

# Wine competition: From between-juries consistency to sensory perception of consumers

Carole Honoré-Chedozeau<sup>1</sup>, Jordi Ballester<sup>2</sup>, Bertrand Chatelet<sup>1</sup>, and Valérie Lempereur<sup>1</sup>

<sup>1</sup> Institut Français de la Vigne et du Vin (IFV) – SICAREX Beaujolais, 210 en Beaujolais, CS60320, 69661 Villefranche sur Saône Cedex, France

<sup>2</sup> Institut Universitaire de la Vigne et du Vin, Université de Bourgogne, Campus Montmuzard, Rue Claude Ladrey, BP. 21078, 21000 Dijon Cedex, France

**Abstract.** Wine is a complex product offering to consumer a multitude of choices. Considering this diversity, wine competitions provide a way of helping consumers in their choices. However, the dramatic increase of wine contests in the last fifteen years has led to an increase in the number of award-winning wines. Is a medal really a guarantee of quality or having a medal is just a matter of chance? Two studies were therefore conducted during 3 years in a French wine competition. The first study assessed the impact of the order in which the wine was assessed, and the variability between juries' judgements and therefore the chances of winning an award. For this purpose, the same wine was placed on all jury tables at the first and penultimate positions of the tasting series. Wines were rated, using the OIV-UIOE score sheet. Score differences were analyzed between tasting juries. The second study sought confirmation by a consumer panel of the medals awarded during the competition. A hedonic test of both award-winning and non-awarded wines was organised, in order to verify whether the awarded wines were indeed preferred by consumers.

## 1. Introduction

A wine consumer may be lost through the multitude of wine choices: colour, region of production, grape variety, appellation, vintage, etc. On certain wine bottles, different medals and other awards from various wine competitions and guides can help consumers in their choices. When consumers are buying a wine, the extrinsic information from the bottle packaging is the most important criteria. Awards are one of these, especially gold medals [1,2]. Moreover, for consumers with a low-involvement with wine, extrinsic information is viewed as a dimension of wine quality [3].

In France, there are 114 competitions, including 45 contests created since 2000 [4]. This dramatic increase in wine competitions in France has led to an increase in the number of award-winning wines on the shelves. Given this increase, can consumer trust all these hallmarks? In other words, does a medal guarantee the quality of a wine?

Some studies have been conducted on wine competitions. For instance, Hodgson studied the chances of winning an award for wines among 13 US wine competitions. Results showed there was no consensus among the 13 wine competitions and the likelihood of winning a gold medal was only a question of chance [5]. Another study on wine judge performance at a major wine competition showed the judges were very consistent for the wines that they did not like (wines which were rejected from the competition) but less for the wines they did like [6]. On the evaluation of the consistency of wine judgements from experienced wine tasters, Gawel and Godden showed a better consistency when the wine ratings were combined from three assessors compared with individual ratings [7].

Another study showed a positive correlation between final ratings given by wine tasters from a wine competition and wine consumers [8].

Within this context, two studies have been conducted over three editions of a wine competition in France: (1) the impact of the jury variability and the tasting order on the chances of winning an award: jury consistency, and (2) consumer preferences on awarded and non-awarded wines from the 2012 wine competition.

These studies were conducted in partnership with the regional Union of French Oenologists (UOEF).

## 2. Material and methods

Our first “between-juries consistency” study aimed at evaluating rating variability during the wine competition between tasting juries. The purposes were threefold to assess: (1) the consistency between juries, (2) the impact of the order in which the wine was served, and (3) the chances of winning an award.

The second study “consumer preferences” sought confirmation by wine consumers of the medals awarded during the 2012 wine competition using a paired preference test of both award-winning and non-awarded wines.

### 2.1. Between-juries consistency

#### 2.1.1. Tasting context: A French wine competition

The study was conducted over three editions of a French wine competition called “Trophée Lyon Beaujolais Nouveau”, from 2011 to 2013. This wine competition

takes place in Lyon, each Sunday before the third Thursday of November and is organized by the regional Union of French Oenologists (UOEF).

