at the sensory evaluation process of coffees, using the methodology of the Specialty Coffee Association to analyze the accuracy and efficiency of the Q-graders.

1. Introduction

Sensory analysis is a complex process because it is based on human perceptions, which means it is highly subjective and dependent on trained Q-Graders. According to Civille & Oftedal (2012), sensory evaluation concerns the human response to physical stimuli, since the sensory process can be simplified as follows: a stimulus (for example: food) reaches the mouth, at which point sensory signals are generated, integrated and sent to the brain. Thus, we have the complexity of the sensory analysis technique.

This process has developed over the last few years as a laboratory technology. The protocols describe minimal or ideal test conditions for sensory evaluation. These conditions generally maximize the control of sensory phenomena and minimize the intrusion of uncontrolled "other variables" in the environments where the analyses are employed. However, standardization has never reached high levels, and the details of how each laboratory and how they perform sensory work is not well documented (Meiselmun, 1993).

Recently, Pereira *et al.* (2017) discussed the relation of the environment (conversation / shift) on the sensory analysis process of specialty coffees, concluding that the level of precision of Q-Graders decreases the more disturbances (talk and noise)

there are during sensory analysis. For Amorim *et al*. (2010) the consonance among the evaluators on the use of the attributes of the same sample is an essential characteristic in the process of sensory analysis.

Several authors (Alvarado & Linnemann, 2010; Bhumiratana, Adhikari, & Chambers, 2011; Bosselmann *et al*., 2009; Cook, Hollowood, Linforth, & Taylor, 2005; Evangelista *et al*., 2014; Parnaiba, Ferreira, & Morais, 2009; Pereira, Chalfoun, Carvalho, & Savian, 2010; Ribeiro, Leitão, Ramalho, & Lidon, 2014) use sensory analysis to validate coffee quality. However, these authors do not discuss aspects inherent to calibration, precision, shift and sorting of the samples for analysis, nor the consistency in the use of the SCA test protocol.

For Dzung & Dzuan (2010), one of the main problems in using experts for sensory evaluation is that the qualification of "expert" (Q-Graders) is not well defined. According to ISO 856-2 (1994), experience is not the only main criterion of a specialist; they must be trained and have high sensory sensitivity. This fact is also verified by Chambers *et al*. (2004), because the authors believe that training time contributes to the level of Q-Graders' accuracy. This reinforces the need for Q-Graders to be highly trained and experienced in the sensory field because Q-Graders, by definition, are trained and tested for sensory sensitivity, unlike experts from other industries.

What is the limit of this sensory sensitivity? According to Alvarado & Linnemann (2010), the "taster" is a judge who performs the sensory evaluation. They are in charge of determining the final quality of the coffee, according to their perceptions. Consequently, they interfere on the price of the product with their personal opinion, accumulating tasting experiences over the years (Feria-Morales, 2002).

Due to the level of complexity and subjectivity in the sensory analysis process, there has been little discussion in the literature regarding the limit of samples that a Q-

grader can evaluate, and how the judgment capacity of these professionals decreases as a function of the use of the technique in daily life.

Therefore, this study was based on two hypotheses, these being inherent to the loss of accuracy of the Q-Graders[1] during the sensory analysis, as a function of the "shifts" (schedules) and their judgment abilities due to the quality of the specialty and non-specialty coffees[2]. The specialty coffees are divided into excellent and outstanding, and the non-specialty coffees[3] as very good.

The aim of this study was to discuss the intrinsic relations of the use of the Specialty Coffee Association (SCA) protocol. This is a new adaptation of the process that uses this tool, seeking to reduce subjectivity, aiming to increase the level of accuracy of the Sensory analysis using Q-Graders.

2. Materials and Methods

2.1 Selection of Samples

Three experiments were carried out with coffees varying between 77 and 87 points, for three consecutive days with two tables per shift (morning and afternoon). Twelve coffee samples were arranged on each of the tables, a total of 48 coffee samples per day of work, respecting the normal practice in coffee tasting rooms around the world.

[1] The Q Coffee System identifies quality coffees and brings them to market through a credible and verifiable system. A common standard for both Q Arabica Coffee, (Specialty Grade) and Q Robusta Coffee (Fine Robusta Grade) have resulted in a universally shared language and standard top-scoring lots.

[2] According to the Specialty Coffee Association (SCA), coffee which scores 80 points or above on a 100-point scale is graded as a specialty coffee. Specialty coffees come from the most exceptional quality beans harvested from the best of the world's Coffea arabica crops, roasted and brewed to enhance their flavor potential.

[3] The Specialty Coffee Association (SCA) defines coffees that do not reach 80 points as non-specialty coffees. However, this definition is not entirely clear and is debated. (SCA, 2018).

All samples were selected by a Q-Grader with more than 10 years of experience in sensory analysis (Head-Judge of international competitions), 15 days before the study.

The role of the Head-Judge is to organize and control the flow of analysis during the coffee tastings, and their grade was not computed for experiments. In the case of the present study, this Head-Judge only worked on the previous selection of the samples used. This process was adapted from the methodology of the Cup of Excellence Program of the Alliance for Coffee Excellence (2017). The selected coffees belong to producers that were awarded in the Cup of Excellence, that is, they come from a sample bank based on the sensory history of an international contest. The Head-Judge performed the screening based on the norms for the definition of sensory standards indicated by the Coffee Quality Institute, since he is trained as a Q-Grader.

A panel of three professionals trained in food engineering, who are experienced in the process of sensory analysis, prepared the sensory analysis samples. The professionals formulated the distribution of the samples, and had no contact with the Q-Graders, only with the Head-Judge.

The samples used in the present study belong to the sample database of the Laboratory of Analysis and Research in Coffee of the Federal Institute of Espírito Santo. All samples used are from the 2016/2017 harvest and were processed by the wet process method. The materials were harvested in regions that have been awarded with the Cup of Excellence. The coffees were selected among the 34 properties that make up the study laboratory group. In addition; the resources were financed through public sponsorship 469058/2014-5 of the National Council of Research and Development (CNPq), n. 17/2014.

The quality scale of the SCA (2013) was adopted to define the coffee standards. The scale it follows is: 70.25 to 79.75, very good coffee; 80.00 to 89.75, excellent coffee; and above 90.00, outstanding coffee. Although the SCA Quality Scale sets a coffee with averages of 70.25 to 79.75 as very good coffee, that same coffee is not framed as special. This practice is adopted in the quality contests of the Cup of Excellence program promoted by the Alliance for Coffee Excellence, where coffees with grades below 80 points may not appear as special. For this study, only the SCA protocol was adopted, standardizing the routine of cupping taste.

2.2 Preparation of samples:

For the sensory evaluation, coffee samples were prepared in the coffee sensory analysis laboratory (LAPC) of the Federal Institute of Espírito Santo, city of Venda Nova do Imigrante, Brazil, following the methodology of the Specialty Coffee Association (SCA). The coffee samples roasting processes were carried out using the Laborato TGP2 roaster and monitored with Agtron-SCA disk array. The roasting point of the samples was between the colors of discs # 65 and # 55 (SCA, 2013).

The roasting process was carried out 24 hours in advance (on each day of experiment) and the grinding respected the time of 8 hours of rest after the 24 hours of roasting. The samples were roasted according to the roasting routine established by the LAPC, with initial temperature of 150 °C (setpoint). The temperature was raised with an increase of 5 °C per minute until reaching 200 °C. The samples were cooled for 1 minute; after that, the air flow cycle remained constant at 50% of the total capacity of the Laboratto TGP2 roaster. All samples were roasted for 9 to 11 minutes and, after roasting and cooling, the samples remained sealed, according to the methodology established by the SCA. The coffee evaluation followed the SCA protocol.

