

Article

CONTROLLING SENSORY ANALYSIS RESULTS: TOOLS AND IMPORTANCE IN ASSESSOR SELECTION. CASE STUDY: PROVA-ALABE PROFICIENCY TEST

CONTROLO DE RESULTADOS DE ANÁLISE SENSORIAL: FERRAMENTAS E IMPORTÂNCIA NA SELEÇÃO DO PROVADOR. ESTUDO DE CASO: TESTE DE APTIDÃO PROVA-ALABE

Manuel Pinto^{1,*}, Paulo Barros², Alice Vilela³, Elisete Correia⁴

¹ALABE – Associação dos Laboratórios de Enologia, Rua de Ferreira Borges, 27, 4050-253 Porto. Portugal.

²ALABE – Associação dos Laboratórios de Enologia, Rua de Ferreira Borges, 27, 4050-253 Porto. Portugal.

³Chemistry Research Center (CQ-VR), Dep. of Agronomy, School of Agrarian and Veterinary Sciences, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal.

⁴Center for Computational and Stochastic Mathematics (CEMAT), Dep. of Mathematics, University of Trás-os-Montes and Alto Douro, Apt. 1013, 5001-801, Vila Real, Portugal.

* Corresponding author: Tel: +351 935 576 384; email: mmnpinto@gmail.com

(Received 27.11.2023. Accepted 27.03.2024)

SUMMARY

The PROVA-ALABE essay aims to characterize the olfactory, tactile and taste changes underlying the sensory evaluation of wines. It allows assessors to understand the extent to which their perception of a wine is harmonized with that of the other assessors in the trial, contributing to their qualification, as well as the corresponding panel, in detecting and recognizing the most prevalent defects in wines. This work aimed to provide a critical summary of outcomes achieved through interlaboratory tests to identify a primary defect in deliberately contaminated wines. Over three years, interlaboratory tests were performed, corresponding to 60 wines mainly altered with a specific compound. These tests aimed to identify different defects, such as mould, lactic, acescent, reduced, oxidized, vegetable, animal, acidic, bitter, astringency, and others (plastic, sulphur dioxide, bitter almond). However, wine's complexity introduces factors that can affect assessors' perception, with the same defect being perceived differently based on the panel's experience, training, and cultural backgrounds, which can be problematic for an objective analysis. Results of rejection rates and assessor identification percentages suggest that oxidation, mould, and animal defects are more easily recognized. Participation in interlaboratory tests is mandatory for accredited laboratories, serving as an interesting tool for demonstrating results consistency across sensory laboratories. Additionally, such participation can provide valuable information about assessors' performance.

RESUMO

O ensaio PROVA-ALABE tem como objetivo caracterizar as alterações olfativas, táteis e gustativas subjacentes à avaliação sensorial dos vinhos. Permite aos provadores compreender até que ponto a sua perceção de um vinho está harmonizada com a dos outros provadores, contribuindo para a sua qualificação e do respetivo painel na deteção e reconhecimento dos defeitos mais frequentes nos vinhos. O objetivo deste trabalho foi obter um resumo crítico dos resultados alcançados através de testes interlaboratoriais destinados a identificar um defeito primário em vinhos intencionalmente contaminados. Foram realizados testes interlaboratoriais ao longo de três anos, correspondendo a 60 vinhos maioritariamente alterados com um composto específico. Esses testes tiveram como objetivo identificar diferentes defeitos, como mofo, láctico, acescente, reduzido, oxidado, vegetal, animal, ácido, amargo, adstringência e outros (plástico, dióxido de enxofre, amêndoa amarga). Contudo, a complexidade do vinho introduz fatores que podem afetar a perceção dos provadores, podendo o mesmo defeito ser percebido de forma diferente em função da experiência, treino e formação cultural do painel, o que pode ser problemático para uma análise objetiva. Os resultados das taxas de rejeição e das percentagens de identificação dos provadores sugerem que a oxidação, o mofo e animal são os defeitos mais facilmente reconhecidos. A participação em ensaios interlaboratoriais é obrigatória para os laboratórios acreditados, constituindo uma ferramenta interessante para demonstrar consistência de resultados entre laboratórios de análise sensorial. Além disso, a participação em testes interlaboratoriais pode fornecer informações importantes sobre o desempenho dos provadores.

Keywords: Sensory analysis, wine defects, wine quality control, interlaboratory test, taste and olfactory alterations.

Palavras-chave: Análise sensorial, defeitos do vinho, controlo da qualidade do vinho, teste interlaboratorial, alterações olfativas e gustativas.

INTRODUCTION

Proficiency testing (PT) programs have received extensive attention and are regularly employed in chemical and instrumental analyses as official inter-comparison studies. However, PT programs have not yet achieved the same level of development in sensory evaluation, primarily because sensory analysis differs significantly from chemical or other instrumental analyses. Nevertheless, engaging in a PT program remains the most efficient means to showcase the capabilities of a sensory panel. Moreover, for accredited laboratories, it is often a mandatory requirement, as referred to by some authors (McEwan *et al.*, 2003; Kilcast, 2010; Pinto and Barros, 2015).

Engaging in PT programs equips laboratories with an impartial instrument to gauge and validate the accuracy of their data, thereby showcasing the laboratory's competence in conducting specific tests. These programs are firmly rooted in the realm of analytical (chemical) analysis, and international standards and guidelines are in place to govern the planning, execution, and interpretation of interlaboratory studies (Thompson and Wood, 1993; Hyldig, 2010; ISO, 2023).

Consistency of results obtained from different sensory panels is frequently addressed, particularly within the industry. With accreditation, the need to demonstrate the reliability of sensory panel results has been brought to the forefront of discussions (McEwan, 1999; Gomis-Bellmunt *et al.*, 2024).

In sensory panels, excellence in PT is crucial to demonstrate the consistency and reliability of results. Although several studies have focused on assessing individual sensory assessor performance (Naæs, 1998; Rossi, 2001), the collective performance of the sensory panel has received less attention. However, the overall panel performance is fundamental in determining the sensory quality of a product. Subpar performance by individual assessors can lead to a decline in panel performance (McEwan *et al.*, 2003). Sensory evaluation, unique for using human perception as the measuring instrument, deals with subjective perceptions that may not align easily with chemical or physical measurements. Unlike chemical analysis, which quantifies the concentrations of substances, sensory analysis assesses responses to stimuli, and results are inherently relative. PT in sensory evaluation faces challenges in establishing criteria for panel performance evaluation. Previous studies (McEwan, 2000a, 2000b; McEwan, 2001b, 2001a) have made progress in developing these criteria, introducing the concept of the 'expected value' instead of the 'true value' used in chemical analysis. The 'expected value' is established through

validation panels before complete inter-comparison trials, as outlined by McEwan *et al.* (2003).

Adapting the process of forming the validation panel is fundamental in order to meet the requirements and goals of the sensory analysis in question. It is important to note that validation panels do not guarantee an absolute true value, as sensory perception can vary between individuals. However, setting up a validation panel provides a systematic and reliable approach to determining the average value and fluctuations in a product's sensory characteristics.

Larger companies managing sensory panels across multiple locations face challenges in achieving comprehensive quality control and product development. Calibrating sensory panels ensures consistency across locations, minimizing discrepancies and facilitating result comparisons. When various panels assess the same samples, global performance issues can amplify local concerns, making result comparisons more complex. Existing proficiency testing techniques developed for chemical comparisons (Thompson *et al.*, 2006) may not adequately address specialized aspects of sensory analysis (Tomic *et al.*, 2010).

Wines and spirits, in addition to their sensory attributes, sometimes have other attributes that are considered harmful and can be classified into visual, olfactory, gustatory, and tactile alterations. However, it should be stressed that not all unpleasant aromas or flavors in a wine are considered defects – some may reflect the wine's variety or style (Miranda *et al.*, 2017). Wine tasting is a subjective process, and "tastes" can vary from person to person (Goode, 2018).

Detecting defects in wines requires skilled assessors with a broad repertoire of reference aromas. Taste defects, often indicative of contamination, are considered severe. Preventing negative olfactory occurrences involves various organoleptic methods, such as sensory analysis, quality control, storage care, raw material evaluation, and production line monitoring (Saxby, 1996).

ALABE (Associação dos Laboratórios de Enologia/Oenology Laboratories Association) established the PROVA-ALABE to harmonize sensory evaluation criteria among assessors for wines, whether defective or not. The test is a tool for assessor qualification, training, and ongoing improvement, ensuring criteria alignment and practical understanding of their performance (ALABE, 2024). The test is conducted under strict confidentiality, involves manipulated wines with common defects, and provides a diverse view of the assessor's perception. Since 2020, the PROVA-

ALABE has evaluated 60 wines altered with specific compounds, testing perceptions of defects such as mould, lactic, acescent, reduced, oxidized, vegetable, animal, acidic, bitter, astringent, and others (plastic, sulphur dioxide, bitter almond).

A PT that focuses on these issues of what exactly constitutes a defect in a wine (and in the absence of a prior validation panel) will be based on the fact that the final result will always be the outcome of a global consensus on the acceptable limits of the various compounds associated with defects present in the wine.

Since 2020, ALABE has used the test format outlined in this study, covering the period from 2020 to 2022. The goal is to characterize olfactory, tactile, and flavor changes in wine sensory evaluation, aiding assessors in understanding the alignment of their perceptions with other participants. The test also contributes to the assessor and panel qualification in detecting and recognizing common wine defects.

The aim of this work was, therefore, to report on the experience accumulated in organizing, guiding, developing, operating, selecting, and using proficiency tests in the field of sensory analysis, with particular emphasis on the performance evaluation that is carried out, which is fundamental to guarantee the accuracy and reliability of the results. In addition to address the state of the art in the organization of proficiency tests and their importance for bodies dedicated to product certification, the purpose of this work was to critically analyze the results of the participation of assessors from different sensory analysis laboratories in the proficiency test PROVA-ALABE organized by ALABE. It aims to qualitatively and quantitatively assess the response given by assessors to olfactory, tactile, and flavor stimuli responsible for wine faults commonly present in wines, providing an instrument for maintaining, perfecting, or training the sensory potential aiding assessors in understanding the alignment of their perceptions with other participants. The objectives of the PROVA-ALABE as a proficiency test are related to the importance of wine faults familiarization and respective decision-making, allowing to improve the ability to get to know products, improving sensory capacity, and obtaining accurate judgments through "standardization," making it possible to generally classify assessors in terms of their ability to perceive and discriminate faults related to their olfactory and gustatory memory.

MATERIALS AND METHODS

Planning of distribution and preparation of wine samples

According to the previously established schedule, ALABE distributed two wine samples to PROVA-ALABE participants in each edition. There were 25

editions between 2020 and 2022, corresponding to 50 wine samples. Each sample of wine, packaged in 0.75 L bottles, was coded as follows: circuit/letter (A or B)/month/year (e.g., PROVA-ALABE - A/January/2023).

