

How to Improve the Monitoring of Subjective Analytical Methods and Assessors - The Case of Wine

Ivana Alpeza*, Ivna Linke, Biserka Runje

Abstract: Due to possible various physiological errors in sensory testing, regardless of knowledge and experience, it is necessary to control the results and assessors. This paper aims the possibility of some statistical tools based on laboratory control charts to monitor assessors in wine sensory analyses done in the certification procedure. The data of control charts and reliable results were processed, and individually the results of "100 points" and "Yes/No" methods. The possibility of checking the consistency and concordance of the assessors using wine faults descriptors is also presented. It has been observed that the Chi-square test and some graphical statistics, such as scatter with correlation analysis, can be a good tool in assessors monitoring. The results of sensory analyses affect the wine declaring, and therefore, continued development of technical conditions and tools in the monitoring of assessor's performance is expected to reduce subjectivity and potential errors.

Keywords: assessors; monitoring; subjective sensory analyses; wine

1 INTRODUCTION

In addition to objectively measurable quality, wine acceptability is strongly influenced by the psychological dimension and the fact that the choice of a particular wine is a complex issue burdened with emotions and like / dislike personal aptitude. According to Jackson [1], as Maynard Amerine once said, wine quality is easier to detect than define, and it is more difficult to corroborate positive quality factors than negative ones. Negative quality, such as unpleasant taste or off-odors, is generally easier to identify and describe [2].

The EU wine production and the market are regulated by common regulations and national legislation, and some of them require pre-marketing protocols that involve objective and subjective, chemical and sensory analysis. The principle of consumer protection implies that Sensory characteristics included in the specification of Protected Designation of Origin (PDO) must be guaranteed [3]. The content of some labels is consequently linked to subjective analysis, e.g. Traditional terms, which are important as quality markers and can influence product positioning [4].

1.1 Wine Quality and Subjective Methods of Analyses

Subjective sensory analysis are a large group of methods used during production and wine maturation, bottle-aging, and official quality control, and they definitely have no alternative. Any kind of instrument cannot replace the human senses because it is limited by the key human ability to connect and recognize quality as a whole, regardless of the potential impact of an ingredient. Unfortunately, subjective methods are associated with a constant risk of error; the measuring instruments are the human senses, and, although all external conditions may be at the highest level, subjectivity may cause some of the known errors. According to Stone et al. [5], there are seven groups of psychological errors in sensory testing, and some of them are central tendency (rating of samples using the middle point of scale

and avoiding the extreme ends), error of expectation (rating of samples according to expectation based on previous knowledge of the product information), halo effect (rating the same attributes when they appear in a series of questions differently from when they were tested separately), and stimulus error (assessors rate samples according to the other stimulus and not on their perception from the samples). It is therefore important to conduct sensory analysis in isolated booths by using the proper number of assessors, by training them adequately by placing them in the proper testing environment and conditions, by proper the order of sample presentation, and by using a carefully designed protocol [6]. It is not surprising that the sensory discipline encompasses dozens of ISO standards related to the field of sensory analysis (<https://www.iso.org/ics/67.240/x/>).

Given that the new professional approach in wine sensory evaluation started in the latest fifties of the last century [7], and it should be based on extensive wine knowledge, it is expected that sensory evaluation of wine is permanently built on this knowledge, experience, and research. This is the case with the methods used in the training and selection of tasters; methods and evaluation of results are well developed and mostly standardized, either at the level of broad-spectrum methods, or for sensory analysis of specific purposes or specific products [8-11].

Quality assurance of sensory testing is a demanding discipline, with the need for serious analysis and finding appropriate support tools. Perhaps the most challenging permanent question is "How to judge the wine judges?". There are several interest groups concerning that question. Most often, papers deal with methods and analyses of the quality of sensory testing in research and food industry [12, 13]. The competencies and the results of assessors on International wine competitions are also the subject of interest [14-17]. However, evaluating the sensory compliance of food products and wine with Protected Designation of Origin (PDO) with the sensory description in official specifications is difficult because there are no standard methods for such testing [18]. Some authors

presented the solutions in panel preparing and sensory evaluation in certification of protected products in Portugal and Greece [19-21], but they cannot be used beyond the specific area and medium for which they were created. Perez-Elortondo et al. [18] analysed the current situation of official sensory control of PDO food products and wines in Spain, France, and Italy, the countries that represent almost 70% of the total PDO products registered in the EU. They found a severe problem because of the different methodological approaches and technical criteria in wine sensory evaluation. This kind of sensory testing can be organized through an internal panel with a permanent assessors who work continuously and professionally as measuring instruments. In such cases, assessor monitoring can be incorporated into the methodology of routine testing [19, 20], especially when it is computerized. The laboratory work is complex when a large number of external assessors conduct testing. Although they all need official references, their practical experiences and engagements differ. It is a potential source of an error. Furthermore, there is no repetition in such sensory testing (except control samples), which is also a potential source of error.