The coffee samples were ground with a Bunn G3 electric grinder, with medium grain size. Each sample had 5 cups, and 8.25 grams of ground coffee was used in 150 ml of water, according to the midpoint of the optimal balance graph for the Golden Cup (SCA, 2013).

The infusion point of water occurred after the water reached 92-95 °C. The fresh water used for the sensory analysis of the coffees in both experiments was in accordance with the resolution by CONAMA n. 357/2005, which deals with the classification of water bodies. The Q-graders started the evaluations when the cup temperature reached 55 ºC, respecting the time of 4 minutes for the tasting after the infusion (SCA, 2013).

2.3 The Sensory Analysis Environment

Seven Q-Graders performed the sensory analysis. This number was chosen considering the minimum quantity of 6 Q-Graders proposed by Pereira *et al*. (2017).

The experiments were carried out in an isolated environment during three consecutive days with 137 specialty coffee samples, as defined by SCA, and 7 samples of non-specialty coffee, to evaluate and test the hypotheses.

Two identical samples (same sample) were arranged randomly at each test table. The test tables were organized with two samples repeated during the morning shift and two samples repeated during the afternoon shift. This organization of the experiment was used to evaluate the level of precision of the Q-Graders in relation to the samples of the three experiments.

The sample distribution during the three days of experiments is shown in Figure 1, in a simplified way. All samples were randomized after each session to prevent the Q-Graders from perceiving any repetition of the coffees.

First experiment: Two sessions of sensory analysis were prepared with 10 coffees with scores ranging from 85.00 to 89.00 points, and two samples with a score of 87.00 points were inserted by the Head-Judge as a repeated sample (same sample). In total, 24 samples were evaluated in the morning shift. The same procedure was followed in the afternoon shift.

Second experiment: Two sessions of sensory analysis were prepared with 10 coffees with scores ranging from 80.00 to 84.00 points, and two samples with a score of 82.00 points were inserted by the Head-Judge as repeated sample (same sample). In total, 24 samples were evaluated in the morning shift. The same procedure was followed in the afternoon shift.

Third experiment: Two sessions of sensory analysis were prepared with 10 coffees with scores ranging from 77.00 to 81.00 points, and two samples with a score of 79.00 points were inserted by the Head-Judge as repeated samples (same sample). In total, 24 samples were evaluated in the morning shift; the same procedure was followed in the afternoon shift.

During the three days of the experiment, the Q-Graders were not allowed to discuss their coffees grades during the process of sensory analysis. At the end of each shift, they were taken to a room to discuss their scores on the protocol. Q-Graders could not change the score after the discussion.

The tasting morning shift was considered from 8:00 a.m. to 12:00 p.m., with an interval of 1:30 hours (one hour and thirty minutes) for lunch, and the afternoon shift, from 1:30 p.m. to 5:30 p.m., totaling 8 hours of work per day.

During the sensory analysis sessions, the Q-Graders were allowed to ingest fruits such as papaya, banana, grape and strawberry, whole-grain bread, biscuit and water

biscuit, to regenerate their senses during the sensory analysis process. This was done according to the routine adopted by the Cup of Excellence program.

The experiment had an external team to prepare the samples, coding and randomization, and assembly of the coffee tables during the studies. The Q-Graders were only responsible for the sensory analysis.

2.4 Experimentation and Statistical Analysis

The three experiments were conducted with stratified coffee samples of the following quality: very good coffee (70.25 to 79.75 points), excellent coffee (80.00 to 89.75 points) and outstanding coffee, with score above 90.00 points. They were conducted with the following objectives:

First, to verify if there was consistency in the Q-Graders' scores in relation to the quality of the coffees tested in the two samples (repeated sample), both in the morning shift and in the afternoon shift. The t-test was used in the comparison of two averages of two samples;

Second, to verify if there was consistency in the Q-Graders' scores regarding the quality of the coffees tested in the two samples (repeated sample) between the morning shift (before) and the afternoon shift (after), using the paired t-test in the comparison between averages before (morning) and after (afternoon); and

Third: to evaluate if there was consistency in the Q-Graders' scores regarding the quality of the coffees tested in the two samples (repeated sample), in the morning and afternoon shifts and between the two shifts. Grouping analysis was used to verify if there was homogeneity within each group of Q-Graders and heterogeneity between them.

The averages of the characteristics of the two samples evaluated in the morning and in the afternoon were compared using the t-test at 5% probability. In the comparison of the averages in the morning shift and afternoon shift, the paired t-test was used at 5% probability.

In the evaluation of the similarity between the Q-Graders, the matrix was elaborated with the averages of the variables and, later, dendrograms were constructed using the Euclidean Mean distance to measure the consistencies between two points and the hierarchical group of complete bonding. The software SPSS version 19 was used for the statistical analysis.

3. Results

The results obtained in the three experiments of the three consecutive days are presented below. To summarize and facilitate the discussion, the data are presented per experiment. After that, the experiment is discussed, followed by a general discussion of the study.

3.1 First day of experiments, analysis and discussion

The Q-Graders evaluated the coffee sample with an 87-point score (according to the Head-Judge) and their repetitions (same sample) on the first day, in the morning shift and in the afternoon shift (without interaction between shifts). The results of this sensory analysis are presented in Table 1.

The results of the paired t-test of the first day, morning and afternoon shifts, with analysis of the tasting environment, are shown in Table 2. Table 3 shows the

contribution of the variables to the dendrogram formation of the Fig. 2, which shows the results of the Q-Graders' sensory analysis on the first day. Fig. 3 shows the dendrogram between all the shifts based on the paired t-test. The first experiment ends with Table 4, with the relative contribution of the variables to the construction of this dendrogram (Fig. 3).

Based on the first hypothesis presented in the study, the sensory analysis results on the first day indicate that the Q-Graders are accurate when evaluating specialty coffees, since the results did not present statistical difference between the sensory analysis of the repeated samples (same sample) in the morning shift and the afternoon shift. The results for comparation between the shifts (morning and afternoon) also did not shows statistical difference.

Table 3 shows the contribution of the variables that formed the dendrograms. Relevant cues are observed in the formation of the Q-graders' set of decisions, based on the morning results for the two repeated samples (same sample). The attributes balance, acidity and overall are the most impacting among the attributes. In the repeated sample the attributes change to fragrance, aftertaste and balance. In this case a reversal of position in the order of preference is noted. This indicative may be related to the gustatory sensitivity of the Q-Graders. In the afternoon, the most outstanding attributes were fragrance, aftertaste and overall, and in the second sample a change in the preference for acidity and balance, with overall remaining. These indicators show how the decision to describe a coffee is complex and subject to change. Although the final result is not altered, the perceptions and decisions about which variables most affect the scores are complex, indicating that any change in coffee composition can be decisive for a high, medium or low grade.

The results of the first day diverge from the data of Scholz et al. (2013), where the authors consider that qualitative evaluations are insufficient, since these Q-Graders need a long time to prepare and train to reach a level of agreement between them. The data for the first day indicates consistency in the methodology proposed in the Q-Graders training program considering that the Q-Graders training program is relatively short. This result confirms the effectiveness of the training and demonstrates the Q-Graders' ability to distinguish a coffee set with excellence by the SCA protocol, based on Q-Grader formation.

The mean tests indicate that the Q-Graders presented accuracy between the analysis of the repeated samples and between the shifts. However, when analyzing the results of Fig. 2 and 3, we note a variation of terminology of the scores between Q-Graders. These results indicate that the scores between Q-Graders are susceptible to oscillations, as long as the central axis of the scores does not deviate from the group, as observed in the statistical tests; this probably shows the preference of each Q-Grader in relation to the coffee that is being analyzed. These results resemble those results of Chollet & Valentin (2001). In that study, trained and untrained beer tasters had the same accuracy, varying only in terminology.

The variables that contributed the most to the construction of dendrograms also oscillated between six variables in the repeated sample and seven variables between the shifts. These results should be better analyzed regarding the number of attributes needed to compose the score of a special coffee, and the need for so many attributes to evaluate a coffee is not proven in the literature.