All the participating organisations received the same wine samples for sensory evaluation. The two samples that ALABE distributes may be intentionally carriers of sensory alterations, corresponding to naturally altered or even purposely altered wines, considering the formative character of this test. The samples that ALABE distributed in each edition may be purposely altered, or only one or neither is deliberately changed. Thus, all combinations can occur randomly.

To facilitate the tasting procedures and also the reading of this work, the IUPAC nomenclature of compounds, in parentheses, will be referred by each corresponding common compound name: Isovaleric acid (3-methylbutanoic acid); Acetaldehyde (Acetaldehyde); 2-Mercaptoethanol (2-Hydroxyethanethiol); 4-Ethylphenol (4-Ethylphenol); Sulfur dioxide (Sulphur dioxide); 2,4,6-Trichloroanisole (1,3,5-Trichloro-2-methoxybenzene); Diacetyl (Butanedione); Geraniol (3,7-Dimethylocta-2,6-dien-1-ol); Quinine sulphate (Quinine sulphate); Grape tannin (Grape tannin); Tartaric acid (2,3-Dihydroxybutanedioic acid); Styrene (Styrene or Ethenylbenzene); Benzaldehyde (Benzenecarbaldehyde); IBMP (2-Isobutyl-3-methoxypyrazine); Acetic acid (Ethanoic acid); dimethyldisulfide (DMDS) (Dimethyldisulphide); Ethyl acetate (Ethyl acetate).

Altered samples shall reflect the changes induced by reference analytes (e.g., acetic acid, acetaldehyde, 2,4,6 – trichloroanisole) with varying concentration values and based on reference literature concerning detection and recognition limits (Baigrie, 2003; Goode, 2018; Grainger, 2021e; Australian Wine Research Institute, 2023; Nakano, 2023). They must be suitably packaged and come from the same batch.

Sample preparation and shipping

As mentioned above, one of the most critical aspects of implementing proficiency tests concerns the production of stable and homogeneous samples. For this trial, all samples of wine that were intentionally intended to cause a fault were distributed (as well as those that were not intended to mean fault). Additionally, the concentrations of the analytes responsible for the odor, taste, and tactile defects were above the known detection limits for alcoholic matrixes/wines. The list of used chemical compounds along with their purities are 2,4,6–trichloroanisole (99% purity), 2-mercaptoethanol (99% purity), 4-ethylphenol (99% purity), acetaldehyde (99% purity), acetic acid (99.88% purity), benzaldehyde (99% purity), diacetyl (99%

purity), dimethyldisulfide (99% purity), ethyl acetate (99.9% purity), geraniol (97% purity), grape tannin (purity information not provided), 2-isobutyl-3-methoxypyrazine (99% purity), isovaleric acid (99% purity), quinine sulphate (99.8% purity), styrene (99% purity), sulphur dioxide (purity information not provided), tartaric acid (99.8% purity).

The quantity of the compound to be added was optimized by laboratory-scale testing for the wine selected for the test (it is usually necessary to prepare intermediate-concentration solutions to minimize the risk associated with measuring tiny amounts). Once defined, the essential quantity of the required compound (for the volume of wine to be prepared) was added to the bottle of 0.75 L, ensuring a sufficient contact time (wine + fault) between 1 and 3 days. Bottling can be done almost immediately in the case of very volatile compounds. The winery has no temperature control (in the laboratory, the temperature is close to 20 °C). Bottling was manual, carried out in glass bottles using a natural cork stopper.

Ideally, verifying the homogeneity and stability of the batch(es) should be performed according to ANNEX B of the ISO 13528 standard (ISO, 2022). For reasons related to the speed of the test and the absence of laboratories capable of analyzing some of the compounds used, this step was not carried out, seeking to ensure the correct homogeneity in the stainless-steel tank used for the contamination of the wine.

If possible, the test materials to be distributed should be generally similar in the matrix to the routinely analyzed samples (concerning the matrix composition and concentration range of the analyte). In this regard, EA-4/09 G:2022 (EA, 2022) states: "Where available, laboratories shall participate in proficiency tests which are relevant to their scope of accreditation; preference should be given to proficiency testing schemes that use appropriate matrices." However, as Wood *et al.* (1998) highlight, this is not always possible concerning stability and homogeneity or other reasons when working with microorganisms and some unstable chemicals. They must be of acceptable uniformity and strength. The prepared bulk material must be effectively homogeneous so that all laboratories receive similar samples that do not show significant differences in analyte concentration. The coordinator shall indicate the procedure to determine the homogeneity of the test material. However, as a guide, the standard deviation between samples should be less than 0.3 times the target value for the standard deviation (Wood *et al.*, 1998; Thompson *et al.*, 2006). Wood *et al.* (1998) state that, when possible, the coordinating laboratory should also provide evidence that the test

material is sufficiently stable to ensure that the material will not change significantly throughout the proficiency test. Before distributing samples, the matrix and analyte stability should be assessed under storage conditions mirroring test duration. Stability tests should consider sample transport and laboratory conditions with unchanged analyte concentrations. For unstable analytes, the coordinating organization may set a deadline for analysis. The same authors (Wood *et al.*, 1998) also suggest that quality controls for samples ideally come from a laboratory other than the one preparing them. However, it may pose difficulties for the coordinating body. The number of distributed test materials depends on covering various compositions, with a practical limit of six per analyte in the food sector. Coordinators should consider potential hazards of test materials and advise parties at risk, such as distributors and test laboratories. The correct procedure involves following the mentioned steps to ensure that proficiency testing accurately reflects the capabilities and performance of analytical laboratories. This includes selecting appropriate test materials, ensuring their homogeneity and stability, implementing quality control measures, and considering safety considerations.

Once the samples have been prepared, they were sent to the laboratories registered in the circuit, always ensuring anonymity, without indicating the molecules used to prepare the samples.

Tasting panel

The number of participating entities varied between 22 and 25, representing 162 and 176 assessors annually.

The entities that take part in this test were directly or indirectly linked to the wine sector, and their assessors were members of trained panels that form part of accredited or non-accredited tasting panels.

Assay methodology

In the first phase of the test, the PROVA-ALABE test enabled assessors to evaluate whether a given wine is acceptable by determining its organoleptic characteristics for non-defective status. If considered defective, the procedure outlined a sensory analysis to identify the specific defects in the wine (second phase of the test).

The test methodology followed an adopted protocol, which is explained below.

Calendar

PROVA-ALABE had ten editions per year, each involving the availability of two wines in 0.75 L bottles, for each entity.

An annual calendar was provided, indicating the sample distribution dates, the deadlines for sending results, and the expected dates for completion of each report.

The distribution of samples took place at the beginning of the month, and the submission of results by the participants to PROVA-ALABE took place during that month. The PROVA-ALABE report was preferably released in the first week after the month of the trial. Given the nature of the product, participants had to organize themselves to carry out the sensory analysis sessions as soon as possible after receiving the samples.

Trial registrations

Only assessors/groups of assessors duly registered in PROVA-ALABE were able to participate in the test. The participation regime in the trial required prior registration and included the distribution of two bottles of wine per edition to ALABE members,

although non-associated entities could also participate. With registration, participants were required to perform the trials according to the specific procedure established for the trial and report the results achieved in the specified periods.

Methodology

Affective tests were used to assess product preference or acceptance. The acceptance/rejection method determines whether a wine can be considered defective because of its organoleptic characteristics. Therefore, in the first step, the test consisted of the expression of acceptability by the sample tested. In the second phase, in case of rejection of the sample, the assessor was asked to comment on the reason for the rejection of the sample concerning the characterization of olfactory, tactile, or taste changes according to the information presented in the evaluation form, available in Figure 1.

PROVA-ALABE

TASTING SHEET CHARACTERIZATION OF OLFACTORY, TACTILE OR TASTE ALTERATIONS						
Date:		Sample identification:				
Taster name:		Taster code:				
Cabin n°:		Glass n°				
TEST - ACCEPTANCE / REJECTION						
Do you accept this wine? Yes <input type="checkbox"/> No <input type="checkbox"/>						
If NO, please complete the table below. You should only mark one change.						
TEST - CHARACTERIZATION OF OLFACTORY, TACTILE OR TASTE ALTERATIONS						
ODOR	Vegetal	1	2	3	4	5
	Mould	1	2	3	4	5
	Acescent	1	2	3	4	5
	Reduced	1	2	3	4	5
	Oxidation	1	2	3	4	5
	Animal	1	2	3	4	5
	Lactic	1	2	3	4	5
TASTE, TACTILE	Acid	1	2	3	4	5
	Bitter	1	2	3	4	5
	Astringent	1	2	3	4	5
Other changes *		1	2	3	4	5
Comments:						
1: hardly noticeable; 2: light; 3: median; 4: strong; 5: extreme						

NOTE 1: If the "other changes" option on the test sheet is selected, it must be clearly defined in the "Characterize change" field by selecting one of the options available in the list provided. ALABE may not consider results that use ambiguous terms and/or are not listed in the "Characterize change" field.

NOTE 2: If the sample is rejected, only one sensory change ("defect") can be recorded, which must correspond to the one the taster considers most relevant. Thus, any record that shows more than one sensory change will be completely eliminated, i.e. it will not be taken into account for the purposes of processing the results.

Figure 1. PROVA-ALABE Test Sheet.

In case of sample rejection, the assessors also had to rate the intensity of the perceived defect on a 5-point Likert scale. Generally, a large number of responses is required for each assessment. Therefore, it was understood that the minimum number of assessors should be higher than 30, based on the fact that the data is statistically expected to follow approximately a normal distribution when n is high ($n > 30$) (Bower, 2013). For the test to be considered valid, ALABE required that a minimum of 30 assessors were considered in the statistical treatment and did not publish any report when the number was lower.

Affective tests refer to a measure of opinion, that is, preference or acceptability, stating how much the assessor/consumer likes or dislikes a product.

Category scales are widely used in sensory analysis for affective responses (related to product preference or acceptance). For example, these are scales where the degree of acceptance is assessed through verbal (nine-point hedonic scale) or numerical (nine-point numerical hedonic scale) labels (Pimentel *et al.*, 2016).

On the implemented tasting sheet, the question that assessors were asked to answer to determine whether a wine is accepted or rejected is: Do you accept this wine? yes no

The acceptance/rejection method was completed by answering the question; then, they answered other questions formulated by executing other methods.

When a defect is detected by at least 50 percent of the panel of assessors, it is common practice in the wine sector to classify the wine as defective. The practice is also similar in the olive oil sector (International Olive Council, 2018). Considering that (i) the population of assessors participating in a given PROVA-ALABE test varied between 100 and 120 assessors, (ii) there is no single universal guideline for the percentage of assessors who should attribute a defect to a sample to consider it defective, (iii) the size of the population is crucial when interpreting statistical results, ALABE took the decision, in line with what underlies the use of a binomial criterion in its decision-making, to consider that if at least 60% of the assessors rejected the sample, ALABE believes that there was a defect in the wine. Its characterization was focused on the most frequent defect. If the sample was rejected and at least 60 % of faults were still not achieved, it was considered that there was insufficient evidence to support the conclusion that the sample was defective.