Trophée Lyon Beaujolais Nouveau is a wine competition in which 2 Protected Denomination of Origin (PDO) (Beaujolais Nouveau, Beaujolais Villages Nouveau), from a same grape variety (*Gamay noir*) and a same vintage, are blind tasted by wine professionals (winemakers, wine and vine professionals, wine waiters, and wine brokers) from France and other countries. About 400 samples are tasted each year by a hundred wine professionals.

Wine professionals have to rate two series from 10 to 12 wines on a 100-point scoring system. All the wines are identified by random 3-digit code and served monadically in clear INAO glasses at room temperature. A 10–15 minute break is taken between the two tasting series.

### 2.1.2. Samples

For the three editions, the wine selected for the study was a Beaujolais Nouveau red wine from the same producer and the same Cuvée. The vintage used corresponded to the year of each edition (i.e. the wine from the 2011 vintage was selected for the 2011 edition). This wine was not registered in the wine competition. Thus, the medals obtained were only hypothetical.

### 2.1.3. Assessors

Each year, wine professionals were recruited by invitation to take part in the wine competition. Assessors were grouped into juries of four to six members. Within each jury, a winemaker from the Beaujolais vineyard was designed as a jury president whose role was to help the other assessors in using the tasting score sheet, to check that all score sheets were correctly filled and to discuss about potential disagreements on wine ratings.

Prior to the beginning of the wine competition, a red wine (from the same PDO and vintage than the year of the edition) was tasted by the assessors to prepare the palate and to familiarize themselves with the tasting score sheet. This wine was commented by the organizing committee with the tasting score sheet. In 2013, three wines were tasted before the wine competition.

### 2.1.4. Tasting score sheet

The tasting score sheet used was the OIV-UIOE score sheet which is recommended for still red wine tasting in wine competitions [9]. There are four tasting criteria: visual, nose, taste and harmony. Each criterion has various descriptors assigned on a five-point discontinuous scale, differently weighted according to descriptors. The final rating, a score on 100, is obtained by combining the score scales by descriptor for each assessor of a same jury. Then, the average of all the final ratings of a same jury for each wine is calculated and converted into medal according to the following award level categories:

- Grand gold medal: at least 93 points
- Gold medal: at least 87 points
- Silver medal: at least 81 points.

### 2.1.5. Procedure

The experiment was conducted during the wine competition: the same wine was placed on all jury tables, at the first and the penultimate positions of the tasting series for the 2012 and 2013 editions and only at the first position of the tasting series for the 2011 edition.

The aims were to assess differences in wine rating between tasting juries and the chances of winning an award.

Assessors were unaware of this experiment during the three editions.

### 2.1.6. Data analysis

#### 2.1.6.1 Consistency between juries

For each edition, the data were analysed using an ANOVA (ANalysis Of VAriance) on final ratings given by all assessors for the wine in the first position. Then, Newman-Keuls's pairwise comparisons of means by jury ( $\alpha = 0.05$ ) were performed on ANOVA results in order to pinpoint the juries rating the wine differently.

For the 2012 and 2013 editions, the same analysis was also performed for the penultimate position.

#### 2.1.6.2 Order effect

A student t-test ( $\alpha = 0.05$ ) was realised on final average rating given by juries at the first and the penultimate positions for the 2012 and 2013 editions in order to highlight the rating differences between the two positions.

#### 2.1.6.3 Chances of winning an award

For the 2012 and 2013 editions, the average ratings from all juries for the two positions were converted into medals based on the award level categories system (i.e. the one used in the wine competition).

## 2.2. Consumer preferences

### 2.2.1. Tasting context: A tasting room

This second study was conducted in two tasting rooms: in Villefranche sur Saône (Rhône-Alpes, France) and in Apremont (Savoie, France), two weeks after the 2012 wine competition.