There is an increasing demand for platforms that facilitate and integrate panel performance measurements in the routine procedures of a sensory laboratory. Some authors found that only 20% of the studies applied some tool to control the assessors [22]. Guideline for sensory analysis of PDO food products and wine is good basic support [23], but an upgrade is needed. According to Sipos et al. [13], the ISO subcommittee for sensory analysis decided to create a working document for panel performance measurement, especially for those methods which are different from the traditional descriptive analysis.

1.2 Subjective Analytical Methods in Wine Certification in Croatia

The details of wine certification with PDO in Croatia are regulated by national legislation. The sensory testing is a part of certification protocol and should confirm that sensory properties of wine comply with the corresponding specification. With that purpose, the samples are coded, and specific information about samples, such as grape variety, and vintage are presented to assessors. The testing is performed by a 7-member commission (5 assessors) and the "100 points" method that evaluates ten different parameters. Each parameter and selected expression from excellence to inadequate is accompanied by a certain number of points; the ideal wine has 100 points. Sensory results, together with other data that precede sensory analysis, determined the rights to use PDO labels and the Traditional terms "Quality wine KZP" and "Top wine KZP". These traditional terms are therefore objective indicators of quality. The method, "Yes/No" is used in the sensory testing of wines without PDO. This is a descriptive method and five parameters are analysed; limpidity and color (visual aspect), smell, taste, and overall quality.

The assessors who participate in the sensory testing are wine professionals; with an academic level of education in

viticulture and enology, and with at least 5 years of work experience in the wine sector. They have a certificate that confirms abilities and knowledge in wine sensory testing, and are obligated to participate periodically in the aptitude tests.

In addition to the technical conditions (ISO 8589), the commission and the sample preparation and analysis are accredited under the ISO/IEC 17025. The indispensable condition in laboratory work is quality, and various external and internal tools are used to ensure the quality of the results.

The control charts are used to control the results and monitor the assessors. The assessors participate in international periodic proficiency testing with blind samples. Although all assessors are certified, the possible physiological influences are respected when combining them in different commissions. It is known that women are more sensory sensitive [24], so assessors are members of both gender. It is also found that sensitivity changes over the years [25], so assessors represent different age groups. The recommended number of assessors is between five and eight in such analyse [23], which is also provided. Assessors use software produced for this purpose to facilitate their work, to be dedicated to analysis and sample.

Based on individual results and their agreement, the reliable descriptive result (yes/no) and the median (100 points) is defined. When the result is unreliable, the second analysis is carried out in a very short time by another commission.

All the elements explained ensure the objectivity of the examination and reduce the distrust of subjective methods. However, subjective, physiological sources of error cannot be eliminated. Each assessor is an individual measuring instrument and should be monitored individually. The tools that are available and could be generally accepted in this kind of analysis are limited, as already explained [23].

The aim of this study was to examine the use of control charts to develop a new model of statistical tools in the monitoring of results and assessors, to reliable the quality of sensory instruments and low variability, and to guarantee the transparency of testing. This paper is the first one dealing with the problem of subjective methods and monitoring the assessors in wine certification in Croatia.

2 MATERIAL AND METHODS

2.1 Assessors and Sensory Analyses

The research was conducted during 2019, and the results of two assessors produced during 2017 and 2018 were analysed. Two assessors of different ages and gender were randomly selected; assessor A is a 53 old male, and assessor B is a 40 old female. They have more than ten years of experience in wine sensory testing.

They used both of explained methods. In the method "100 points", the result is a median. In the case of the "Yes/No" method, a result is defined by a minimum of three (out of five) same answer. The negative result had to be explained; the assessors could indicate one or more quality descriptors from the list in software created for this purpose. All technical requirements for sensory analyses follow

ISO/DIS 8589 standard, and the procedure is accredited to HRN EN ISO/IEC 17025.

2.2 Data analyses

The results of analyses and the data of control charts were used. Control charts are based on transformed values of results in relation to the points intervals ("100 points" method) and descriptive answers ("Yes/No" method). The intervals of points are as follows: 1-59 (negative result), 60-71 (positive result, but insufficient for PDO), 72-81 (PDO, Traditional term "Kvalitetno vino KZP", minimum median 72), and 82-100 (PDO, Traditional term "Vrhunsko vino KZP", minimum median 82). A value of 0 is recorded for each individual result that is in the same points interval as the final result, and possible deviations were in the range of ± 1 to ± 3 , according to the score ranges for the quality groups and Traditional terms. Deviations can be 1, 2, 3, -1, -2, and -3. For example, with a final result of 85 and an assessor result of 70, the relative deviation is -2, otherwise, the deviation would be 2. In the case of the "Yes/No" method with two possible answers, the deviations -1 and 1 are possible. A value 0 is recorded for each descriptive result equal to the final result. On the horizontal axis are the ordinal numbers of the samples, and on the vertical are the deviations. Fig. 1 presented the first control chart of assessor A used in this study. He tested 50 samples, and had 8 deviations.