If the sample was manipulated to cause a defect and the sample was rejected, the defect most frequently

reported was expected to be initially intended ("true value"). If this was not the case, for various reasons (interactions, among others), each entity should pay attention to the initial intention and not limit itself only to the most frequent defect.

Selection of descriptors

The tasting sheet lists typical wine defects (Figure 1). The descriptors of each defect, the corresponding molecule(s) that simulate the defect, and its (their) concentration levels required for detection or recognition thresholds in wine were researched in various literature (Goode, 2018; Australian Wine Research Institute, 2023).

Sample conservation and preparation

To ensure no external effect may alter the matrix, it is essential to store samples properly. Samples should be kept away from light and stored in a cool place at a constant temperature. It was recommended to refrigerate the samples, if possible, at a temperature of 6°C to 8°C and store them in a vibration-free place. The samples should be removed from the refrigerator at least 24 hours before the test and thermally balanced to achieve accurate results.

Execution of the test

Entities involved in the circuit were requested to perform the test according to the instructions provided in the test procedure: Use tasting glasses that comply with ISO 3591:1977 (ISO, 1977); follow the test procedure guidelines to perform the test; the panel leader should ensure that their team of assessors is familiar with the methodology described in the test procedure and that everyone has all the necessary information for the correct execution of the test; prepare a sufficient number of glasses that match the number of assessors who will participate in the test; pour the same volume of wine equally into each glass (approximately 40 mL per glass); place the tasting glasses in the tasting booths and ensure that the temperature of the wine is $20\text{ }^{\circ}\text{C} \pm 3\text{ }^{\circ}\text{C}$ (except for sparkling wines, which should be $10 \pm 3\text{ }^{\circ}\text{C}$) at the time of tasting.

Communication/submission of results

To communicate results to ALABE, each entity had a "participation code" (password) sent promptly. ALABE analyzed the individual and/or group and sent the report to the assessor or group coordinator.

When entering data online, the assessor(s) was coded with an 8-character password of letters and numbers (e.g., A345Y97F). The first character was a letter and was kept in each edition. This was the code with

which the assessor was recognize himself in the ALABE report.

Each entity entered the data directly on ALABE's website. To this end, ALABE provided a "username" and "participation code" (password). The codes and results submitted by the participants were kept confidential by ALABE.

Statistical analysis and treatment

Acceptance/Rejection assay

Distribution of the sample proportion

The distribution of all possible values that some statistic can assume, computed from samples of the same size randomly drawn from the same population, is called the sampling distribution of that statistic (Daniel and Cross, 2018). Sampling distributions allow answering probability questions about sample statistical measures and providing the theory needed to make statistical inferences.

Suppose (X_1, X_2, \dots, X_n) is a random sample of dimension n of a Bernoulli random variable with unknown probability success p ($X_i \sim B(p)$), here X_i takes the value 1 if it is a success and 0 if it is a failure. When the intention is to estimate the proportion p of elements possessing a certain characteristic in the sample (proportion of successes), the estimate $\hat{p} = \bar{X} = X/n$ should be used, where X represents the number of elements in the sample that possess the characteristic. Therefore, the distribution of X will be binomial, $X \sim B(n, p)$ and its probability distribution is given by the Equation 1.

$$P(X = k) = \binom{n}{x} p^x (1-p)^{n-k} \quad \text{Eq. 1}$$

$$x = 0, 1, \dots, n.$$

When the sample size is large enough ($n \geq 30$), the central limit theorem (CLT) allows us to establish the Equation 2 and 3.

$$Z = \frac{X - np}{\sqrt{np(1 - np)}} \sim N(0, 1) \quad \text{Eq. 2}$$

or

$$Z = \frac{\bar{X} - p}{\sqrt{\frac{\bar{X}(1 - \bar{X})}{n}}} \sim N(0, 1) \quad \text{Eq. 3}$$

because $\hat{p} = \bar{X} \sim N\left(p, \frac{p(1-p)}{n}\right)$

Thus $\hat{p} = \bar{X}$ is a point estimator for the proportion (Murteira *et al.*, 2007).

Confidence interval (CI) for the proportion

Point estimation allows us to obtain estimates for the unknown parameter p ; however, relying on a single estimate is insufficient to establish that value's accuracy. Sometimes, it is preferable to determine a

confidence interval for the parameter. This is particularly useful if the sample size is large enough ($n \geq 30$) for the Central Limit Theorem (CLT) to apply (Equation 4).

$$Z = \frac{\bar{X} - p}{\sqrt{\frac{\bar{X}(1 - \bar{X})}{n}}} \sim N(0, 1) \quad \text{Eq. 4}$$

So, an interval estimate for p may be expressed using the Equation 5.

$$\left[\bar{X} - b \sqrt{\frac{\bar{X}(1 - \bar{X})}{n}}, \bar{X} + b \sqrt{\frac{\bar{X}(1 - \bar{X})}{n}} \right] \quad \text{Eq. 5}$$

where b is the value such that $P(Z \leq b) = 1 - \alpha/2$, with $Z \sim N(0, 1)$ (Murteira *et al.*, 2007).

It is important to note that interval estimation does not have a specific value for the unknown parameter. Instead, there is a range of values to which the parameter, p , can belong with a certain degree of confidence. However, asserting that the parameter belongs to the confidence interval (CI) with a probability of $(1 - \alpha) \%$ is inaccurate due to the unknown actual value of p . The parameter either falls within the provided interval or does not. Therefore, one can only state that in repeated sampling, approximately $100(1 - \alpha)$ percent of these intervals will include the population proportion p .

Example: If one wishes to find the proportion p of individuals who answered "yes" to the question "Do you accept this wine?". To this purpose, a survey was carried out with 100 individuals regarding their preference (or not) for wine, allowing possible answers:

Yes (1) I accept the wine | No (0) I do not accept the wine.

Of the 100 respondents, 40 answered "yes" and 60 "no". Assuming that the favorable event is to "answer yes," then a point estimate for the proportion would be

$$\hat{p} = \frac{40}{100} = 0.4$$

Similarly, in seeking a 95% confidence interval for the proportion, p , in example above, the following is applicable: $\alpha = 0.05$, $P(Z \leq b) = 0.025$, so the b value corresponding to a confidence coefficient of 95% will be 1.96.

Thus, the 95 percent confidence interval for the population proportion p , based on this data, is

$$\left[0.4 - 1.96 \sqrt{\frac{0.4(1 - 0.4)}{100}}, 0.4 + 1.96 \sqrt{\frac{0.4(1 - 0.4)}{100}} \right]$$

$$]0.395, 0.405[$$

Based on the results obtained, with 95 percent confidence, it would be expected that between 39.5% and 40.5% of individuals would answer "yes" to the question "Would you accept this wine?".

Binomial criterion

In the absence of a proportion of interest (yes/no), the criterion used for determining the maximum number of "not accepted" responses that a wine can receive without being rejected or the maximum number of "accepted" responses that a wine can receive to be rejected is based on a specific threshold.

Only the results of the assessors who correctly identified whether the test wine (Sample A and Sample B) should be accepted or rejected will be considered for calculating the test results.

When determining whether a test is accepted or rejected, the results follow a binomial distribution with only two possible outcomes - "Accepted" or "Rejected" - each with a probability of 0.5. The calculation involves determining the smallest value required for the cumulative binomial distribution to equal or exceed a specified criterion. The Microsoft Excel INV.BINOM() function helps determine the maximum number of successful test results needed to meet a specific acceptance criterion for the binomial distribution that is represented by the Equation 6.

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k} \quad \text{Eq. 6}$$

where:

- $P(X = k)$ is the probability that exactly k successes will occur on n independent trials.
- $\binom{n}{k}$ is the binomial coefficient, which represents the number of different ways to pick k successes in n trials. It is calculated as $\frac{n!}{k!(n-k)!}$ where $n!$ represents the factorial of n
- p is the probability of success (of making a correct decision) in a single attempt.
- $(1-p)$ is the probability of failure (of making an incorrect decision) in a single attempt.

Identification of sensory alteration

In the first phase of the evaluation process, the assessors were asked to assess the acceptability of the wine sample, deciding whether the sample was fit for approval. In the second phase, if the sample was rejected, the assessors had to identify the sensory alteration (also known as "defect") and rate its intensity on a Likert scale from 1 to 5. They could only record one sensory alteration they consider the most relevant. Any record with more than one sensory alteration was eliminated and not considered for the treatment of results.

Using high-performing taste assessors in various contexts can lead to a better understanding of

consumer preferences, improved product quality, differentiation in the market, and fact-based decision-making (Beeren, 2018). The PROVA-ALABE test's typology highlights the importance of using top-performing assessors in sensory tests. The overall performance indicator for entities and assessors, respectively, based on acceptance/rejection tests, was calculated using a specific formula (Equation 7).

$$\frac{n \text{ successes} * n \text{ total valid (successes + failures)}}{\text{weighted average}} \quad \text{Eq. 7}$$

The formula adopted corresponds to a mechanism developed internally by the ALABE team to find a methodology that makes it possible to establish a ranking in terms of the performance of organizations and assessors (average of specific % hits), but at the same time considering the number of registrations (participations), and dividing by the weighted average, which consists of multiplying each value (average of specific % hits) by its weight (participations). Finally, all the values are added together, and the result is divided by the sum of the weights (participations).

The weighted average is likely calculated by multiplying each entity's specific % hits by its corresponding number of registrations, summing up these products, and then dividing by the total number of registrations. This accounts for the fact that entities with more registrations significantly impact the overall average because their larger representation in the dataset amplifies their impact on the calculation. When calculating a weighted average, each entity's contribution is proportional to both its percentage of hits and the number of registrations it represents. Therefore, entities with more registrations contribute more significantly to the final result due to their larger weight in the calculation.

RESULTS AND DISCUSSION

Quality Assurance (QA) and Quality Control (QC) test results in sensory analysis should be promptly communicated for effective decision-making. Clear data presentation is crucial for user interpretation. Sensory tests must deliver fast, actionable results compatible with standard QC formats, typically numerical data. Exceptions, like pass/fail data from PROVA-ALABE, may employ word-based results. Complementary to the primary results, brief descriptions or pre-developed lists can specify defects or aid decision-making (Stapleton, 2021). The initial acceptance/rejection decision in the test's first phase relies on a binomial criterion, a statistical

tool modeling the probability of success or failure in independent trials with two possible outcomes (Næs *et al.*, 2011).

ALABE published a report after each edition of the assay. The report contained the individual (assessors) and group (entity) results. In the first phase, the test consisted of the expression of acceptability for the sample tested, whose statistical analysis of the results

obtained for assessors and panel (entities) are shown in Table I and Table II, respectively.