### 2.2.2. Samples

We selected 3 pairs of wines each one formed by one awarded wine (Grand gold medals) and one non-awarded wine (less than 80 points). Each pair of samples came from a different jury (Table 1). These wines were selected from the PDO Beaujolais Nouveau of the 2012 edition.

### 2.2.3. Consumer panel

A total of 115 wine consumers were recruited from the Savoie (Apremont, 31 consumers) and the Beaujolais (Villefranche sur Saône, 84 consumers) vineyards by an external recruitment. The wine consumers' characteristics are shown in Table 2.

**Table 1.** Wines selection.

Wines	Jury table	Type of producer	Average (score on 100)	Medals
AW1	2	Cooperative winery	93.83	Grand gold medal
AW2	9	Cooperative winery	93.00	Grand gold medal
AW3	17	Winemaker	93.00	Grand gold medal
NW1	2	Negociant	78.50	No medal
NW2	9	Winemaker	74.50	No medal
NW3	17	Winemaker	70.00	No medal

AW: awarded wine, NW: non-awarded wine.

The wine consumers were also questioned on their consumption habits: 83.5% of all consumers reported that they consumed mainly red wine and 47.8% of these consumed Beaujolais red wines.

#### 2.2.4. Procedure

For this second study, a hedonic paired test was conducted in order to check if Grand gold medal awarded wines were indeed preferred by wine consumers. The six wines were blind tasted in clear INAO glasses by pairs: each pair was randomly composed of an awarded wine (AW) and a non-awarded wine (NW). Assessors were asked to choose their preference between each pair, without any information about wines. Three pairs were served to each assessor, at ambient temperature.

#### 2.2.5. Data analysis

The data were analysed using a Chi-square test ( $\alpha = 0.05$ ).

**Table 2.** Wines consumers' characteristics ( $n = 115$ ).

Theme	Segment	Total panellists
Gender	Female	46
	Male	69
Age	18–25	18
	26–35	19
	36–45	22
	46–60	38
	>60	18
Wine consumption frequency	Never	0
	Rarely	27
	1–2 times per week	65
	Every day	23

### 3. Results and discussion

#### 3.1. Between-juries consistency

##### 3.1.1. Consistency between juries

Table 3 shows ANOVA summary and the post-hoc mean comparison between juries for each edition of the wine competition. Significant differences between juries were found for the 2011 and the 2012 editions but not for the 2013 one.

Results in Table 3 show that for 2011, 6 juries scored significantly higher than jury number 2, indicating that the rating was not completely homogeneous among juries. Moreover the difference between the highest and

**Table 3.** ANOVA summary and post-hoc average comparison (Newman-Keuls,  $\alpha = 0.05$ ) for each edition and each jury at the first position.