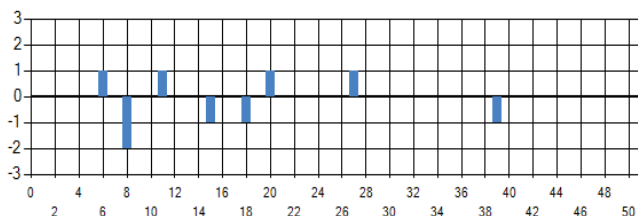


Figure 1 Control chart

Control charts contain all results, including unreliable ones. Therefore, the results were filtered and only reliable results were used in this study. The reliable result is when at least three assessors have a result in the same points interval. A reliable negative result is when at least three assessors have the same descriptors of wine faults. An unreliable result is an error caused by a measuring instrument, but it is not known with which one. To eliminate this unknown source of error and its impact on the reliability of a particular assessor in the study, we included only reliable results in analysis.

The graphical statistics, correlation analyse and Chi-square were used to test the results and the assessors. The data were analysed in Minitab Statistical Software, 2019 (Minitab, LLC, Pennsylvania, USA).

3 RESULTS AND DISCUSSION

3.1 Control Charts and Consistency of Assessor in Relation to Time; Graphical Analysis Use of Graphs

The number of testing with agreement and disagreement for assessors and per year is presented in Tab. 1. The data were calculated from the control charts, and all results were

analysed. The distribution of deviation related to intervals of points is shown in Fig. 2 and Fig. 3.

Table 1 The data of testing and results

Assessor	Year	Samples (No)	Agreement (No, %)	Deviation (No, %)
A	2017	474	409 / 86.3	65 / 13.7
	2018	433	382 / 88.2	51 / 11.8
B	2017	243	199 / 81.9	44 / 18.1
	2018	629	494 / 78.5	135 / 21.5

Most results of both assessors, in both years, were in the same points interval as results of the commissions (Tab. 1). Assessor A had a deviation in 13.7% of testing compared to the total number of testing in 2017, while in 2018, the numbers of variations were smaller. The decrease of the total number of deviations is evident in all points intervals, not only in one specific (Fig. 2).

In the case of assessor B, we found more deviations; 18.1% in 2017 and 21.5% in 2018. Variability of his results is observed in the number of variations and intervals (Fig. 3). More precisely, assessor B had a worse matching with commissions in 2018 than in 2017, and the most significant deviation was observed in intervals 1 and -1. While the share of deviations in interval -1 was significantly reduced (assessor perceives the samples as lower quality), the share of deviations in interval 1 increased in 2018 (assessor perceives the samples as better quality than the commissions).

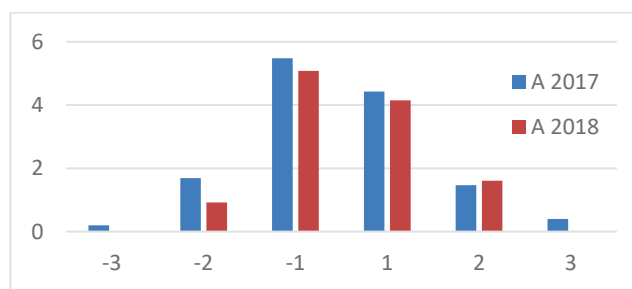


Figure 2 Assessor A; results with interval disagreement (%)

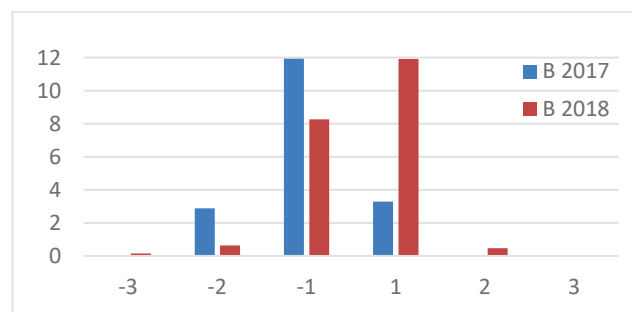


Figure 3 Assessor B; results with interval disagreement (%)

3.2 Assessor and 100-Point Method; Use of Scatter Diagram, Correlation Analysis and Chi-Square Test

A scatter plot is a graphical presentation of values in a coordinate system; the "scattering" of values determines the shape, direction, and intensity of the connection between the two variables. These diagrams are also useful for monitoring

a variable over time. In this analyse the horizontal axis presents the results of the assessor, and the vertical axis the results of the commissions for the same samples. Scatter diagrams confirm the existence of a linear relationship between the assessors and the corresponding commissions during both years (Fig. 4).