The analysis of each entity concerning the identification of defects is presented in Table III. The identification of the most frequently observed defect (mode) is shown in Table IV; the frequency and respective percentage of defects reported in a given test is presented.

Table I

Example of descriptive statistics of proportions for a given sample of the PROVA-ALABE trial

Proportion:	
Count (yes)	26
Count (no)	85
Count (total)	111
Proportion (p=yes)	0.234
Proportion (p=no)	0.766
Average ratio to "yes"	0.234
Variance (p-var)	0.179
Standard error (p-se)	0.04
CI 95%	0.08
Binomial criterion	64

Table II

Example of the most frequent value (mode) for acceptance/rejection test Panel – PROVA-ALABE test

Entity code	Acceptance Testing Panel ¹	Number of assessors in the panel
1	2	11
2	1	4
3	1	2
4	1	8
5	1	9
6	1	10
7	1	9
8	1	9
9	1	14
10	1	11
11	1	4
12	1	6
13	2	9
14	1	5

1 (1 =" No"; 2 =" Yes").

Table III

Example of the most frequently observed value (mode) for testing defect identification Panel – PROVA-ALABE test

Entity code	Defect Identification Test ²	Number of assessors in the panel
1	0	11
2	0	4
3	1	2
4	0	8
5	0	9
6	1	10
7	0	9
8	1	9
9	1	14
10	0	11
11	0	4
12	1	6
13	0	9
14	0	5

2 (1=detected; 0=not detected).

The data collected from each tasting panel was processed statistically. For each wine and each defect (if applicable), the percentage of the intensity level of the perceived defect was calculated. An example is shown in Table V.

An overview of the test results corresponding to adulterated samples (that is, samples with "defect") is summarized in Table VI.

The results were generally satisfactory, with the deliberately altered samples rejected by most of the

assessors, and the defect most frequently pointed out by most assessors was found in the alignment of the initial intention desired. In most cases, the assessors found the defect associated with the presence of the contaminating molecule in almost all the proposed samples. Even if wines with a neutral sensory profile were chosen for the tests, adding some molecules can alter the sensory balance and highlight defects that were not the intended target.

Table IV

Example of a frequency table and respective percentage of defects reported by the assessors for the PROVA-ALABE test

Defect	Frequency	Percentage
acescent	38	34.2
animal	1	0.9
lactic	4	3.6
mould	4	3.6
oxidation	11	9.9
reduced	16	14.4
vegetal	3	2.7
others_ Sulphur dioxide	0	0.0
others_ Benzaldehyde ("Bitter almond. ")	1	0.9
others_ Peracetic acid	1	0.9
acid	3	2.7
bitter	3	2.7
astringent	0	0.0
No defects	26	23.4
Total	111	100.0

Table V

Example of a table of counts and percentage frequencies per wine for the scale's different intensity levels and each defect

Defect	Defect intensity					Total
	1 n (%)	2 n (%)	3 n (%)	4 n (%)	5 n (%)	
acescent			21 (55.3)	11 (29)	2 (5.3)	38
animal				1 (100)		1
lactic		1 (25)	3 (75.0)			4
mould	1 (25)	1 (25)	1 (25.0)		1 (25)	4
oxidation		3 (27)	3 (27.3)	4 (36)	1 (9.1)	11
reduced		6 (38)	5 (31.3)	5 (31)		16
vegetal			1 (33.3)	2 (67)		3
others_ Benzaldehyde ("Bitter almond")			1 (100.0)			1
others_ Peracetic acid					1 (100)	1
acid			2 (66.7)	1 (33)		3

Analyzing Table VI, it is evident that 23 out of 25 samples were rejected by the assessors, indicating the presence of defects in the samples. One of the tests (sample) was considered statistically inconclusive, and another test (sample) was accepted by the assessors where it would have been expected to be rejected. Regarding the accuracy of identifying defects, the majority were identified and determined by the assessors. Even in cases in which the percentage of assessors who identified the proposed defect was low, it corresponded to the most frequent value reported. This suggests that, overall, the assessors were consistent in identifying defects,

aligning with the initial intention of the test organizer. However, in the case of four samples (17%), there was a deviation from this pattern, indicating a discrepancy between the assessors' identifications and the intended outcomes of the test organizer.

From a global perspective, considering the entities and their performance on the acceptance/rejection test, regardless of the direction taken, whether acceptance or rejection, the percentage of correct answers was 77% (2020), 89% (2021), and 84% (2022). Regarding defect identification, the rate of

correct answers was 65% (2020), 57% (2021) and 66% (2022).

It was impossible to determine the effective concentration of several compounds used in the tests carried out due to various contingencies. Analyzing the amount of compound added to a wine and the concentration retained helps understand how different factors affect the final concentration. This insight is valuable for optimizing processes and making informed decisions and allows researchers to compare results across experiments, enhancing the reliability and generalizability of findings.

Examples of a radar graph representing the intensity of each defect perceived in a specific wine is displayed in Figure 2 and Figure 3. In these graphs, the filled polygon represents the average intensity of the defects perceived by the assessors.

An overview of the results obtained in the proficiency tests to identify a significant defect in artificially contaminated wines is presented below (Table VI).

Two samples were prepared for the detection of **mould** alteration in white and red wine, respectively: sample B January 2021; sample A, June 2022. Both wines were contaminated with 2,4,6-trichloroanisole (TCA). The amounts of TCA used to contaminate the wines were 100 ng/L for white wine and 250 ng/L for red wine. The moldy olfactory alteration was indicated by 89.5% of the assessors for the first test with white wine, and 93.1% for the second test with red wine. Both tests had high rejection rates (94.7% for white wine and 98% for red wine). The mean intensity of the perceived change of the selected assessors was 4.3 and 3.9, respectively. The results of both tests were satisfactory; a larger population of assessors detected the proposed defect, and the average perceived intensity found was higher on the first test with white wine, although the amount of compound added was lower than that added to red wine. The olfactory identification of trichloroanisole can be difficult due to the influence of the matrix (the environment in which the compound is present), the difference in panel sensitization (how sensitive individuals are to detecting the compound), the training of the panel preparation and education of individuals tasked with identifying odors, which can vary in effectiveness and consistency among different panels), and, finally, the rapid saturation of the sensory receptors that can quickly become overwhelmed or saturated, reducing their ability to detect odors accurately over time, factors that extend to the olfactory identification of other compounds (Takeuchi *et al.*, 2013; Grainger, 2021b). Long and deep sniffs of a wine affected by TCA would result in the taster's nose being desensitised for some time (Grainger, 2021b). TCA sensory threshold levels are often about 1.4 - 1.5 ng/L or lower (especially for white or sparkling wines) and typically vary up to 3-4 ng/L (Tarasov *et al.*, 2022).

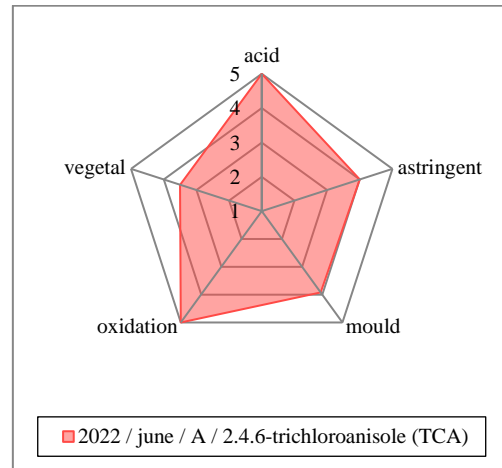


Figure 2. Example of a sample of red wine added with 2,4,6 - trichloroanisole (sample A, June 2022).

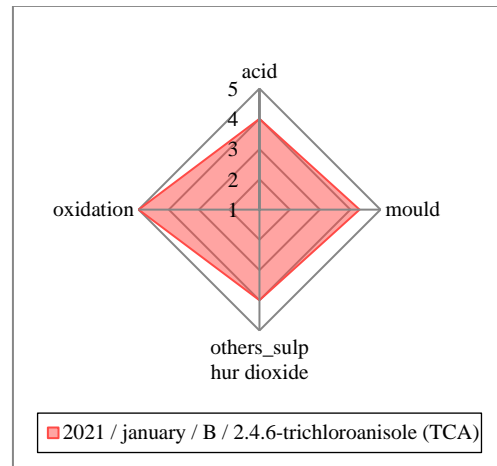


Figure 3. Example of a sample of white wine added with 2,4,6 - trichloroanisole (sample B, January 2021).

Two distinct white wine samples were examined for the detection of **oxidation** alteration in sequential tests. Both wines were supplemented with varying amounts of acetaldehyde. In the initial test, the addition of acetaldehyde to the white wine was 150 mg/L, while in the subsequent test, it was 100 mg/L. The detection rates for oxidation defects were 68.7% in the first test and 84.3% in the second. The mean intensity of perception for this defect was 4.0 and 4.1 in the first and second tests, respectively. Additionally, a test was conducted on a red wine sample treated with acetaldehyde at 200 mg/L. In this case, 60% of assessors detected an olfactory alteration and 97% rejected the sample. The mean intensity of the perceived defect among the selected assessors was 4.0. A considerable percentage of assessors detected the olfactory alteration in the red wine sample, and a high percentage rejected it.

Table VI

Summary of the results obtained in the PROVA-ALABE proficiency test for identifying defects in adulterated wine samples.