2011		2012		2013	
<i>ddl:</i> 17		<i>ddl:</i> 18		<i>ddl:</i> 18	
<i>F:</i> 1.854		<i>F:</i> 2.004		<i>F:</i> 1.264	
<i>p:</i> 0.038		<i>p:</i> 0.018		<i>p:</i> 0.233	
Juries	Average score	Juries	Average score	Juries	Average score
13	90.40 <sup>a</sup>	2	84.83 <sup>a</sup>	16	87.33 <sup>a</sup>
10	87.60 <sup>a</sup>	4	83.20 <sup>ab</sup>	3	85.60 <sup>a</sup>
11	87.20 <sup>a</sup>	8	83.20 <sup>ab</sup>	18	83.67 <sup>a</sup>
3	86.20 <sup>a</sup>	7	82.83 <sup>ab</sup>	7	83.60 <sup>a</sup>
12	86.20 <sup>a</sup>	1	82.80 <sup>ab</sup>	6	83.33 <sup>a</sup>
4	86.16 <sup>a</sup>	19	82.50 <sup>ab</sup>	1	83.20 <sup>a</sup>
8	84.40 <sup>ab</sup>	9	82.00 <sup>ab</sup>	5	82.60 <sup>a</sup>
14	84.40 <sup>ab</sup>	15	81.20 <sup>ab</sup>	19	82.33 <sup>a</sup>
15	84.25 <sup>ab</sup>	11	79.00 <sup>ab</sup>	11	82.00 <sup>a</sup>
17	83.60 <sup>ab</sup>	3	79.00 <sup>ab</sup>	8	81.80 <sup>a</sup>
7	83.00 <sup>ab</sup>	16	78.67 <sup>ab</sup>	2	81.40 <sup>a</sup>
16	83.00 <sup>ab</sup>	13	78.50 <sup>ab</sup>	17	80.83 <sup>a</sup>
1	81.80 <sup>ab</sup>	5	78.40 <sup>ab</sup>	9	80.80 <sup>a</sup>
18	81.40 <sup>ab</sup>	14	78.40 <sup>ab</sup>	13	80.60 <sup>a</sup>
6	81.20 <sup>ab</sup>	18	78.20 <sup>ab</sup>	10	80.20 <sup>a</sup>
9	81.20 <sup>ab</sup>	6	78.00 <sup>ab</sup>	15	79.40 <sup>a</sup>
5	80.80 <sup>ab</sup>	17	77.50 <sup>ab</sup>	14	79.40 <sup>a</sup>
2	71.00 <sup>b</sup>	10	74.40 <sup>ab</sup>	4	79.00 <sup>a</sup>
		12	73.20 <sup>b</sup>	12	77.50 <sup>a</sup>

Averages sharing the same letter (a,b) within a column were not significantly different.

the lowest score was almost 20 points. The 2012 edition showed an improvement on the rating homogeneity with only one significant difference between two extreme juries and a narrower range of scores (11.6 points). For the 2013 edition, no significant differences were found among juries suggesting a very good rating homogeneity.

A second ANOVA with the post-hoc mean comparison between juries was performed for the 2012 and 2013 editions of the wine competition, when the wine was placed at the penultimate position (Table 4). Significant differences between juries were also found.

When the same wine was presented in the penultimate position of the tasting series, the consensus was less strong. However, like before, results showed a clear improvement also between 2012 and 2013 editions.

**Table 4.** ANOVA summary and post-hoc average comparison (Newman-Keuls,  $\alpha = 0.05$ ) for the 2012 and 2013 editions and each jury at the penultimate position.

2012		2013	
<i>ddl:</i> 18		<i>ddl:</i> 18	
<i>F:</i> 5.676		<i>F:</i> 3.281	
<i>p:</i> <0.0001		<i>p:</i> <0.001	
Juries	Average score	Juries	Average score
8	93.60 <sup>a</sup>	17	90.67 <sup>a</sup>
10	88.80 <sup>ab</sup>	18	88.00 <sup>ab</sup>
3	87.33 <sup>ab</sup>	15	88.00 <sup>ab</sup>
2	86.17 <sup>ab</sup>	19	87.83 <sup>ab</sup>
19	85.00 <sup>abc</sup>	6	87.67 <sup>ab</sup>
16	84.83 <sup>abc</sup>	13	86.20 <sup>ab</sup>
6	84.33 <sup>abc</sup>	7	86.00 <sup>ab</sup>
7	84.33 <sup>abc</sup>	3	85.80 <sup>ab</sup>
17	84.00 <sup>abc</sup>	16	84.83 <sup>ab</sup>
13	84.00 <sup>abc</sup>	9	84.60 <sup>ab</sup>
14	83.60 <sup>bc</sup>	14	84.00 <sup>ab</sup>
4	82.40 <sup>bcd</sup>	12	82.50 <sup>ab</sup>
11	82.00 <sup>bcd</sup>	5	82.00 <sup>ab</sup>
1	81.80 <sup>bcd</sup>	1	81.00 <sup>ab</sup>
9	81.17 <sup>bcd</sup>	10	80.80 <sup>ab</sup>
12	80.60 <sup>bcd</sup>	2	79.60 <sup>b</sup>
5	75.80 <sup>cde</sup>	4	79.50 <sup>b</sup>
18	74.40 <sup>de</sup>	11	78.33 <sup>b</sup>
15	72.80 <sup>e</sup>	8	77.40 <sup>b</sup>