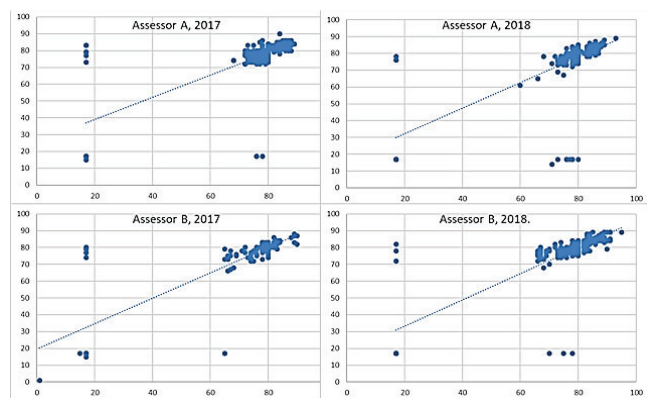


Figure 4 Scatter plot, "100 points" method

The scatter plots are used in determining the correlation and we chose the Pearson correlation coefficient as the most common measure of the direction and strength of the linear statistical correlation of the two variables. The correlation strength is defined by the range of the Pearson coefficient within the interval $[-1, 1]$. In all cases in this study, the risk of error is less than 1% given that the p-values in all cases are 0.000. In all four cases, the obtained coefficients are in the range of medium and high correlations and indicate good compliance between the assessors and the commissions, and in the appropriate direction. In the case of assessor A, Pearson's correlation coefficient was 0.68 for 2017 (257 results) and 0.66 for 2018 (264 results), which means a medium correlation. The correlation strength in the case of Assessor B was high, with a coefficient of 0.83 in 2017 (142 results) and 0.81 in 2018 (362 results). This analyse found a stronger correlation between the assessor B and the corresponding commissions, compared to assessor A, in the case of the 100-point method.

A Chi square test (χ^2) was used to test the consistency of assessors during the time. Hypotheses for the test were:

H_0 : There is no statistically significant difference between the assessor's results compared to the commissions' results during the observed period.

H_1 : There is a statistically significant difference between the assessor's results compared to the commissions' results during the observed period.

In both tests, the P-value is greater than 0.05 (Fig. 5 and 6), which means there is no evidence that there is a difference between the results of the assessor and the commissions during the observed period. Thus, it can be concluded that assessors A and B were consistent throughout the observed period in case of the "100 points" method.

It can be concluded that both assessors presented the sensitivity and consistency in case of the "100 points" method. However, a longer period than two years should be analysed to make a more significant conclusion.

These examples presented that the Chi-square test can be applicable in monitoring of assessors. It could be a good tool

for analysing the differences between assessors when they participate in multiple commissions, or when they are the permanent members of the panel in evaluating certain properties. In our study, this was not possible because assessors A and B were not members of the same commissions. Furthermore, it would be recommended to analyze a more extended period to ensure the relevance and the significance of the test.

Category	Observed	Historical Counts	Test proportion	Expected	Contribution to Chi-Sq
-3	0	2	0,007782	2	2
-2	3	4	0,015564	4	0,25
-1	22	17	0,066148	17	1,47059
0	206	215	0,836576	215	0,37674
1	18	14	0,054475	14	1,14286
2	8	5	0,019455	5	1,8
N	DF	Chi-Sq	P-Value		
257	5	7,04019	0,218		

Figure 5 χ^2 test (observed and expected counts and Chi-Square test), assessor A, 100 points method

Category	Observed	Historical Counts	Test proportion	Expected	Contribution to Chi-Sq
-2	3	4	0,011111	2	1,37178
-1	24	47	0,130556	18	1,89
0	107	287	0,797222	111	0,13126
1	5	19	0,052778	7	0,74391
2	0	3	0,008333	1	1,16
N	DF	Chi-Sq	P-Value		
257	5	7,04019	0,218		

Figure 6 χ^2 test (observed and expected counts and Chi-Square test), assessor B, "100 points" method

3.3 The Descriptors of Wine Faults as a Tool in Monitoring of Assessors

Since it is mandatory to explain the negative result and unacceptable wine quality with sensory descriptors from the available application menu, we compared the most commonly used descriptors of the commissions with the descriptors of the assessors. The wine oxidation, volatile acidity combined with the ethyl acetate, and reductive smells of hydrogen sulphide and mercaptans were tested.