This observation about the variables of a sensory protocol has a logical basis, according to Sáenz-Navajas et al. (2010), in the case of wine, imbalances created by excessive acidity, astringency and aftertaste, are often the first notes of deficiency of an

evaluation. This occurs because gustatory sensations are critical to the perception of the quality of beverages such as wine and coffee. In this way, it is difficult to estimate the limit of acidity, intensity of flavor, sweetness and body that each Q-Grader understands as a desirable attribute in a sample. These oscillations are amenable to the nature of the taster in function of his sensory experience.

In the case of coffee, the results of the variables that contribute to the formation of the dendrogram indicate how these variables are fundamental in the formation of the clusters.

3.2 Second day of experiments, analysis and discussion

The results of the second day (experiment 2) with repetition of samples with a score of 81.5 points (according to the Head-Judge) are presented in Tables 5 - 8 and Fig. 4 and 5.

The results observed in Tables 5 and 6 do not present statistical differences between the means of the two samples, indicating that on the second day of analysis, the Q-Graders maintained the level of accuracy.

These data diverge from the results found by Amorim *et al.* (2010), where the authors used simulations and Pearson's correlations to measure the consistency of tasters in a quality contest. From the authors' perspective, the tasters were expected to have high consistency because they carry out frequent analyses and accumulate experience. However, the results do not discuss the adjustments and sensitivities observed between the attributes of the protocol and how these attributes are interpreted by the Q-Graders.

When the results of Tables 5 and 6 and Fig. 4 and 5 are observed, the consistency of the Q-Graders is verified, and again, the results show that the Q-Graders transit in groups of different sensory preferences. In other words, even with oscillations

between groups, the Q-Graders arrive at an optimum result. This is another indicator that the error may be in the excess of protocol variables, or even on the definition of what the attributes represent for a Q-Grader when he evaluates a coffee, and not on the Q-Graders' sensory analysis. The dendrograms show (Fig. 4 and 5) that the contributions of the variables is taken into account, according to Tables 7 and 8, indicating that the attributes oscillate in their weight in the construction of the presented model.

These oscillations reinforce that the results of food preference, as any human behavior, is complex, and can be influenced by different factors. Whenever a food is presented for consumption, the consumer has expectations about what the sensory qualities of the item may present and perceptions are influenced by factors external to the product (Geel et al. 2005).

In the same way, Q-Graders can base their choice among the best coffees, based on the quality history they practice in their daily lives, to serve different markets, customers and types of coffees. This can be a determining factor for the Q-Grader's decision at the time of the sensory analysis. As well as the perception of taste, as this results from the stimulation of the taste cells that make up the palate. These compounds correspond to sensations of "sour" (hydrogen ion concentration), "salty" (salt ions), "sweet" (organic compounds) and "bitter" (various types of compounds, including alkaloids and glycosides). Umami accreditation has been introduced as one of the basic taste categories along with sweet, sour, bitter and salty (Fuke & Shimizu, 1993; Jinap et al., 2010; Temussi, 2009). The differentiation by taste of individual foods is based, to a certain extent, on the balance between these five aspects, along with the contribution of smell (Mccorkindale & Mccorkindale, 2000).

In this way, the construction of the dendrograms is more than evident and explained, since the Q-Graders have their own evaluations, interpretations and perceptions. Such factors will be affected by the palate and by smell. Consequently, such changes in the organization of sensory groups are plausible, depending on the individual perceptions of each Q-grader.

3.3 Third day of experiments, analysis and discussion

The results of the third experiment are presented in Tables 9 - 12 and Fig. 6 and 7.

The results described in Table 9, referring to the coffee sample evaluated by the Q-Graders in the morning and afternoon shifts, did not present statistical difference according to the t-test. However, when the morning and afternoon shift effect was evaluated (Table 10), the results present statistical differences among all attributes of the protocol.

These results are directly related to the loss of precision when samples with lower quality coffees were evaluated, because in the samples with superior quality, no such differences were found. The results give us two new questions: Is there an effect of the turn? Could the accuracy have been reduced as a function of the level of the samples being very close to the cut-off score of the protocol (the limit recommended by the SCA protocol for specialty coffees currently is above 80 points)?

The observations collected with the data from the last experiment show a relevant perspective from the scientific point of view: the training and the formation of a Q-Grader must demonstrate that intermediate coffees require a lot of attention in the process of sensory analysis.

Inherent to these points, the results obtained by Pereira *et al*. (2017) indicated that in two days of experiments, sampling coffee in the morning and in the afternoon, the Q-Graders did not lose their accuracy either in the morning shift or in the afternoon shift. This reinforces that the shifts are not a problem in the sensory analysis of specialty coffees. However, the data on the third day shows another reality. The shift had a significant effect on the sensory analysis because the samples had lower marks in relation to the aromas and flavors. The Q-Graders may have evaluated the coffees with more haste, without maintaining the level of concentration due to the simplicity of the coffee attributes. This factor may be associated with fatigue in some Q-Graders.

According to Ross (2009), while methods have been developed which allow dynamic measurement of texture and flavor, more research needs to be conducted in relating these methods to dynamic sensory methods, that analyze other parameters, such as fatigue and stress. On this aspect, consumption tests are typically limited in the number of products that can be evaluated in a day due to consumer fatigue because they are not trained to evaluate many samples Lawless & Heymann, (2010). Demonstrating how fatigue can affect the accuracy of Q-Graders when they are subjected to a large number of samples.

4. Discussion

Many scientific studies use sensory analysis based on the SCA protocol and certified Q-Graders from the Coffee Quality Institute (CQI), for quality validation. Consequently, the flaws left in the studies made room for the theoretical evolution of science and technology to make it possible to consider replacing sensory evaluation with instrumental analyses.

However, so far, it was not possible to completely replace the sensory evaluation. The instruments may be sophisticated, sensory evaluation depends to a great extent on the practitioner's physiological and psychological characteristics (Zhang & Chen, 1997). This was evidenced by the presentation of the multivariate results,

indicating how the Q-Graders are sensitive and attentive to the small variations in the evaluation of the coffees. This is relevant if one has a high level of training for the correct use of the sensory analysis technique with the use of Q-Graders (Coulon-Leroy, Symoneaux, Lawrence, Mehinagic, & Maitre, 2017).

On the first and second day of the study, the data indicated that the quality measured and perceived by Q-Graders did not present statistical difference between these professionals. According to Giacalone *et al*. (2016), taste preferences are learned mainly after repeated exposure. Q-Graders, when trained and familiar with a certain quality standard, will tend to the right ones. In this way, the coffees presented on the first and second day had more special, refined and exotic characteristics to the palate and this may have contributed to the increase in the precision and concentration of the sensory analysis.

Unlike the third and final day of study, the quality of the coffees was lower at the limit of what the Specialty Coffee Association (SCA) considers to be specialty coffees. This double combination, fatigue and low quality of the coffees, was fatal to the loss of accuracy in the sensory analysis.

In the study by Alvarado & Linnemann (2010), the authors found variations among the evaluators. The cupper results showed changes in four aspects: fragrance, aroma, body and acidity, as well as some changes in results due to the cuppers' experience. However, the authors do not describe the relative effects of shift among other factors that should be more controlled.

The results observed in all dendrograms are in part in agreement with the results presented by Alvarado & Linnemann (2010) since the changes that were observed and the contribution of each variable to the construction of the model undergo drastic changes as the analyses are performed. In other words, the sensibility of human sensory

organs and subjective judgment played a decisive role in the results. Besides this factor, it is necessary to emphasize that all dendrograms always form two groups of Q-Graders, both between the mirror samples and between the paired t-test.