Year	Edition	N° Assessors	Sample	Grape(s) variety(ies)	Harvest year	Wine	Descriptor defect	Compound spiked in wine	Quantity added	Effective concentration	Units	Assessors (labs.) that reject the wine (%)	Assessors (labs) who identified the proposed defect (%)	The average intensity of the perceived defect [1 – 5] (SD ⁶)
2020	June	96	B	Viosinho	2017	white	lactic	isovaleric acid	10	na ¹	mg/L	92.8 (100)	60.0 (61.1)	3.8 (0.78)
2020	July	100	A	Touriga Nacional + Cabernet Sauvignon	2019	red	oxidation	acetaldehyde	200	106	mg/L	97.0 (100)	60.0 (50)	4.0 (0.75)
2020	October	118	B	Encruzado	2017	white	reduced	2-mercaptoethanol	2	na	mg/L	44.1 ²	10.0	3.2 (0.91)
2020	November	103	A	na ⁴	na	white	animal sulfur dioxide (excess)	4-ethylphenol	0.5	0.292	mg/L	94.2 (100)	58.3 (57.1)	3.5 (0.81)
2020	December	95	A	Encruzado	2017	white	sulphur dioxide (excess)	sulphur dioxide	250	127 ⁵	mg/L	87.4 (92.3)	66.3 (61.5)	3.8 (0.83)
2021	January	76	B	na ⁴	2017	white	mould	2,4-6-trichloroanisole (TCA)	100	>20	ng/L	94.7 (100)	89.5 (83.3)	4.3 (0.68)
2021	March	114	A	Touriga Nacional + Cabernet Sauvignon	2020	red	lactic	diacetyl	10	na ¹	mg/L	21.9 (68.8)	0.0	--- ³
2021	April	123	B	na ⁴	na	white	other sensory alterations	geraniol	30	na ¹	mg/L	80.5 (88.2)	22.8 (0)	4.0 (0.86)
2021	May	117	A	na ⁴	2019	white	bitter (excess) astringent	quinine sulphate	20	na ¹	mg/L	74.4 (87.5)	23.9 (31.3)	3.4 (0.97)
2021	June	127	B	Trincadeira + Syrah	2020	red	(excess) acid	grape tannin	3	na ¹	g/L	75.6 (87.5)	39.4 (43.8)	3.8 (0.90)
2021	July	103	B	na ⁴	2019	white	(excess)	tartaric acid	3	6.4	g/L	82.0 (100)	62.0 (62.5)	3.6 (0.92)
2021	September	83	A	na ⁴	2019	white	oxidation	acetaldehyde	150	82.7	mg/L	98.8 (100)	68.7 (66.7)	4.0 (0.81)
2021	October	112	B	Syrah + Touriga Nacional + Cabernet Sauvignon + Trincadeira	2020	red	plastic	styrene	1	na ¹	mg/L	87.5 (93.3)	3.6 (20)	3.5 (0.88)

¹ Not analysed; ² Statistically inconclusive trial; ³ Proposed defect not identified; ⁴ Not available (unknown grape); ⁵ Free SO₂ = 127; Total SO₂ = 219; ⁶ SD = Overall Standard Deviation; ⁷ Free SO₂ = 135; Total SO₂ = 245

Table VI (continuation)

Year	Edition	N° Assessors	Sample	Grape(s) variety(ies)	Harvest year	Wine	Descriptor defect	Compound spiked in wine	Quantity added	Effective concentration	Units	Assessors (labs.) that reject the wine (%)	Assessors (labs) who identified the proposed defect (%)	The average intensity of the perceived defect [1 – 5] (SD ⁶)
2021	November	108	A	Arinto + Macabeo + Moscatel Galego	2019	white	bitter (excess)	quinine sulphate	25	na ¹	mg/L	70.4 (71.4)	32.4 (21.4)	3.2 (0.82)
2021	December	114	B	Viosinho	2019	white	bitter	benzaldehyde	20	na ¹	mg/L	96.0 (100)	85.0 (78.6)	3.6 (0.86)
2022	January	106	B	Arinto + Macabeo + Moscatel Galego	2019	white	almond sulfur dioxide (excess)	sulphur dioxide	250	135 ⁷	mg/L	96.2 (100)	84.0 (73.3)	4.2 (0.84)
2022	February	123	B	Cabernet Sauvignon. Syrah. Touriga Nacional. Trincadeira Cabernet	2021	red	vegetal	IBMP (2-isobutyl-3-methoxy-pyrazine)	100	na ¹	ng/L	91.1 (93.3)	67.5 (73.3)	3.7 (0.82)
2022	March	120	B	Sauvignon. Syrah. Touriga Nacional. Trincadeira	2021	red	acescent	acetic acid	2	1.88	g/L	88.2 (93.8)	62.2 (56.3)	3.7 (0.82)
2022	April	116	A	na ⁴	2021	white	bitter (excess)	quinine sulphate	30	na ¹	mg/L	61.2 (50)	44.8 (52.9)	3.4 (0.80)
2022	May	100	B	Moscatel Graúdo + Macabeu	2021	white	reduced	dimethylsulphide (DMDS)	200	na ¹	ug/L	71.0 (78.6)	38.0 (42.9)	3.0 (0.80)
2022	June	102	A	Touriga Nacional. Syrah. Cabernet Sauvignon. Trincadeira	2021	red	mould	2,4,6 – trichloroanisole (TCA)	250	33	ng/L	98.0 (100)	93.1 (85.7)	3.9 (0.91)
2022	July	102	B	Encruzado	2021	white	oxidation	acetaldehyde	100	na ¹	mg/L	99.0 (100)	84.3 (84.6)	4.1 (0.77)
2022	October	114	A	Moscatel Graúdo + Macabeu	2021	white	plastic	styrene	0.25	na ¹	mg/L	60.5 (69.2)	0.0 (7.1)	--- ³
2022	November	111	B	Encruzado	2021	white	acescent	ethyl acetate	150	134.8	mg/L	76.6 (85.7)	34.2 (35.7)	3.3 (0.85)
2022	December	102	A	na ⁴	2021	red	lactic	diacetyl	20	na ¹	mg/L	35.3 (69.2)	6.9 (0)	3.3 (0.68)

¹ Not analysed; ² Statistically inconclusive trial; ³ Proposed defect not identified; ⁴ Not available (unknown grape); ⁵ Free SO₂ = 127; Total SO₂ = 219; ⁶ SD = Overall Standard Deviation; ⁷ Free SO₂ = 135; Total SO₂ = 245

Despite different amounts of acetaldehyde in the white wine tests, the intensity of perception of oxidation alteration was comparable. The red wine test showed a lower detection rate but a higher rejection rate, indicating a strong aversion to the perceived defect in the red wine sample. Adding acetaldehyde seems to have a noticeable impact on the perceived quality of white and red wines. The tests highlight the importance of the detection rate, the intensity of perception, and the overall acceptance or rejection by assessors.

These findings suggest that the type of wine and the amount of acetaldehyde added can influence how assessors perceive and respond to oxidation alterations. At sub-ppm levels, acetaldehyde can play positive roles in some specific aromatic contexts. In contrast, at higher levels, it enhances the negative effects associated with the generic presence of other aldehydes by enhancing “green vegetable” notes and “itching” character (Arias-Pérez *et al.*, 2021). Acceptable levels in white wines are generally around 20 mg/L, and in red wines around 40 mg/L. The impact upon aromas and flavors begins as low as 0.5 mg/L. The presence of the compound in wine at a level above 50 mg/L, although well below the sensory detection threshold, generally indicates that the wine has been oxidized (Grainger, 2021g).

Two separate evaluations were carried out to identify **acescent** olfactory alterations. In the first test, a red wine sample was enriched with acetic acid at a concentration of 2 g/L, while the second test involved a white wine sample with ethyl acetate added at 150 mg/L. In the initial test, 62.2% of assessors identified the defect due to acetic acid defect. However, in the second test, in which ethyl acetate was added at a level typically detectable, assessors did not categorize the acescent defect as obvious, with only 34.2% recognizing the change. Differences in matrix influence, panel awareness, and training might account for these disparities. For the first and second tests, 88.2% and 76.6% of assessors rejected the respective samples. The average intensity of the perceived defect was 3.7 and 3.3 for the first and second tests, respectively. These findings suggest that while the defect was less evident in the second test, it still led to a substantial rejection rate, and the intensity of perception was slightly lower compared to the first test. The question of when a high level of volatile acidity becomes a fault or flaw is not necessarily simple or straightforward. Sensory detection of acetic acid is not necessarily straightforward since ethyl acetate will invariably also be present, and the contents and combination of these compounds will contribute to the sensory effects. Ethyl acetate and acetic acid detection thresholds vary from wine to wine, depending on the grape variety, style of wine, and its matrix, including the alcoholic strength, among others. The sensory detection threshold of total volatile acidity generally

varies between 0.5 and 0.7 g/L (500-700 mg/L) depending on the grape variety, residual sugar, alcohol level, among others. The detection threshold may be as high as 0.9 g/L in wines from some varieties. Above threshold levels, a wine may illustrate an acrid, even bitter finish. At levels above 1.2 g/L, the aroma of the wine will generally be unpleasant, and above 1.5 g/L, there will usually be a pronounced vinegar aroma and flavor (Grainger, 2021d). The threshold at which ethyl acetate impacts wine aromas begins as low as 10 -14 mg/L, but it is often regarded as making a positive contribution when levels are below 60 mg/L. At such levels, the presence of ethyl acetate can give elements of fruitiness, a hint of cherries, richness, or even slight sweetness to a wine. When accompanied by volatile acidity below 0.6 g/L, low ethyl acetate levels can add to wine complexity with a positive effect of up to 80 mg/L ethyl acetate. However, ethyl acetate may negatively impact aromas at levels below the sensory detection threshold but above 60 mg/L, making the wine seem flat and demonstrating a loss of fruitiness (Grainger, 2021d).

Two white wine samples were presented to identify **reduction** alterations, first in an initial and subsequent in a second test. The initial white wine (first test) underwent the addition of 2-mercaptoethanol, while the second white wine (second test) was treated with dimethyldisulfide (DMDS), both belonging to the reduction defect category. The quantity of 2-mercaptoethanol added to the white wine in the first test was 2 mg/L, and for the second test, the added amount of dimethyldisulfide (DMDS) was 200 µg/L. Despite the added amounts typically sufficient to detect the reduction defect (Grainger, 2021c), assessors participating in both tests did not categorize it as an apparent defect. Only 10% of assessors identified the change in the first test with 2-mercaptoethanol and 38% in the second test with dimethyldisulfide (DMDS). The corresponding rejection rates for the first and second tests were 44.1% and 71.0%, with mean defect intensities of 3.2 and 3.0, respectively. The use of different compounds (2-mercaptoethanol and dimethyldisulfide) belonging to the same reduction defect category suggests that the specific chemical composition can influence the perception of the defect. The low detection rates (10% in the first test and 38% in the second test) highlight a potential challenge in assessor perception. It could indicate the need for more specific training or enhanced awareness of subtle alterations associated with reduction defects. The relatively high rejection rates (44.1% in the first test and 71.0% in the second test) despite the low detection rates suggest that even if not all assessors identified the reduction defect, there was a significant impact on the overall acceptability of the wine samples. The mean defect intensities values of 3.2 and 3.0 for the first and second tests indicate a moderate level of perceived defect

intensity. This suggests that while the defect might not be consistently identified, its impact on perceived quality is noteworthy.

The test to identify the **animal** defect involved adding 4-ethylphenol at a concentration of 0.5 mg/L to a white wine sample. Results from the test revealed that 58.3% of assessors could detect this olfactory alteration. Moreover, 94.2% of assessors rejected the wine sample, indicating a strong aversion to the perceived defect. The mean intensity of the defect, as perceived by the participating assessors, was recorded at 3.5, suggesting a moderate level of intensity in the identified animal aroma. Overall, these findings emphasize the impact of 4-ethylphenol on both detection rates and the rejection of the wine sample due to its association with the animal defect. When reporting the number of assessors who identified a descriptor (citation frequencies or percentage), one should be aware that the rate of assessors who identify an attribute is not a measure of the intensity of the attribute (Pérez-Elortondo and Zannoni, 2021).