It can be observed that assessor B is more sensitive in the perception of the discussed wine faults compared to the commissions (Fig. 8). In both years, assessor B has more results with the three most common wine faults than commissions. Monitoring of all negative results confirmed that she has a lower tolerance for wine quality problems in general; both in frequency and in the type of descriptor. She found more samples, with different descriptors, compared to the commissions. The choice of the same descriptors concerning assessor/commissions was more pronounced in the case of assessor A than B. There are fewer differences in general, and assessor A presented a more moderate approach in the perception of wine faults (Fig. 7). He had a smaller number of negative results compared to the commissions. This analysis pointed that the reductive smells of hydrogen

sulphide and mercaptans are the main cause of disagreement between assessor B and commissions. At the same time, it was the descriptor with the smallest deviation between assessor A and the commissions. Furthermore, this analysis shows where the problem of disagreement is most pronounced. Assessor A presented more sensitivity to oxidation than commissions, and very good matching with the commissions in case of reductive smells and volatile acidity, while assessor B definitely has a different perception of hydrogen sulphide and mercaptans compared to the commissions. This kind of analysis can be a good tool, not only in monitoring the assessors, but also in targeting of thematic trainings, and in planning of periodical aptitude tests.

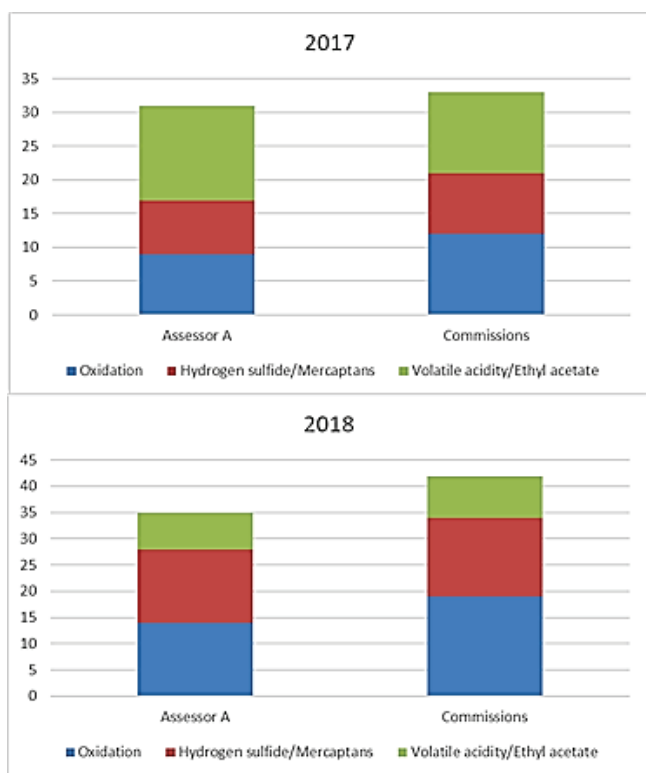


Figure 7 Descriptors of most frequency wine faults, assessor A and commissions

The assessors training, selection, and monitoring are subject of different international standards, like ISO 11037 (colour vision), ISO 3972 (taste perception), and ISO 5496 (odour sensitivity). However, using these standards, it is possible to establish a panel that can detect and describe sensory differences between the products, but not to establish and evaluate the quality of products. A problem with wine sensory testing is a huge variety of products and creations that are not accompanied by appropriate product descriptions, and it limits and complicates the training of assessors, as well as control of their work. Once the assessor selection is completed, other questions such as the monitoring of the assessors during their regular activities have to be dealt with [11]. Monitoring of assessors is not a problem when a permanent panel performs sensory analysis, or when the product has specification with detailed described sensory parameters, and different statistical models can be found in the literature [13, 26].



Figure 8 Descriptors of most frequency wine faults, assessor B and commissions

Unfortunately, we could not find the references on monitoring assessors in laboratory work when the testing is a part of pre-market wine certification performed with the larger number of external assessors in a different combination, as in our case. It is presented in this study that control charts can be used for the new statistical analysis thanks to intervals of result values. We chose the Chi-square statistic and showed how this test could be an excellent tool to upgrade the control charts; for assessors and or methods results. In similar sensory analysis, it is possible to use some other control methods, but when repetition of sample testing is performed. Pinto [27] recommended pre-testing panellists with the Triangle test and Cronbach's Alpha coefficient as a measure of assessor's consistency. Some other authors [17] also presented Cronbach's Alpha criterion for positional analysis because it allows a deeper look at the consistency and variability of assessors. This model could be useful when assessors rarely participate in testing or are limited to a specific product or parameter. However, in everyday laboratory work, as in our case, this model is complicated, increases the number of samples, and is not appropriate due to pre-expected error.