This indicates that sometimes there is a relative agreement and homogeneity between pairs, and at other times the Q-Graders invert groups. This finding confirms the subjectivity behind the sensory analysis, regarding individual choices and preferences. In another line of analysis, the work by DiDonfrancesco *et al*. (2013) describes that individual cuppers use different terms to describe the same coffee, indicating a certain inconsistency between the description of the coffee and the final quality given by the protocol. However, the study sought to address the relation between descriptive analysis and global analysis, making it clear that these divergences would be predictable. Describing the attributes of a product with so many volatile compounds cannot be easily standardized. Even so, sensory analysis has been a very important tool in the characterization of the different attributes of beverages in the coffee chain (Oliveira et al., 2013).

If the coffees are in a range considered to be of poor quality, it is evidenced that the effect of the shift can interfere with the final result of the Q-Graders analysis, which is a great scientific contribution. Instead of transferring the error to the Q-Grader, could the error reside on the threshold of what may or may not be a special coffee?

The results of the third day show the need for adaptations of the protocol for coffees below 80 points that do not present defects in the clean cup, sweetness and uniformity. Considering that none of the samples presented defects in these aspects, they were considered non-specialty coffees because of the coffee profile itself. The second perspective is related to the excess of variables in the protocol. The construction of the dendrograms evidences this, showing that the attributes “clean cup”, “sweetness”

and “uniformity” have no impact on the construction of the dendrograms. The other attributes present insignificant impact on the decision according to the data in Table 13.

This study opens a perspective of creating a system of scores with weighted averages of the attributes with the aim of creating more distant classes between the coffees, and even propose a reduction in the excess of variables of the protocol of the SCA.

5. Conclusion

Q-Graders were efficient in the sensory analysis for the first and second experiments without any alteration between the attributes evaluated by the SCA protocol of sensory analysis.

In the third experiment, the Q-Graders presented inconsistencies in their analysis, indicating that the lower quality of the coffees compromised the capacity of judgment in the sensory analysis.

A new reformulation for sensory analysis in scientific studies should be carried out, adjusting the protocol and calibrating the Q-Graders in order to maintain the accuracy of the method.

The construction of the dendrograms indicated that, despite the constant changes and formations of groups of Q-Graders, the variations exist, but in function of the individual preferences of each Q-Grader at the moment when the sensory analysis is being carried out.

The changes of groups between the Q-Graders evidenced how sensitive and complex the process of sensory analysis of specialty coffees is. This indicates that to achieve high levels of precision, the Q-Grader should keep the sensory variables used in the day to day in mind, so that errors are reduced throughout the process.

Acknowledgements

The authors thank the Federal Institute of Espírito Santo for supporting this research, and the translation and review of this article, as well as the Q-graders, who dedicated themselves to the realization of this study. We also thank the National Council for Scientific and Technological Development (469058/2014-5) - CNPq and Secretariat of Professional and Technological Education of the Ministry of Education - SETEC for making resources available for research.

6. Ethical statements

Conflicts of interest:

The authors declare no conflicts of interest.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent:

Written informed consent was obtained from all study participants.

References

Alliance for Coffee Excellence - ACE. (2017) Available in: <https://www.allianceforcoffeeexcellence.org/en/cup-of-excellence/>.

Alvarado, R. A., Linnemann, A. R. (2010). The predictive value of a small consumer panel for coffee- cupper Judgment., *112*(9), 2010.

Amorim, I. S., Ferreira, E. B., Lima, R. R., & Pereira, R. G. F. A. (2010). Monte Carlo based test for inferring about the unidimensionality of a Brazilian coffee sensory panel. *Food Quality and Preference*, *21*(3), 319–323. https://doi.org/10.1016/j.foodqual.2009.08.018

Bhumiratana, N., Adhikari, K., & Chambers, E. (2011). Evolution of sensory aroma attributes from coffee beans to brewed coffee. *LWT - Food Science and Technology*, *44*(10), 2185–2192. https://doi.org/10.1016/j.lwt.2011.07.001

Bosselmann, A. S., Dons, K., Oberthur, T., Olsen, C. S., Ræbild, A., & Usma, H. (2009). The influence of shade trees on coffee quality in small holder coffee agroforestry systems in Southern Colombia. *Agriculture, Ecosystems and Environment*, *129*(1–3), 253–260. https://doi.org/10.1016/j.agee.2008.09.004

Chollet, S., & Valentin, D. (2001). IMPACT OF TRAINING ON BEER FLAVOR PERCEPTION AND DESCRIPTION: ARE TRAINED AND UNTRAINED SUBJECTS REALLY DIFFERENT? *Journal of Sensory Studies*, *16*(6), 601–618. https://doi.org/10.1111/j.1745-459X.2001.tb00323.x

Civille, G. V., & Oftedal, K. N. (2012). Sensory evaluation techniques — Make "good for you" taste "good." *Physiology & Behavior*, *107*(4), 598–605. https://doi.org/10.1016/j.physbeh.2012.04.015

Cook, D. J., Hollowood, T. A., Linforth, R. S. T., & Taylor, A. J. (2005). Correlating

instrumental measurements of texture and flavour release with human perception. *International Journal of Food Science and Technology*, *40*(6), 631–641. https://doi.org/10.1111/j.1365-2621.2005.00973.x

Coulon-Leroy, C., Symoneaux, R., Lawrence, G., Mehinagic, E., & Maitre, I. (2017). Mixed Profiling: A new tool of sensory analysis in a professional context. Application to wines. *Food Quality and Preference*, *57*, 8–16. https://doi.org/10.1016/j.foodqual.2016.11.005

DiDonfrancesco, B., Guzman, N. G., Chambers Iv, E., Donfrancesco, D., Guzman, G., Chambers, N., … Gutierrez Guzman, N. (2013). A comparison of cupping and descriptive sensory analysis of Colombian brewed coffee.

Dzung, N. H., Dzuan, L. (2010). The role of sensory evaluation in food quality control, food research and development: a case of coffee study. Disponível em: < www4.hcmut.edu.vn/~dzung/Sensoryrole.pdf>. (2010), 2010.

Evangelista, S. R., Silva, C. F., Miguel, M. G. P. da C., Cordeiro, C. de S., Pinheiro, A. C. M., Duarte, W. F., & Schwan, R. F. (2014). Improvement of coffee beverage quality by using selected yeasts strains during the fermentation in dry process. *Food Research International*, *61*, 183–195. https://doi.org/10.1016/j.foodres.2013.11.033

Feria-Morales, A. M. (2002). Examining the case of green coffee to illustrate the limitations of grading systems/expert tasters in sensory evaluation for quality control. *Food Quality and Preference*, *13*(6), 355–367. https://doi.org/10.1016/S0950-3293(02)00028-9

Fuke, S., & Shimizu, T. (1993). Sensory and preference aspects of umami. *Trends in Food Science and Technology*, *4*(8), 246–251. https://doi.org/10.1016/0924-2244(93)90139-2

Geel, L., Kinnear, M., & de Kock, H. L. (2005). Relating consumer preferences to sensory attributes of instant coffee. *Food Quality and Preference*, *16*(3), 237–244. https://doi.org/10.1016/j.foodqual.2004.04.014

Giacalone, D., Fosgaard, T. R., Steen, I., & Münchow, M. (2016). "Quality does not sell itself." *British Food Journal*, *118*(10), 2462–2474. https://doi.org/10.1108/BFJ-03-2016-0127

Jinap, S., Ilya-Nur, A. R., Tang, S. C., Hajeb, P., Shahrim, K., & Khairunnisak, M. (2010). Sensory attributes of dishes containing shrimp paste with different concentrations of glutamate and 5′-nucleotides. *Appetite*, *55*(2), 238–244. https://doi.org/10.1016/j.appet.2010.06.007.

Lawless, H.T., Heymann, H. Sensory Evaluation of Food. Principles and Practices. 2° ed. Springer New York Dordrecht Heidelberg London. 2010. DOI 10.1007/978-1-4419-6488-5.