In the examination aimed to identify the **sulphur dioxide** defect, a sample of white wine was subjected to the addition of sulfur dioxide at a concentration of 250 mg/L. The results revealed that 66.3% of assessors were able to discern the olfactory alteration attributed to excess sulfur dioxide. Furthermore, 87.4% of assessors rejected the wine sample, indicating a notable aversion to the perceived defect. The mean intensity of the defect, as perceived by the assessors, was recorded at 3.8, suggesting a moderate to high intensity associated with excess sulfur dioxide. These outcomes underscore the impact of sulfur dioxide on both the detection rates and the overall rejection of the wine sample due to undesirable olfactory alterations.

Three distinct tests were conducted in the assessment aimed to detect the **lactic** defect. The first involved introducing isovaleric acid at a concentration of 10 mg/L into a white wine sample. In this scenario, 60% of assessors could identify the olfactory alteration associated with isovaleric acid, leading to a 92.8% rejection rate. The average intensity of the perceived defect among the selected assessors was recorded at 3.8. For the tests involving red wine and the addition of diacetyl at 10 mg/L and 20 mg/L in the first and second tests, respectively, the defect was not overtly identified by participating assessors. Specifically, none of the assessors identified the alteration in the first test, and only 6.9% recognized it in the second. Despite the low detection rates, the rejection rates were relatively reasonable (particularly in the second test), with 21.9% and 64.7% of assessors rejecting the respective samples in the first and second tests. The most frequently reported defect in the test carried out in December 2022 (sample A) was oxidation (11.8%). There were references to other defects not objectified in the test, with very high

dispersion in their characterization. These findings underscore the potential impact of diacetyl on overall acceptability, even when all assessors do not explicitly perceive the olfactory alteration. The results were quite good considering the amount of isovaleric acid added to the wine. The sensory threshold of isovaleric acid (3-methylbutanoic or 3-methylbutyric acid) is 33 - 72 µg/l (Grainger, 2021a). The threshold for diacetyl flavors depends on the wine type and matrix; Martineau *et al.* (1995) found that the flavor detection threshold was 0.2 mg/L for a lightly aromatic 'Chardonnay' wine, 0.9 mg/L for a low tannic aromatic 'Pinot Noir' wine, and 2.8 mg/L in a full-flavored, full-bodied 'Cabernet Sauvignon' wine. Concentrations of diacetyl over 5 - 7 mg/L may be regarded as detrimental to wine quality due to off-flavors, and high levels may cause spoilage (Grainger, 2021f).

In the initial test to identify **plastic** alteration, a red wine sample was used and treated with styrene at 1 mg/L. Only 3.6% of participating assessors were able to detect the plastic defect. The average intensity of perception for this defect was reported as 3.5. Surprisingly, even with the low detection rate, 87.5% of assessors rejected the sample. In the subsequent test using a white wine sample treated with a lower styrene concentration of 0.25 mg/L, none of the assessors identified the plastic change. However, 60.5% of assessors rejected the sample, indicating a significant impact on overall acceptability, even without explicit detection. These results suggest that styrene, even at lower concentrations, can influence the rejection of wine samples, emphasizing the perceptual nuances associated with plastic alterations. The sensory perception threshold of styrene in wine is 50 - 210 µg/L (Grainger, 2021h). The fact that it is not a common fault may explain the low fault detection values. The defects most reported in the test carried out in October 2021 (sample B) were those related to reduction (23.2%) and oxidation (21.4%). There was a reference to other defects, not objectified in the test, with very high dispersion in their characterization.

In evaluating **vegetal** alterations, the test consisted of presenting a single sample of red wine to assess the presence of this defect. The red wine was treated with IBMP (2-isobutyl-3-methoxypyrazine) at a concentration of 100 ng/L. A high percentage of assessors detected the alteration, with 67.5% identifying it. Rejection rates were 91.1%, indicating a pronounced aversion to the defect detected. The average intensity of the defect detected was rated at 3.7, suggesting a moderate to high level of intensity associated with the vegetal alteration. These results highlight the perception of vegetal defects and underline their significant impact on the rejection of the sample.

When assessing additional sensory alterations, specifically involving **geraniol** compound, the test comprised presenting a sample of white wine to detect the presence of this defect. The white wine was enriched (30 mg/L) with geraniol. The assessors did not overtly identify the defect. Only 22.8% of assessors found the alteration. The rejection rates were 80.5%. The average intensity of the perceived defect was rated at 4.0. Typically, the sensory descriptors for geraniol include spicy, flowery and citrus with a detection threshold level of 30 µg/L (Australian Wine Research Institute, 2023). Considering the quantities added, the level of the reference detection threshold, the fact that it is not a common fault and is usually associated with pleasant descriptor aromas may explain the low fault detection values.

In the assessment focused on detecting the **bitter almond** defect, a white wine sample was treated with benzaldehyde at 20 mg/L. The results indicated a high detection rate, with 85.0% of the assessors identifying the olfactory alteration associated with the bitter almond defect. Furthermore, 96.0% of the assessors rejected the sample, emphasizing a solid aversion to the perceived defect. The mean intensity of the defect, as perceived by the selected assessors, was reported at 3.6, indicating a moderate to a high level of intensity. These findings underline the significant impact of benzaldehyde on both detection rates and the wine sample's overall rejection due to the bitter almond defect.

Tartaric **acid** was introduced to white wine at a concentration of 3 g/L to amplify acidity. In this evaluation, 62% of assessors were able to discern this alteration, reporting an average perceived intensity of 3.6. Subsequently, 82% of assessors rejected the sample, indicating a significant aversion to the heightened acidity introduced by the tartaric acid. These results emphasize the impact of tartaric acid on both detection rates and overall sample rejection due to the perceived excessive acidity. Perception thresholds for various acids vary according to the individual, with under 50% of tasters detecting tartaric acid in concentrations of 0.1 g/L or less and the remainder between 0.1 and 0.2 g/L. However, sweetness negates the impact of acidity, and vice versa, and the relationship between them is one of the considerations when considering balance (Grainger, 2021i).

Three distinct tests were designed to identify changes in **bitter** flavor in white wine samples by incorporating quinine sulfate at varying concentrations: 20 mg/L for the first test, 25 mg/L for the second, and 30 mg/L for the third. Despite the typically detectable amount added to the wine, the bitter taste defect was not identified by assessors in these tests. Only 23.9% of assessors detected the alteration in the first test, while the second and third tests demonstrated a somewhat higher detection rate

with 32.4% and 44.8%, respectively. Interestingly, the higher concentration of quinine sulfate in the third test resulted in more consistent results than the first and second tests despite similar average intensities of perception (3.4, 3.2, and 3.4, respectively). The rejection rates were 74.4%, 70.4%, and 61.2% for the first, second, and third tests. The findings suggest that fortifying the wine with more than 30 mg/L of quinine sulfate may be necessary to achieve a higher percentage of defect detection, closer to the desired threshold of 60%. This conclusion emphasizes the importance of considering the concentration of added substances in sensory tests to ensure reliable and consistent results in defect detection.

In an assessment targeting the detection of the **astringent** sensation defect, a red wine sample was treated with grape tannin at a concentration of 3 g/L. Despite the relatively high tannin content, only 39.4% of assessors were able to discern this alteration, concluding that it was not an apparent defect. The mean intensity of the perceived defect in the selected assessors was reported at 3.8. However, it is noteworthy that 75.6% of assessors anticipated the likely rejection of the sample, indicating a notable aversion to the perceived astringent defect. This discrepancy between the detection rate and the anticipation of rejection suggests that while not all assessors identified the defect, those who did had a significant impact on the overall acceptability of the wine sample. A reported average detection threshold for astringency was 0.14 g/L (Medel-Marabolí *et al.*, 2017). According to the results of another study, the astringency threshold was significantly higher in the elderly group (age = 75 ± 4.2 years) than in the young group (age = 29.4 ± 3.8 years) (Wang *et al.*, 2022), which may explain the values found. According to another study conducted by Pavez *et al.* (2022), the perception of overall astringency and dryness reaches a saturation point, and therefore, a greater concentration of tannins does not necessarily result in a higher score of overall astringency by the panelists during a sample evaluation.

Analyzing the results obtained in the PROVA-ALABE proficiency test for identifying defects in adulterated wine samples it was found that most of the wines were clearly rejected by most assessors, except for the following cases: white wine added with 2 mg/L of 2-mercaptoethanol, which 44.1% of assessors rejected, with only 10% identifying the reduction alteration and red wine that was added with 10 mg/L of diacetyl, which 21.9% of assessors rejected, with none identifying the lactic alteration. However, it is essential to remember that wine is a very complex matrix, and the perception of defects can be influenced by the characteristics of the "contaminated" product (Mazzoni *et al.*, 2022).

When assessing the intensity of the defects, the actual values were confined to a restricted part of the scale,

mostly between 3.5 and 4.0. The scale ranged from 1, representing low intensity, to 5, representing maximum intensity. So, it is essential to interpret the results presented with caution. Aromatic intensity does not increase linearly with the concentration value. According to Stevens' power law of psychophysics (Baldovini and Filippi, 2017), there is a relationship between odor intensity (I) and odorant concentration (C) that follows a power curve $I=f(C)$, $f(C) = kC^a$, where C is the intensity of the stimulus, $f(C)$ is the magnitude of the sensation, a is an exponent that depends on the type of stimulation, and k is a proportionality constant that depends on the units used. Nose saturation at high concentrations becomes noticeable, and the curve before saturation can be plotted on a logarithmic scale ($\text{Log}(I)=\text{Log}(f(C))$), it gives a line with a given slope (Figure 4). If the same inclination does not necessarily characterize different odorants in a given mixture, a compound with a high dilution factor may induce a weaker olfactory stimulus than another component with a lower dilution factor when that mixture is evaluated at a high concentration. In Figure 4, an odorant A is perceived as stronger than an odorant B at a high concentration C_1 ($I_A > I_B$). However, the reverse is observed at lower concentrations of C_2 ($I_A < I_B$) due to differences in the slopes of their psychometric functions.

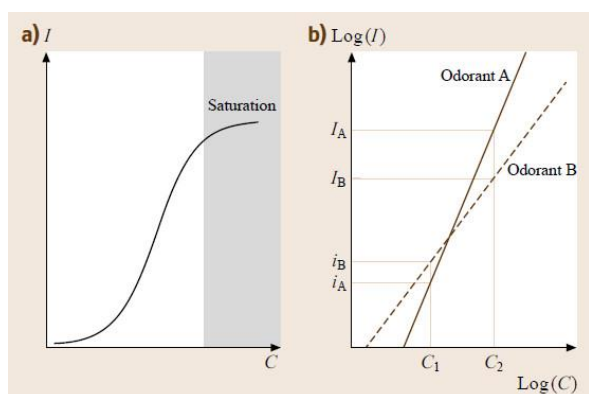


Figure 4. Psychometric function. The intensity of the olfactory stimulus (I) is represented graphically as a function of the odor concentration (C) and gives a power curve (a) or a line (b) if the scale is logarithmic. The saturation of the nose at high concentrations is visible on grey area. (Baldovini and Filippi, 2017).