It is known that there is no repetition of samples in wine sensory analysis in certification. Sample replication is desirable but not considered in routine laboratory testing. According to Moskowitz et al. [28], when well-trained assessors are instruments with low variability, the data are still valid. Despite all the elements of quality assurance of these sensory tests based on which the result of one testing is reliable, the risk of error still exists. Replication has been found to be a good way to monitor the assessors and their mutual agreement in the case of descriptive sensory analysis. A replicated test is defined as a new independent

measurement taken under the same set of conditions as the original one. Repetition is helpful when it is possible to control samples for qualitative properties and limit the sample information presented to the assessor. For example, Pinto [29] proposed a new methodology to determine an estimate of uncertainty to be applied in interpreting results related to sensory testing of wines for the certification of some specific geographical names in Portugal. The basis of his research was the repetition of sample analysis. Unfortunately, in the case of this and older study [27], it is not clear what sample data were presented to the assessors in the test, which may be a source of error and limit the possibility of repeating the test in routine work.

However, there are test conditions when replication does not guarantee the reliability of the results. In our study, assessors in sensory testing have some crucial information about the sample; grape variety, vintage, specific technology, and some analytical data. We did some experiments, but they recognized the same sample in replicating. This is probably reinforced by the fact that the sample should be presented in the same group concerning the recommendation of the order of analysis [30]. In such conditions, the sense of applying repeat testing of the same sample is questionable due to an error that is difficult to determine, and we know it exists.

One of the assessor's work requirements is an agreement with the panel in quality descriptors and their intensity evaluation (reliability/repeatability). It is very difficult to find any information in the primary literature about the panel or assessor monitoring based on descriptors of wine faults. According to Hodgson's examination of judge reliability at a major US wine competition, the assessors tend to be more consistent in wine's sensory testing they do not like than opposite [2]. The detection of wine faults in interlaboratory tests is presented to support the improvement of the assessor's sensibility and performance [31].

According to Perez-Elortondo and Zannoni [32], if the results of one of the assessors fall outside the acceptable limit of performance, the re-training should be recommended and the assessor should provide adequate results. It is presented in this study that the analysis of attributes used in the description of wines with faults can be useful in routine laboratory work. Several aspects of control can be included; assessor sensitivity analysis, perception analysis, and consistency analysis. Sensitivity is incorporated in the final perception of the intensity of an attribute, which directly affects the result and is therefore very important in the training and selection of assessors.

The analysis of two assessors in our study raises the question of what is more important to use as a basis for evaluating the assessor's reliability; consistency or concordance. This analysis offers the thesis that the consistency of assessors is a better basis for estimating error, which is further important in their selection. Indeed, a deeper analysis (more results and a more extended period) can upgrade the graphical model with an appropriate and significant statistical tool that will define the limits of error acceptability.

Ensuring the reliability of the results requires constant monitoring of all factors that may be the source of the error. Although the reliability of the results in the analysis we have studied here is very high due to the numerous input insurance

references, including assessors who are wine experts, it is necessary to raise the scale of quality constantly.

4 CONCLUSION

Sensory analyses are burdened with many potential sources of error. Although the external causes of errors can be completely controlled, the measuring instruments (assessors) are a constant risk due to physiologically caused errors regardless of knowledge and experience. Monitoring the results is therefore not only a recommendation but also the "*Conditio sine qua non*". The monitoring of results aims to uncover the factors that produce the inconsistency. This paper presented the possibilities of some statistical tools in control of the wine assessor's work in the case of two methods of sensory analyses; "100 points" and "Yes/No". Control charts have been used as a basis for testing other options. It is presented that correlation analysis and non-parametric Chi-square test are effective tools to check the reliability and consistency of assessors. It was pointed out that graphical analysis of the type and frequency of selected wine faults descriptors is an excellent basis for monitoring the sensitivity and consistency of assessors and can be supported in the planning of training programs. These tests can be used to compare the assessors of the same panel, test the assessor's results compared to the panel, and test the assessor's consistency over time. This research needs to continue; it is proposed to include a larger number of assessors and a longer period of time in the analysis. The next level of development of quality assurance in laboratory work could include upgrading the software with selected statistical tools to monitor the results of assessors in routine work that will provide significant indicators of the quality of sensory analysis.