Mccorkindale, L. M., & Mccorkindale, L. M. (2000). What Is Taste ?

Meiselmun, H.L. (1993). Critical Evaluation of sensory techniques. Food Quality and Preference , Vol. 4, p. 353-358. (1993), *4*, 1993.

Oliveira, P., Oliveira, P. D., Borém, F. M., Isquierdo, E. P., Giomo, G. da S., Lima, R. R. de, & Cardoso, R. A. (2013). *Coffee science*. *Coffee Science* (Vol. 8). Universidade Federal de Lavras.

Parnaiba, P. F.; Ferreira, D. F.; Morais, A. R. (2009). TAMANHO OTIMO DE PARCELAS EXPERIMENTAIS: PROPOSICAO DE METODOS DE ESTIMACAO. *Rev. Bras. Biom.*, *27*, 255–268.

Pereira, L. L., Cardoso, W. S., Guarçoni, R. C., da Fonseca, A. F. A., Moreira, T. R., & Caten, C. S. ten. (2017). The consistency in the sensory analysis of coffees using Q-graders. *European Food Research and Technology*.

https://doi.org/10.1007/s00217-017-2863-9

Pereira, M. C., Chalfoun, S. M., Carvalho, G. R. de, & Savian, T. V. (2010). Multivariate analysis of sensory characteristics of coffee grains (Coffea Arabica L.) in the region of upper Paranaíba. *Acta Scientiarum. Agronomy*, *32*(4), 635–641. https://doi.org/10.4025/actasciagron.v32i4.4283

Ribeiro, V. S., Leitão, A. E., Ramalho, J. C., & Lidon, F. C. (2014). Chemical characterization and antioxidant properties of a new coffee blend with cocoa, coffee silverskin and green coffee minimally processed. *Food Research International*, *61*, 39–47. https://doi.org/10.1016/j.foodres.2014.05.003

Ross, C. F. (2009). Sensory science at the human-machine interface. *Trends in Food Science and Technology*. https://doi.org/10.1016/j.tifs.2008.11.004

Sáenz-Navajas, M.-P., Ferreira, V., Dizy, M., & Fernández-Zurbano, P. (2010). Characterization of taste-active fractions in red wine combining HPLC fractionation, sensory analysis and ultra performance liquid chromatography coupled with mass spectrometry detection. *Analytica Chimica Acta*, *673*(2), 151–159. https://doi.org/10.1016/j.aca.2010.05.038

SCA. Specialty Coffee Association. Protocols. (2013). January, 23, 2013. Disponível em: < http://www.scaa.org/PDF/resources/cupping-protocols.pdf>. Acessado em 03/02/2016. (2013), 2016.

Scholz, M. B. dos S., Figueiredo, V. R. G. de, Silva, J. V. N. da, & Kitzberger, C. S. G. (2013). *Coffee science*. *Coffee Science* (Vol. 8). Universidade Federal de Lavras.

Temussi, P. A. (2009). Sweet, bitter and umami receptors: a complex relationship. *Trends in Biochemical Sciences*, *34*(6), 296–302. https://doi.org/10.1016/j.tibs.2009.02.005

Zhang, J., & Chen, Y. (1997). Food sensory evaluation employing artificial neural

networks. *Sensor Review*, *17*(2), 150–158. https://doi.org/10.1108/02602289710170320

Table 1: Averages of the sensory variables evaluated in experiment 1 in the samples of the morning and afternoon shifts.

Sensory characteristics and standard deviation			Sensory characteristics and standard deviation		
Attributes	Average ± SD (n=7 Q-Graders) [(1)]		Attributes	Average ± SD (n=7 Q-Graders) [(1)]	
	Sample 1 (Morning)	Sample 2 (Morning)		Sample 1 (Afternoon)	Sample 2 (Afternoon)
Fragrance	8.21 A / 0.2249	8.07 A / 0.2378	Fragrance	8.14 A / 0.1967	8.32 A / 0.2378
Flavor	8.46 A / 0.2673	8.17 A / 0.3450	Flavor	8.32 A / 0.4261	8.21 A / 0.4190
Aftertaster	8.25 A / 0.4564	7.92 A / 0.4725	Aftertaster	8.00 A / 0.4082	8.00 A / 0.2887
Acidity	8.18 A / 0.2782	8.04 A / 0.3363	Acidity	8.04 A / 0.4190	8.29 A / 0.3037
Body	8.04 A / 0.2673	7.96 A / 0.2673	Body	7.86 A / 0.4046	8.14 A / 0.2835
Uniformity	10	10	Uniformity	10	10
Balance	8.36 A / 0.5373	8.25 A / 0.5401	Balance	8.25 A / 0.5590	8.11 A / 0.3181
Clean Cup	10	10	Clean Cup	10	10
Sweetness	10	10	Sweetness	10	10
Overall	8.54 A / 0.3660	8.25 A / 0.4564	Overall	8.29 A / 0.4190	8.29 A / 0.3660
Global Score	88.04 A / 1.3650	86.68 A / 2.1199	Global Score	86.89 A / 1.4709	87.36 A / 1.1534

[(1)] The averages of the characters measured in the two samples followed by the same letter in the row do not differ according to the t test, at 5% probability.

Table 2: Averages of the sensory variables of experiment 1 by paired T-test, in the morning and afternoon shifts.

	Sensory characteristics and standard deviation	
	Average ± SD (n=7 Q-Graders) [(1)]	
	Before (Morning)	After (Afternoon)
Fragrance	8.14 A / 0.2344	8.23 A / 0.2292
Flavor	8.32 A / 0.3315	8.27 A / 0.4098
Aftertaster	8.09 A / 0.4764	8.00 A / 0.3397
Acidity	8.11 A / 0.3056	8.16 A / 0.3748
Body	8.00 A / 0.2594	8.00 A / 0.3669
Uniformity	10,00	10,00
Balance	8.30 A / 0.5205	8.18 A / 0.4432
Clean Cup	10,00	10,00
Sweetness	10,00	10,00
Overall	8.39 A / 0.4242	8.29 A / 0.3780
Global Score	87.36 A / 1.8520	87.13 A / 1.2925

[(1)] The averages of the before and after characters followed by the same letter in the line do not differ according to the t-test, at 5% probability.

Table 3: Contribution of the variables in the construction of the dendrograms of Fig. 2, in relation to experiment 1, under the variables evaluated in the morning and afternoon shifts.

Attributes	Sample 1 Morning Contribution (%)	Attributes	Sample 2 Morning Contribution (%)	Attributes	Sample 1 Afternoon Contribution (%)	Attributes	Sample 2 Afternoon Contribution (%)
Global Score	4.76	Global Score	0.0	Global Score	14.3	Global Score	0.00
Fragrance	9.52	Fragrance	28.6	Fragrance	19.0	Fragrance	9.52
Flavor	9.52	Flavor	0.00	Flavor	0.00	Flavor	9.52
Aftertaster	9.52	Aftertaster	14.3	Aftertaster	19.0	Aftertaster	4.76
Acidity	19.0	Acidity	9.52	Acidity	9.52	Acidity	28.6
Body	0.00	Body	4.76	Body	4.76	Body	9.52
Balance	33.3	Balance	28.6	Balance	4.76	Balance	14.3
Overall	14.3	Overall	14.3	Overall	28.6	Overall	23.8

Table 4: Total contribution of the variables for dendrogram formation based on the results of the shifts, morning and afternoon for experiment 1.

Attributes	Contribution (%)	Attributes	Contribution (%)
Global Score	0.00	Global Score	4.76
Fragrance	23.8	Fragrance	0.00
Flavor	14.3	Flavor	4.76
Aftertaster	4.76	Aftertaster	23.8
Acidity	19.0	Acidity	9.52
Body	4.76	Body	14.3
Balance	19.0	Balance	9.52
Overall	14.3	Overall	33.3

Table 5: Averages of the sensorial variables evaluated in experiment 2 in the samples of the morning and afternoon shifts.