The utilization of a 5-point numerical ordinal scale, yielding mean values of perceived sensory alterations typically ranging from 3.5 to 4.0, can be attributed to the "central tendency" effect, as exemplified by Cardello and Jaeger (2010). In their study, a 9-point hedonic scale was employed, illustrating that the uneven scale intervals diminish its mathematical precision to an ordered metric. This aligns with the observed "central tendency" phenomenon inherent in category scales (Xiang *et*

al., 2021). It leads assessors to avoid extreme categories, effectively transforming the 9-point scale into a 7-point scale, thereby restricting its capacity to discriminate between highly favored or disfavored samples.

The definition of the number of assessors in sensory analysis directly impacts the project's costs. For this reason, knowing precisely what the minimum number is necessary to obtain reliable data is highly relevant for the field of sensory research. The sample sizes recommended range from 25 to 300 participants, depending on the exact objective and setting (Stone and Sidel, 2004; Lawless and Heymann, 2010; Meilgaard *et al.*, 2015). Discrimination tests rarely use fewer than 20 to 25 panel members (and often up to 40 panel members) unless the products differ with fewer numbers (Wolf, 2020). ISO/IEC 17043:2023 (ISO, 2023) emphasizes the importance of considering the minimum number of participants necessary for proficiency testing objectives. If the participant number is insufficient for statistical significance, alternative approaches must be documented and shared. Considering the indication of point 5.4.2 of ISO/IEC 13528:2022 (ISO, 2022), the minimum number of participants need for the various statistical methods will depend on: the statistical method used (e.g. robust method or outlier removal); the experience of the participants with the particular PT scheme; the experience of the PT provider with the matrix, measurand, methods and group of participants and whether the intent is to determine the assigned value or standard deviation (or both). The standard recommendation is a minimum of 60 evaluators for a homogeneous target audience, or 100 to 120 evaluators following ISO 11136:2014 (ISO, 2014) for hedonic tests with consumers in a controlled area.

The results of the PROVA-ALABE proficiency test are robust since they involve a significant number of participants, and a generalized variation in the various participants' olfactory (and gustatory) sensitivity was observed.

Sensory analysis involves the use of humans as a measurement tool. This poses an immediate problem, since individuals are variable due to their experiences, expectations, and sensitivity. Thus, each person could genuinely perceive the same product quite differently. Therefore, it is essential in all sensory tests that all variables, except the one under analysis, are controlled as carefully as possible to minimize this variability (Sell, 2006).

According to Tempère *et al.* (2014), few studies have focused on the olfactory perception of wine professionals. However, they demonstrated considerable variations in the olfactory sensitivity of specialists to essential compounds in wine, including ethylphenols (Tempère *et al.*, 2011). This study also

revealed that 1% of the 134 specialists tested had hyposmia (partial loss of the sense of smell) specific to ethylphenols. This particular sensory deficit certainly affects identifying this olfactory defect within the wine's aroma profile.

According to Spence (2019), the findings from Tempère's study (Tempère *et al.*, 2011) align in this respect, as these researchers emphasized widespread differences in detection thresholds among wine experts (suggesting hyposmias or hyperosmias) for a variety of odorants relevant to wine, including β -ionone (often described as a fruity, raspberry or floral, violet aroma) in wine, 2-isobutyl-3-methoxypyrazine, TCA and linalool (which is characteristic of wines made with the 'Moscatel' grapevine variety). Individual differences in chemosensory perception prompt two crucial questions. On the one hand, Tempère *et al.* (2011) draw attention to the issue of statistical power, particularly with the frequently small sample sizes in research comparing experts and non-experts. On the other hand, it is easy to imagine how those with superior sensory abilities with respect to the chemical senses might be more likely to be drawn to, and/or more successful in achieving recognition in the world of wine-tasting.

Exceptional assessors offer significant benefits across various industries. In the food sector, their keen sensory skills are crucial for evaluating product quality, flavor, aroma, and texture, ensuring consistency and excellence. These assessors play a vital role in providing valuable feedback on consumer preferences, aiding in new product development and formulation adjustments. Their expertise extends to market research, product evaluations, and comparative testing, offering insights for strategic decision-making and marketing strategies. In fragrance and cosmetics, assessors with refined olfactory skills contribute to raw material selection, formulation development, and sensory quality assurance. Expert assessors in clinical and pharmaceutical studies evaluate medicines acceptability based on taste, texture, and aroma, especially relevant for pediatric medicines. Qualified assessors are indispensable in the beverage industry, assessing the sensory quality of wines, beers, spirits, and other drinks and providing essential information for producers and winemakers.

The overall performance indicator was calculated for the organizations and the assessors based on the acceptance/rejection test, showing performances ranging from 0.2% to 1445.3% for organizations and 0% to 80.4% for assessors. In addition, the overall performance indicator was also established for the entities and the assessors for the defect identification test where performance values between 0% and 611.5% were observed for entities and between 0% and 38.9% for tasters.

Despite analyzing a more significant number of wines in this study, the results were similar to those of a previous study by Mazzoni *et al.* (2022). This confirms that sensory analysis is an effective tool for ensuring wine quality as it examines the product as a whole. However, it requires rigorous implementation to yield reliable and high-quality results. The authors recommend participation in interlaboratory testing as an exciting tool for laboratories to meet the ISO/IEC 17025 (ISO, 2017) standard and gain trust from accreditation bodies and customers.

CONCLUSIONS

The primary goal of this work was to document and share the accumulated experience in organizing, guiding, developing, operating, selecting, and utilizing proficiency tests in sensory analysis. Special attention was given to the crucial performance evaluation aspect, which is essential for ensuring the accuracy and reliability of results. Additionally, the work aimed to explore the current state of organizing proficiency tests and their significance for organizations involved in product certification. Specifically, the study critically analyzed the outcomes of assessors' participation in the PROVA-ALABE proficiency test organized by ALABE. The focus was evaluating assessors' qualitative and quantitative responses to olfactory, tactile, and flavor stimuli associated with common wine faults. The intention is to provide a tool for maintaining, refining, or training sensory skills, helping assessors align their perceptions with those of other participants. The objectives of PROVA-ALABE as a proficiency test revolved around enhancing familiarity with wine faults and decision-making, improving sensory capabilities, and achieving accurate judgments through "standardization." This process allows for the general classification of assessors based on their ability to perceive and discriminate faults, leveraging their olfactory and gustatory memory.

Considering the percentage of assessors who rejected the wine and the percentage of assessors who identified the defect, the results suggest that some faults are more easily identified by them, mainly oxidation, mould, and animal. In general, assessors better detect the artificially added defect when the molecule responsible for the change has already been identified in previous tests. Proficiency tests support this observation, as assessors perform better when they found modifications for the second time.

The results of the intensity of the primary perceived defect confirm this observation, as there is a low dispersion of the panel members' responses when the assessors face a change for the second time. These outcomes show that interlaboratory tests can support laboratories in their continuous improvement process and can be helpful as sensory training to improve the

sensitivity of assessors, providing valuable information about evaluators' performance.

There are at least two main factors related to the efficiency of a sensory panel. The first factor, which turns out to be the most important, is the individual's inability to perceive and analyze sensory stimuli. The second aspect is the experience gained through ongoing training.

Although continuous training or repetitive exposure to the same stimulus likely improves the performance of assessors, the results obtained in the proficiency tests do not allow us to infer these facts fully. However, the uncertainty remains regarding whether extended training would benefit more participants. However, the impact of repeated sensory contact with the odorant varied widely between individuals; some exhibited a slight increase in sensitivity, while others experienced a more pronounced effect. This individual functional plasticity is potentially a decisive factor in assessing experts' skills.

In conclusion, sensory analysis is an excellent tool in wine quality control, allowing the product to be evaluated as a whole. However, like any scientific method, this discipline requires high rigor to ensure reliable and high-quality results. Accreditation bodies only accredit adequately documented and validated objective sensory tests. Participation in interlaboratory testing is a relevant strategy for laboratories to comply with the requirements established by the ISO/IEC 17025 standard and to gain the trust of accreditation bodies and their customers.

FUNDING

This study was funded by the CQ-VR [grant number UIDB/00616/2020 and UIDP/00616/2020, <https://doi.org/10.54499/UIDB/00616/2020>] and by the CEMAT/IST-ID [grant number UIDB/04621/2020 and UIDP/04621/2020, <https://doi.org/10.54499/UIDB/04621/2020>].

ACKNOWLEDGEMENTS

Acknowledgments are due to ALABE for making this study possible. Without the support, this work would not have been viable.

CONFLICTS OF INTEREST: The authors declare no conflict of interest.

REFERENCES

ALABE., 2024. Informações PROVA-ALABE. Associação dos Laboratórios de Enologia. Available at: https://www.alabe.pt/pt/prova/info_pa.php (accessed on 29.01.2024)

Arias-Pérez I., Sáenz-Navajas M. P., de-la-Fuente-Blanco A., Ferreira V., Escudero A., 2021. Insights on the role of acetaldehyde and other aldehydes in the odour and tactile nasal perception of red wine. *Food Chem.*, **361**, 130081.

Australian Wine Research Institute, 2023. Wine flavours, faults and taints. Australian Wine Research Institute. Available at: https://www.awri.com.au/industry_support/winemaking_resources/sensory_assessment/recognition-of-wine-faults-and-taints/wine_faults/ (accessed on 30.10.2023)

Baigrie B., 2003. Taints and off-flavours in foods. 206 p. Woodhead Publishing, Cambridge.

Baldovini N., Filippi J.-J., 2017. Natural Fragrant Raw Materials. In: *Springer Handbook of Odor*. 11-12. Buettner A. (ed.), Springer International Publishing, New York.

Beeren C., 2018. Application of Descriptive Sensory Analysis to Food and Drink Products. In: *In Descriptive Analysis in Sensory Evaluation*. 609-646. Kemp S.E., Hort J., Hollowood T. (eds.), Wiley Blackwell, West Sussex.

Bower J.A., 2013. The Nature of Data and Their Collection. In: *Statistical Methods for Food Science*. 15-43. Wiley-Blackwell, West Sussex.

Cardello A.V., Jaeger S.R., 2010. Hedonic measurement for product development: new methods for direct and indirect scaling. In: *Consumer-Driven Innovation in Food and Personal Care Products*. 135-174. Jaeger S. R., MacFie H. (eds.), Woodhead Publishing, Cambridge.

Daniel W.W., Cross C. L., 2018. Biostatistics: A Foundation for Analysis in the Health Sciences. 720 p. Wiley Blackwell. West Sussex.

EA, 2022. Accreditation for sensory testing laboratories. In: *European co-operation for Accreditation, Laboratory Committee*. (EA-4/09 G: 2022). European Accreditation, Utrecht.