5 REFERENCES

- [1] Jackson, R. S. (2017). *Wine Tasting*. A Professional Handbook. Chapter 8 - Nature and Origins of Wine Quality. 337-370. Academic Press, Elsevier Ltd. <https://doi.org/10.1016/B978-0-12-801813-2.00008-2>
- [2] Hodgson, R. T. (2008). Evaluation of wine judges at a major U.S. wine competition. *Journal of Wine Economics*, 3(2), 105-113. <https://doi.org/10.1017/S1931436100001152>
- [3] Regulation (EU) No. 1151/2012 on quality schemes for agricultural products and foodstuffs.
- [4] Commission delegated Regulation (EU) No. 2019/33 supplementing Regulation (EU) No 1308/2013 regarding applications for protection of designations of origin, geographical indications and traditional terms in the wine sector. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019R0033>
- [5] Stone, H., Bleibaum, R. N., Thomas, H. A., & Sidel, J. L. (2012). *Sensory Evaluation Practices*. 4th edition. Academic Press, Inc., San Diego, CA.
- [6] Meilgaard, D., Civille, G. V., & Carr, B. (2007). *Sensory Evaluation Techniques*, 4th edition. CRC Press, Boca Raton, FL. <https://doi.org/10.1201/b16452>
- [7] Amerine, M. A., Roessler, E. B., & Filippello, F. (1959). Modern sensory methods of evaluating wine. *Hilgardia*, 28(18), 477-567. <https://doi.org/10.3733/hilg.v28n18p477>
- [8] Perez Elortondo, F. J., Ojeda, M., Albisu, M., Salmerón, J., Etayo, I., & Molina, M. (2007). Food quality certification: An