Sensory characteristics and standard deviation			Sensory characteristics and standard deviation		
	Average ± SD (n=7 Q-Graders) [1]			Average ± SD (n=7 Q-Graders) [1]	
Attributes	Sample 1 (Morning)	Sample 2 (Morning)	Attributes	Sample 1 (Afternoon)	Sample 2 (Afternoon)
Fragrance	7.32 A / 0.4725	7.21 A / 0.3934	Fragrance	7.2143 A/ 0.3934	7.3929 A/ 0.3780
Flavor	7.68 A / 0.4499	7.32 A / 0.4499	Flavor	7.4286 A/ 0.5537	7.2857 A/ 0.4190
Aftertaster	7.36 A / 0.2835	7.00 A / 0.5401	Aftertaster	7.0000 A/ 0.3819	7.1071 A/ 0.3181
Acidity	7.18 A / 0.4009	7.11 A / 0.4756	Acidity	7.2857 A/ 0.2673	7.0714 A/ 0.3450
Body	7.11 A / 0.4756	7.21 A / 0.6196	Body	7.2143 A/ 0.2673	7.2857 A/ 0.3934
Uniformity	10	10	Uniformity	10	10
Balance	7.61 A / 0.4756	7.36 A / 0.4756	Balance	7.3214 A/ 0.4725	7.1071 A/ 0.3493
Clean Cup	10	10	Clean Cup	10	10
Sweetness	10	10	Sweetness	10	10
Overall	7.57 A / 0.3740	7.25 A / 0.4330	Overall	7.3571 A/ 0.4970	7.4643 A/ 0.3037
Global Score	81.82 A / 1.1701	80.46 A / 2.3867	Global Score	80.82 A / 2.1492	80.71 A / 1.3419

[1] The averages of the characters measured before and after, followed by the same letter in the line do not differ according to the t test, at 5% probability.

Table 6: Averages of the sensory variables of experiment 02 by paired t-test, morning and afternoon shifts.

	Sensory characteristics and standard deviation	
	Average ± SD (n=7 Q-Graders) [1]	
	Before (Morning)	After (Afternoon)
Fragrance	7.27 A / 0.4214	7.30 A / 0.3820
Flavor	7.50 A / 0.4703	7.36 A / 0.4775
Aftertaster	7.18 A / 0.4539	7.05 A / 0.3422
Acidity	7.14 A / 0.4242	7.18 A / 0.3167
Body	7.16 A / 0.5336	7.25 A / 0.3252
Uniformity	10,00	10,00
Balance	7.48 A / 0.4750	7.21 A / 0.4144
Clean Cup	10,00	10,00
Sweetness	10,00	10,00
Overall	7.41 A / 0.4230	7.41 A / 0.3996
Global Score	81.14 A / 1.9383	80.77 A / 1.7222

[1] The averages of the characters measured before and after, followed by the same letter in the line do not differ according to the t-test, at 5% probability.

Table 7: Contribution of the variables for the construction of the dendrograms of Fig. 4, in relation to experiment 2, under the variables evaluated in the morning and afternoon shifts.

Attributes	Contribution (%)	Attributes	Contribution (%)	Attributes	Contribution (%)	Attributes	Contribution (%)
Fragrance	14.3	Fragrance	38.1	Fragrance	28.6	Fragrance	28.6
Flavor	23.8	Flavor	0.00	Flavor	9.52	Flavor	9.52
Aftertaster	4.76	Aftertaster	0.00	Aftertaster	19.0	Aftertaster	19.0
Acidity	4.76	Acidity	0.00	Acidity	9.52	Acidity	9.52
Body	9.52	Body	38.1	Body	4.76	Body	4.76
Balance	14.3	Balance	19.0	Balance	14.3	Balance	14.3
Overall	23.8	Overall	4.76	Overall	14.3	Overall	14.3
Global Score	4.76	Global Score	0.00	Global Score	0.00	Global Score	0.00

Table 8: Contribution of the variables for the construction of the dendrograms of Fig. 5, in relation to experiment 2, under the variables evaluated in the morning and afternoon shifts.

Attributes	Contribution (%)	Attributes	Contribution (%)
Fragrance	23.8	Fragrance	9.52
Flavor	28.6	Flavor	19.0
Aftertaster	0.00	Aftertaster	19.0
Acidity	0.00	Acidity	0.00
Body	14.3	Body	9.52
Balance	9.52	Balance	4.76
Overall	23.8	Overall	38.1
Global Score	0.00	Global Score	0.00

Table 9: Averages of the sensory variables evaluated in the experiment 3 in the samples of the morning and afternoon shifts.

Sensory characteristics and standard deviation			Sensory characteristics and standard deviation		
Attributes	Average ± SD (n=7 Q-Graders) [1]		Attributes	Average ± SD (n=7 Q-Graders) [1]	
	Sample 1 (Morning)	Sample 2 (Morning)		Sample 1 (Afternoon)	Sample 2 (Afternoon)
Fragrance	7.46 A / 0.5089	7.18 A / 0.3740	Fragrance	7.07 A / 0.5147	6.96 A / 0.4661
Flavor	7.39 A / 0.3493	7.36 A / 0.2835	Flavor	7.14 A / 0.3780	7.07 A / 0.1220
Aftertaster	7.14 A / 0.2440	7.07 A / 0.3134	Aftertaster	6.79 A / 0.4880	6.71 A / 0.466
Acidity	7.18 A / 0.5147	7.14 A / 0.3780	Acidity	6.86 A / 0.3780	6.82 A / 0.5147
Body	7.07 A / 0.5345	7.39 A / 0.5175	Body	6.82 A / 0.4261	6.79 A / 0.5669
Uniformity	10	10	Uniformity	10	10
Balance	7.43 A / 0.3450	7.50 A / 0.4330	Balance	7.11 A / 0.4532	6.89 A / 0.4046
Clean Cup	10	10	Clean Cup	10	10
Sweetness	10	10	Sweetness	10	10
Overall	7.14 A / 0.4046	7.39 A / 0.2835	Overall	7.04 A / 0.5089	6.96 A / 0.3660
Global Score	80.82 A / 1.5660	81.04 A / 1.6673	Global Score	78.82 A / 1.7000	78.21 A / 2.0383

[1] The averages of the characters measured before and after, followed by the same letter in the line do not differ according to the t-test, at 5% probability.

Table 10: Averages of sensory variables of experiment 3 by paired T-test, morning and afternoon shifts.

	Sensory characteristics and standard deviation	
	Average ± SD (n=7 Q-Graders) [1]	
	Before (Morning)	After (Afternoon)
Fragrance	7.32 A / 0.4539	7.02 B / 0.4750
Flavor	7.38 A / 0.3062	7.11 B / 0.2724
Aftertaster	7.11 A / 0.2724	6.75 B / 0.4599
Acidity	7.16 A / 0.4342	6.84 B / 0.4342
Body	7.23 A / 0.5323	6.80 B / 0.4822
Uniformity	10,00	10,00
Balance	7.46 A / 0.3780	7.00 B / 0.4274
Clean Cup	10,00	10,00
Sweetness	10,00	10,00
Overall	7.27 A / 0.3598	7.00 B / 0.4274
Global Score	80.93 A / 1.5579	78.52 B / 1.8305

[1] The averages of the characters measured before and after, followed by the same letter in the line do not differ according to the t-test, at 5% probability.

Table 11: Contribution of the variables for the construction of the dendrograms of Fig. 6, in relation to experiment 3, under the variables evaluated in the morning and afternoon shifts.