Gomis-Bellmunt A., Claret A., Guerrero L., Pérez-Elortondo F.J., 2024. Sensory evaluation of Protected Designation of origin Wines: Development of olfactive descriptive profile and references. *Food Res. Int.*, **176**, 113828.

Goode J., 2018. Flawless: Understanding Faults in Wine. 240 p. University of California Press, Oakland.

Grainger K., 2021a. Brettanomyces (Dekkera) and Ethyl Phenols. In: *Wine Faults and Flaws* 117-170. Wiley-Blackwell, West Sussex.

Grainger K., 2021b. Chloroanisoles, Bromoanisoles, and Halophenols. In: *Wine Faults and Flaws*. 57-116. Wiley-Blackwell, West Sussex.

Grainger K., 2021c. Excessive Sulfur Dioxide, Volatile Sulfur Compounds, and Reduced Aromas. In: *Wine Faults and Flaws*. 218-256. Wiley-Blackwell, West Sussex.

Grainger K., 2021d. Excessive Volatile Acidity and Ethyl Acetate. In: *Wine Faults and Flaws*. 257-285. Wiley-Blackwell, West Sussex.

Grainger K. 2021e. Faults, Flaws, Off-Flavours, Taints, and Undesirable Compounds. In: *Wine Faults and Flaws*. 1-21. Wiley-Blackwell, West Sussex.

Grainger K., 2021f. Lactic Acid Bacteria-Related Faults. In: *Wine Faults and Flaws*. 318-330. Wiley-Blackwell, West Sussex.

Grainger K., 2021g. Oxidation, Premox, and Excessive Acetaldehyde. In: *Wine Faults and Flaws*. 171-217. Wiley-Blackwell, West Sussex.

Grainger K., 2021h. Sundry Faults, Contaminants, Including Undesirable Compounds from a Health Perspective and Flaws

- Due to Poor Balance. In: *Wine Faults and Flaws*. 352-368. Wiley-Blackwell, West Sussex.
- Grainger K., 2021i. Wine Tasting. In: *Wine Faults and Flaws*. 22-56. Wiley-Blackwell, West Sussex.
- Hyldig G., 2010. Proficiency testing of sensory panels. In: *Sensory Analysis for Food and Beverage Quality Control*. 37-48. Kilcast D. (Ed.), Woodhead Publishing, Cambridge.
- International Olive Council, 2018. Sensory analysis of olive oil: method of the organoleptic assessment of virgin olive oil. COL/T. 20/Doc. No 15/Rev. 10.
- ISO 3591, 1977. Sensory analysis- Apparatus-Wine-tasting glass. <https://www.iso.org/standard/9002.html>
- ISO 11136, 2014. Sensory analysis - Methodology - General guidance for conducting hedonic tests with consumers in a controlled area. <https://www.iso.org/standard/50125.html>
- ISO/IEC 17025, 2017. General requirements for the competence of testing and calibration laboratories. <https://www.iso.org/standard/66912.html>
- ISO 13528, 2022. Statistical methods for use in proficiency testing by interlaboratory comparison. <https://www.iso.org/standard/78879.html>
- ISO/IEC 17043, 2023. Conformity assessment-General requirements for the competence of proficiency testing providers. <https://www.iso.org/standard/80864.html>
- Kilcast D., 2010. Sensory Analysis for Food and Beverage Quality Control: A Practical Guide. 400p. Woodhead Publishing, Cambridge.
- Lawless H.T., Heymann H., 2010. Sensory Evaluation of Food: Principles and Practices (2nd Edition). 596p. Springer, New York.
- Martineau B., Acree T.E., Henick-Kling T., 1995. Effect of wine type on the detection threshold for diacetyl. *Food Res. Int.*, **28**, 139-143.
- Mazzoni C., Tirard A., Boubetra B., 2022. Detection of Wine Alterations by Sensory Analysis: Overview of Results obtained from Interlaboratory Tests. *MCAES*, **2**, 04-13.
- McEwan J.A., 1999. Comparison of sensory panels: a ring trial. *Food Qual. Prefer.*, **10**, 161-171.
- McEwan J.A., 2000a. Proficiency Testing for Sensory Profile Tests: Statistical Guidelines. Part 1. R&D Report No. 119. CCFRA. Chipping Campden, Gloucestershire.
- McEwan J.A., 2000b. Proficiency Testing for Sensory Ranking Tests: Statistical Guidelines. Part 1. R&D Report No. 118. CCFRA. Chipping Campden, Gloucestershire.
- McEwan J.A., 2001a. Proficiency Testing for Sensory Profile Tests: Statistical Guidelines. Part 2. R&D Report No. 127. CCFRA. Chipping Campden, Gloucestershire.
- McEwan J.A., 2001b. Proficiency Testing for Sensory Ranking Tests: Statistical Guidelines. Part 2. R&D Report No. 126. CCFRA. Chipping Campden, Gloucestershire.
- McEwan J.A., Heiniö R.L., Hunter E.A., Lea P., 2003. Proficiency testing for sensory ranking panels: measuring panel performance. *Food Qual. Prefer.*, **14**, 247-256.
- Medel-Marabolí M., Romero J.L., Obreque-Slier E., Contreras A., Peña-Neira A., 2017. Effect of a commercial tannin on the sensorial temporality of astringency. *Food Res. Int.*, **102**, 341-347.
- Meilgaard M.C., Civille G.V., Carr B.T., 2015. Sensory Evaluation Techniques (5 ed.). 589 p. CRC Press, Boca Raton.
- Miranda A., Pereira V., Pontes M., Albuquerque F., Marques J.C., 2017. Acetic acid and ethyl acetate in Madeira wines: Evolution with ageing and assessment of the odour rejection threshold. *Ciência Téc. Vitiv.*, **32**, 1-11.
- Murteira B., Ribeiro C.S., Silva J.A., Pimenta C., 2007. Introdução à estatística (2 ed.). 682 p. McGraw-Hill. Lisboa.
- Naës T., 1998. Detecting individual differences among assessors and differences among replicates in sensory profiling. *Food Qual. Prefer.*, **9**, 107-110.
- Næs T., Brockhoff P.B., Tomic O., 2011. Statistics for sensory and consumer science. 283 p. Wiley-Blackwell, West Sussex.
- Nakano B., 2023. A Guide to: Wine Faults. Available at: <https://www.hawaiiibevguide.com/wine-faults.html> (accessed on 20.12.2023).
- Pavez C., González-Muñoz B., O'Brien J.A., Laurie V.F., Osorio F., Núñez E., Vega R.E., Bordeu E., Brossard N., 2022. Red wine astringency: Correlations between chemical and sensory features. *LWT*, **154**, 112656.
- Pérez-Elortondo F. J., Zannoni M., 2021. Guidelines for sensory analysis of protected designation of origin food products and wines (1 ed.). Editorial Acribia S.A. Available at: https://www.e3sensory.eu/wp-content/uploads/2022/05/Guidelines-for-sensory-analysis_Seguridad_Digital.pdf (accessed on 20.12.2023).
- Pimentel T.C., Gomes da Cruz A., Deliza R., 2016. Sensory Evaluation: Sensory Rating and Scoring Methods. In: *Encyclopedia of Food and Health*, 744-749. Caballero B, Finglas P.M, Toldra F. (eds.) Elsevier Academic Press, Oxford.
- Pinto M.M., Barros P., 2015. Ensaio de aptidão sensorial. In: *Química Enológica - métodos analíticos*. 497-505. Curvelo-Garcia A.S., Barros P. (eds.), Publindústria Edições Técnicas, Agrobook.
- Rossi F., 2001. Assessing sensory panelist performance using repeatability and reproducibility measures. *Food Qual. Prefer.*, **12**, 467-479.
- Saxby M.J., 1996. Food taints and off-flavours (2 ed.). 326 p. Springer Science & Business Media, Glasgow.
- Sell C.S., 2006. The Chemistry of Fragrances: From Perfumer to Consumer (2 ed.). 321 p. Royal Society of Chemistry, Cambridge.
- Spence C., 2019. Perceptual learning in the chemical senses: A review. *Food Res. Int.*, **123**, 746-761.
- Stapleton L., 2021. Data Presentation for QC Programs. In: *The Role of Sensory Analysis in Quality Control (2nd ed.)* 55-67. Ojeh O. (ed.), ASTM International, West Conshohocken, PA.
- Stone H., Sidel J.L., 2004. Sensory Evaluation Practices (3rd ed.). 365 p. Elsevier Academic Press, Oxford.
- Takeuchi H., Kato H., Kurahashi T., 2013. 2,4,6-Trichloroanisole is a potent suppressor of olfactory signal transduction. *PNAS*, **110**, 16235-16240.
- Tarasov A., Cabral M., Loisel C., Lopes P., Schuessler C., Jung R., 2022. State-of-the-Art Knowledge about 2,4,6-Trichloroanisole (TCA) and Strategies to Avoid Cork Taint in Wine. In: *Grapes and Wine*. 324-353. Morata A., Loira I., González C. (eds.), IntechOpen, London.
- Tempère S., Cuzange E., Malak J., Bougeant J.-C., Revel G., Sicard G., 2011. The Training Level of Experts Influences their Detection Thresholds for Key Wine Compounds. *Chem. Percept.*, **4**, 99-115.
- Tempère S., Cuzange E., Schaaper M.-H., Lescar R., Revel G., Sicard G., 2014. "Brett character" in wine: Is there a consensus among professional assessors? A perceptual and conceptual approach. *Food Qual. Prefer.*, **34**, 29-36.

Thompson M., Ellison S.L.R., Wood R., 2006. The International Harmonized Protocol for the proficiency testing of analytical chemistry laboratories: (IUPAC technical report). *Pure Appl. Chem.*, **78**, 145-196.

Thompson M., Wood R., 1993. The international harmonized protocol for the proficiency testing of (chemical) analytical laboratories (Technical Report). *Pure Appl. Chem.*, **65**, 2123-2144.

Tomic O., Luciano G., Nilsen A., Hyldig G., Lorensen K., Næs T., 2010. Analysing sensory panel performance in a proficiency test using the PanelCheck software. *Eur. Food Res. Technol.*, **230**, 497-511.

Wang M., Septier C., Brignot H., Martin C., Canon F., Feron G., 2022. Astringency Sensitivity to Tannic Acid: Effect of Ageing and Saliva. *Molecules*, **27**, 1617.

Wolf M.B., 2020. Sensory Testing Methods (3 ed.). 157 p. ASTM International, West Conshohocken, PA.

Wood R., Nilsson A., Wallin H., 1998. Role of Proficiency Testing in the Assessment of Laboratory Quality. *In : Quality in the Food Analysis Laboratory*. 172-201. Royal Society of Chemistry, Cambridge.

Xiang Y., Graeber T., Enke B., Gershman S.J., 2021. Confidence and central tendency in perceptual judgment. *Atten Percept Psycho*, **83**, 3024-