- approach for the development of accredited sensory evaluation methods. *Food Quality and Preference*, 18(2), 425-439. <https://doi.org/10.1016/j.foodqual.2006.05.002>
- [9] González, M. M., Navarro, T., Gómez, G., Pérez, R. A., & de Lorenzo, C. (2007). Sensory assessment of table olive: I. Set up of a panel test and use of standardised scales. *Grasas Aceites*, 58(3). <https://doi.org/10.3989/gya.2007.v58.i3.176>
- [10] López-Aguilar, J. R., Valerio-Alfaro, G., Monroy-Rivera, J. A., Medina-Juárez, L. A., O'Mahony, M., & Angulo O. (2007). Evaluation of a simple and sensitive sensory method for measuring rancidity in soybean oils. *Grasas Aceites*, 57, 149-154. <https://doi.org/10.3989/gya.2006.v57.i2.30>
- [11] Etaio, I., Albisu, M., Ojeda, M., Gil, P. F., Salmerón, J., & Pérez Elortondo, F. J. (2010). Sensory quality control for food certification: A case study on wine. Panel training and qualification, method validation and monitoring. *Food Cont*, 21, 542-548. <https://doi.org/10.1016/j.foodcont.2009.08.011>
- [12] King, M. C., Hall, J. W., & Cliff, M. A. (2001). A comparison of methods for evaluating the performance of a trained sensory panel. *Journal of Sensory Studies*, 16, 567-581. <https://doi.org/10.1111/j.1745-459X.2001.tb00321.x>
- [13] Sipos, L., Nyitrai, Á., Hitk, G., Friedrich, L., & Kókai, Z. (2021). Sensory Panel Performance Evaluation - Comprehensive Review of Practical Approaches. *Appl. Sci.*, 11, 11977. <https://doi.org/10.3390/app112411977>
- [14] Cicchetti, D. (2006). The Paris 1976 Wine Tastings Revisited Once More: Comparing Ratings of Consistent and Inconsistent Tasters. *Journal of Wine Economics*, 1(2), 125-140. <https://doi.org/10.1017/S193143610000016X>
- [15] Gawel, R. & Godden, P. W. (2008). Evaluation of the consistency of wine quality assessments from expert wine tasters. *Australian Journal of Grape and Wine Research*, 14, 1-8. <https://doi.org/10.1111/j.1755-0238.2008.00001.x>
- [16] Hedozeau, H. C., Ballester, J., Chatelet, B., & Lempereur, V. (2015). Wine competition: From between-juries consistency to sensory perception of consumers. *BIO Web of Conferences*, 5, 03009-03015. <https://doi.org/10.1051/bioconf/20150503009>
- [17] Khalafyan, A. A., Temerdashev, Z. A., Abakumov, A. G., & Akinshina, V. A. (2021). Study of consistency of expert evaluations of wine sensory characteristics by positional analysis. *Heliyon*, 7(2), e06162. <https://doi.org/10.1016/j.heliyon.2021.e06162>
- [18] Pérez-Elortondo, F. J., Symoneaux, R., Etaio, I., Coulon-Leroy, C., Maître, I., & Zannoni, M. (2018). Current Status and Perspectives of the Official Sensory Control Methods in Protected Designation of Origin Food Products and Wines. *Food Control*, 88, 159-168. <https://doi.org/10.1016/j.foodcont.2018.01.010>
- [20] Etaio, I., Gil, P., Ojeda, M., Albisu, M., Salmerón, J., & Elortondo F. P. (2012). Improvement of sensory quality control in PDO products: An example with txakoli white wine from Bizkaia. *Food Quality and Preferences*, 23(2), 138-147. <https://doi.org/10.1016/j.foodqual.2011.03.008>
- [21] Pinto, M. M. (2021). Evaluating uncertainty in sensory analysis. A case study of the panel of tasters of the Dão Regional Wine Commission. *Ciência e Técnica Vitivinícola*, 36(1), 22-31. <https://doi.org/10.1051/ctv/ctv2021360122>
- [22] Tataridis, P., Kannelakopoulos, C., Gatselos, L., & Logothetis, S. (2013). Sensory Evaluation of Wines as a part of Certification of Protected Designation of Origin in Greece and the EU. *AMPELOS 2013, Trends in world vitiviniculture development*. Santorini Island, Greece. <http://ampelos2013.conferences.gr/ampelos2013-proceedings.html>
- [23] Djekic, I., Lorenzo, J. M., Munekata, P. E. S., Gagaoua, M., & Tomasevic, I. (2021). Review on Characteristics of Trained Sensory Panels in Food Science. *J. Texture Stud.*, 52, 501-509. <https://doi.org/10.1111/jtxs.12616>
- [24] Doty, R. L., Applebaum, S., Zusho, H., & Settle, R. G. (1985). Sex differences in odor identification ability: A cross-cultural analysis. *Neuropsychologia*, 23(5), 667-672. [https://doi.org/10.1016/0028-3932\(85\)90067-3](https://doi.org/10.1016/0028-3932(85)90067-3)
- [25] Corwin, J., Louri, M., & Gilbert, A. N. (1995). Workplace, age, and sex as mediators of olfactory function: data from the National Geographic Smell Survey. *J. Gerontol. Ser. B Psychol. Sci. Soc. Sci.*, 50, 179-186. <https://doi.org/10.1093/geronb/50B.4.P179>
- [26] Bi, J. & Kuesten, C. (2012). Intraclass Correlation Coefficient (ICC): A Framework for Monitoring and Assessing Performance of Trained Sensory Panels and Panelists. *Journal of Sensory Studies*, 27(5), 352-364. <https://doi.org/10.1111/j.1745-459X.2012.00399.x>
- [27] Pinto, F. S. T., Fogliatto, F. S., & Qannari, E. M. (2014). A method for panelists' consistency assessment in sensory evaluations based on the Cronbach's alpha coefficient. *Food Quality and Preference*, 32, Part A, 41-47. <https://doi.org/10.1016/j.foodqual.2013.06.006>
- [28] Moskowitz, H. R., Muñoz, A. M., & Gacula, M. C. (2004). *Viewpoints and Controversies in Sensory Science and Consumer Product Testing*. Food & Nutrition Press, Connecticut, USA. <https://doi.org/10.1002/9780470385128>
- [29] Pinto, M. M. (2021). Evaluating uncertainty in sensory analysis. A case study of the panel of tasters of the Dão Regional Wine Commission. *Ciência Téc. Vitiv.*, 36(1), 22-31. <https://doi.org/10.1051/ctv/ctv2021360122>
- [30] OIV. (2009). Standard for international wine competitions and spirituous beverages of vitivinicultural origin OIV-CONCOURS 332A-2009 Retrieved from <http://www.oiv.int/public/medias/4661/oivconours-332a-2009-en.pdf>
- [31] Mazzoni, K., Tirard, A., & Boubetra, B (2022). Detection of Wine Alterations by Sensory Analysis: Overview of Results obtained from Interlaboratory Tests. *Medicon Agriculture & Environmental Sciences*, 2(1), 04-13.
- [32] Perez-Elortondo, F. J. & Zannoni, M. (2021) Guidelines for Sensory Analysis of Protected Designation of Origin Food Products and Wines. *European Sensory Science Society (E3S)*. E3S-PDO Product working group. Retrieved from: <https://www.e3sensory.eu/guidelines-for-the-sensory-evaluation/>

Authors' contacts:

Ivana Alpeza, PhD
(Corresponding author)
Croatian Agency for Agriculture and Food,
Centre for Viticulture, Enology and Edible Oils,
10000 Zagreb, Jandrićeva 42, Croatia
099 7054843, ivana.alpeza@haphi.hr

Ivna Linke
Croatian Agency for Agriculture and Food,
Centre for Viticulture, Enology and Edible Oils,
10000 Zagreb, Jandrićeva 42, Croatia
Ivna.linke@haphi.hr

Biserka Runje, PhD, Full Professor
University of Zagreb,
Faculty of Mechanical Engineering and Naval Architecture,
10000 Zagreb, Ivana Lučića 5, Croatia
biserka.runje@fsb.hr