Attributes	Contribution (%)	Attributes	Contribution (%)	Attributes	Contribution (%)	Attributes	Contribution (%)
Global Score	4.76	Global Score	4.6	Global Score	9.52	Global Score	0.00
Fragrance	19.0	Fragrance	0.00	Fragrance	42.9	Fragrance	42.9
Flavor	14.3	Flavor	28.6	Flavor	14.3	Flavor	9.52
Aftertaster	19.0	Aftertaster	4.76	Aftertaster	0.00	Aftertaster	0.00
Acidity	0.00	Acidity	0.00	Acidity	4.76	Acidity	0.00
Body	0.00	Body	19.0	Body	9.52	Body	0.00
Balance	42.9	Balance	33.3	Balance	19.0	Balance	19.0
Overall	0.00	Overall	9.52	Overall	0.00	Overall	28.6

Table 12: Total contribution of the variables for dendrogram formation based on the results of the shifts, morning and afternoon, for experiment 3.

Attributes	Contribution (%)	Attributes	Contribution (%)
Fragrance	4.76	Fragrance	52.4
Flavor	28.6	Flavor	9.52
Aftertaster	4.76	Aftertaster	0.00
Acidity	0.00	Acidity	0.00
Body	4.76	Body	0.00
Balance	38.1	Balance	19.0
Overall	9.52	Overall	14.3
Global Score	9.52	Global Score	4.76

Table 13: Summary of the relative contribution of the sensorial variables for the construction of the dendrograms between the morning and afternoon analyses in the three days of study.

Attributes	Occurrences of variables in the construction of dendrograms	Maximum accumulation (%)
Fragrance	5	52.4
Flavor	5	28.6
Aftertaste	4	23.8
Acidity	2	19.0
Body	5	14.3
Balance	7	38.1
Overall	6	38.1
Clean Cup	0	0.0
Sweetness	0	0.0
Uniformity	0	0.0
Global Score	4	9.5

Figure 1: Distribution of samples for sensory analysis.
*S.S. – Same sample.

Figure 2: Dendrograms obtained between the Q-Graders, from the sensory analysis performed on the morning shift (left side) for sample 1 (1A) and sample 2 (1B), and the afternoon shift (right side) for sample 1 (1C) and sample 2 (1D) relative to experiment 1 on the first day.

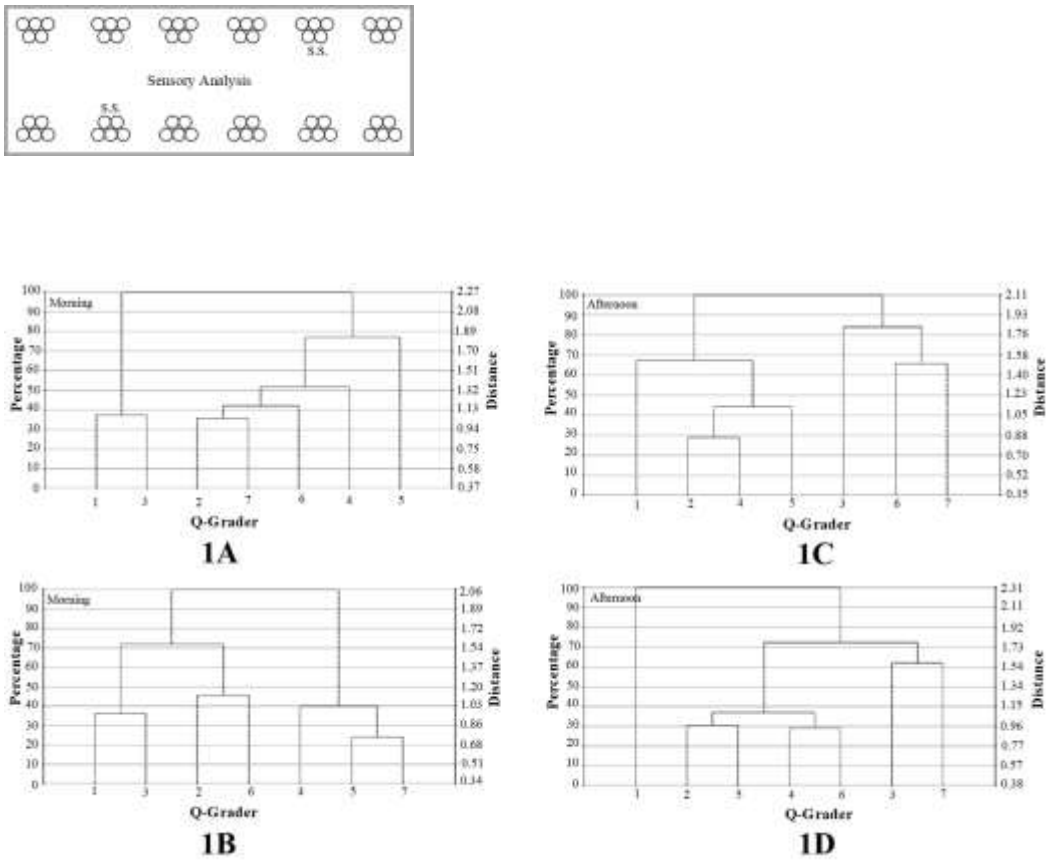
Figure 3: Dendrograms obtained between the Q-Graders from the sensory analysis performed between all the replicates shifts based on the paired T test – morning (2A) and afternoon (2B) relative to experiment 1 on the first day.

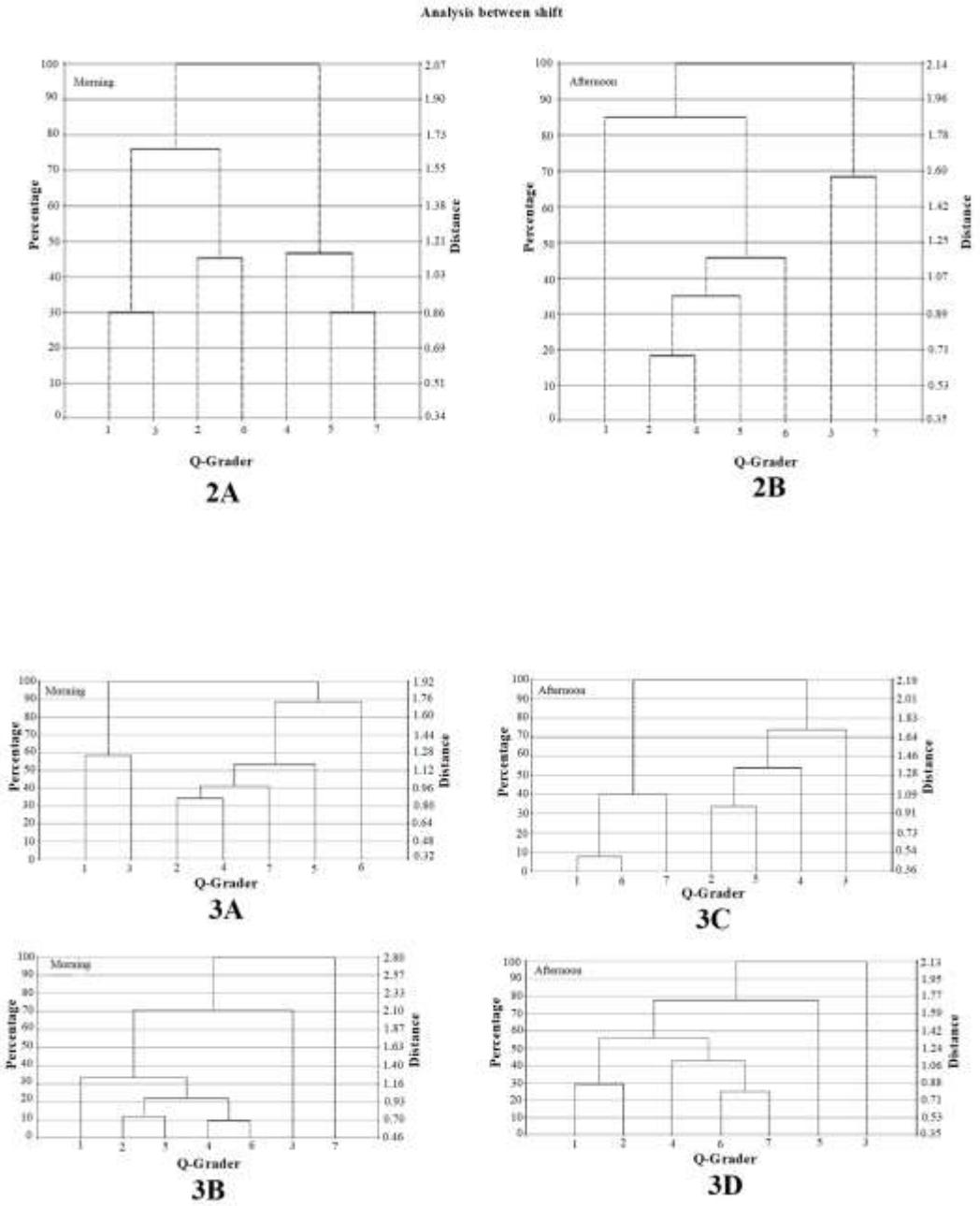
Figure 4: Dendrograms obtained between the Q-Graders, from the sensory analysis performed on the morning shift (left side) for sample 1 (3A) and sample 2 (3B), and the afternoon shift (right side) for sample 1 (3C) and sample 2 (3D) relative to experiment 2 on the second day.

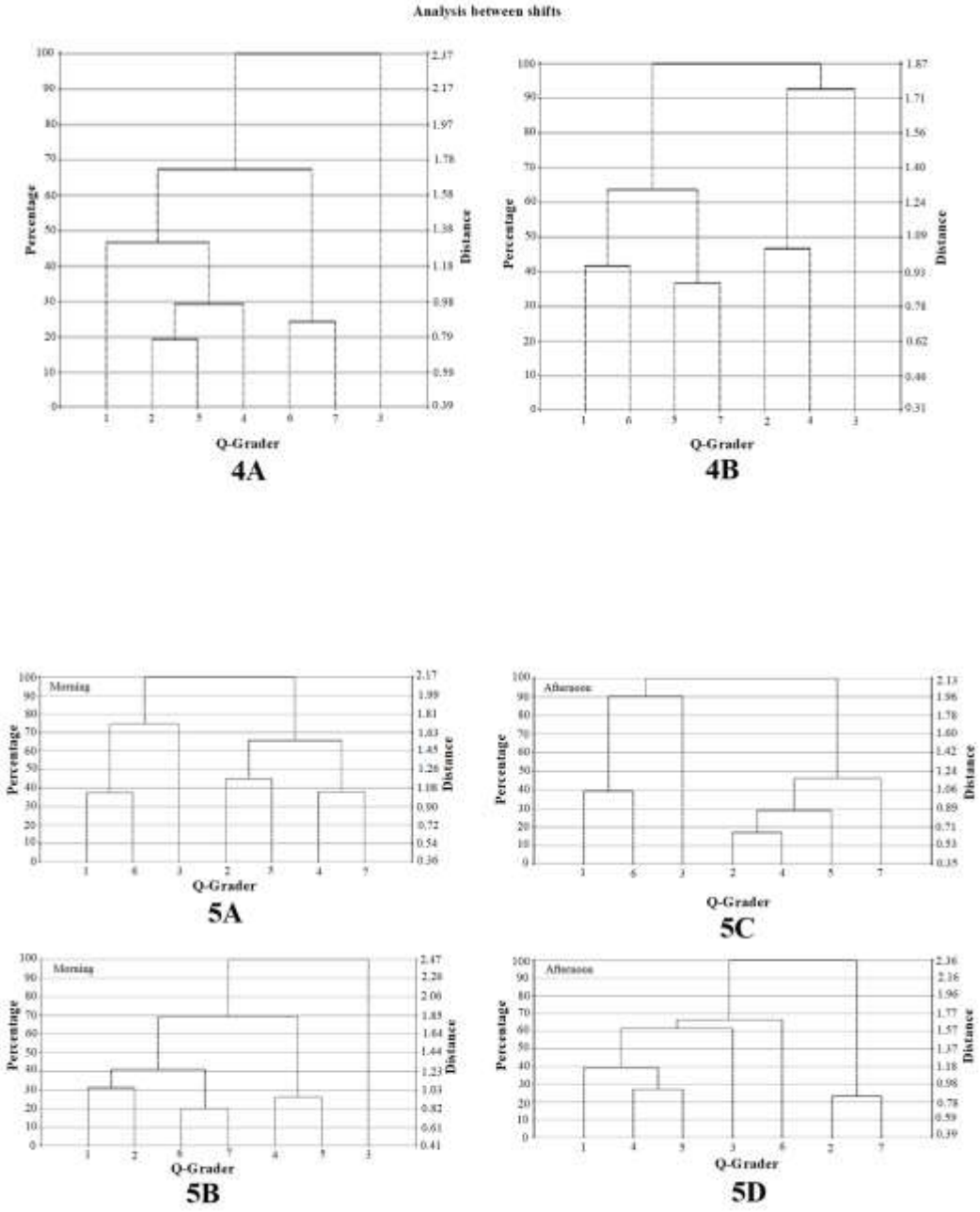
Figure 5: Dendrograms obtained between the Q-Graders from the sensory analysis performed between all the replicates shifts based on the paired T test – morning (4A) and afternoon (4B) relative to experiment 2 on the second day.

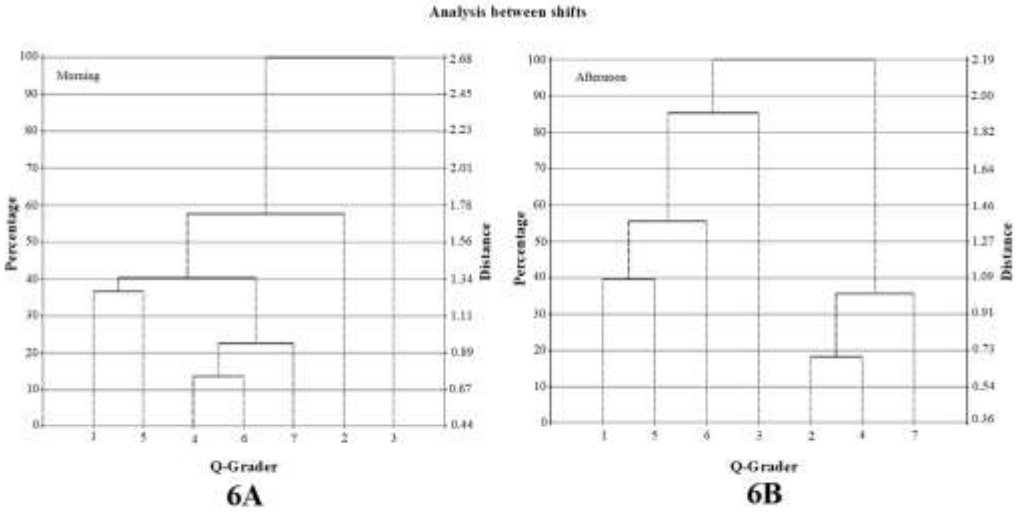
Figure 6: Dendrograms obtained between the Q-Graders, from the sensory analysis performed on the morning shift (left side) for sample 1 (5A) and sample 2 (5B), and the afternoon shift (right side) for sample 1 (5C) and sample 2 (5D) relative to experiment 3 on the third day.

Figure 7: Dendrograms obtained between the Q-Graders from the sensory analysis performed between all the replicates shifts based on the paired T test – morning (6A) and afternoon (6B) relative to experiment 3 on the third day.









6A 6B