Averages sharing the same letter (a,b,c,d,e) within a column were not significantly different.

A possible explanation of the improvement of the rating homogeneity in 2013 could be the fact that this year three warming up red wines with three different quality levels were proposed to all the participants. The three wines were publicly commented and scored by the organizers as an illustration of how to use the score sheet. This probably had some sort of calibration effect on the assessors.

On the other hand, the lower rating homogeneity for assessment of the wine on the penultimate position with respect to the same wine on the first position, could be explained simply by the fact that, besides the common wine, each jury had different wines in the series and therefore the carryover effects on the assessment of the penultimate wine were logically different.

### 3.1.2. Order effect

For both editions, the average rating for all juries was significantly higher when the wine was tasted in the penultimate position of the tasting series (Table 5). This may be due to assessors' conservative scoring at the beginning of the series. A solution to avoid this bias would be to use different tasting orders for each assessor within a jury, as it is done in few contests [8]. Unfortunately, this is quite difficult to achieve considering the complexity of the logistics in this specific event. However, this effect is less strong for the 2013 edition, with a deviation of less than 2 points (out of 100) between the two positions. As for the rating homogeneity, the three wines used as warm up wines before the beginning of the official tasting could be responsible of the greater stability of the scores in 2013.

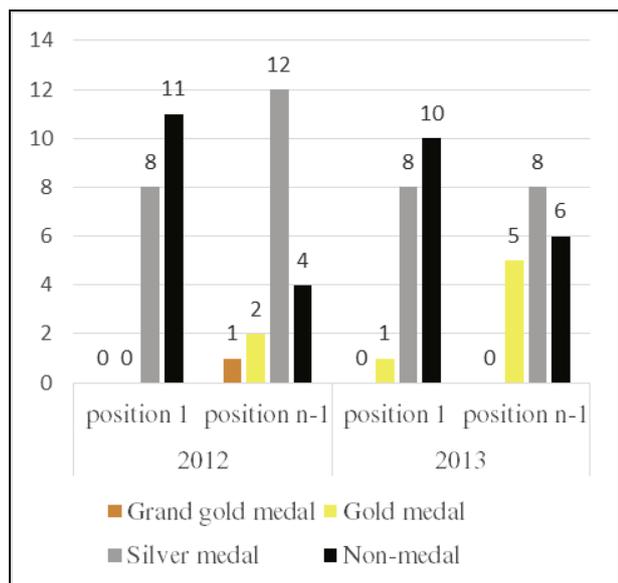
### 3.1.3. Chances of winning an award

The previous result therefore led to an increase on the chances of winning an award, multiplied by a factor of 1.9 in 2012 and 1.4 in 2013, when the same wine was tasted in penultimate position rather than in first position (Fig. 1). For 9 juries in 2012 and 7 juries in 2013, the same categories of medals were awarded for the same wine in the two positions. For the others juries in these two editions, different categories of medals were awarded by the juries in each position. Thus, there is consistency for 47.4% of the juries in 2012 and 36.8% in 2013, as far as wine rating and awarding medals are concerned. These results are higher compared to Hodgson's results in a US wine competition: about 10% of the judges were "consistently consistent" to an award level category [6].

**Table 5.** Student *t*-test summary ( $\alpha = 0.05$ ) on average ratings given by juries at the first and the penultimate positions for the 2012 and 2013 editions.

Editions	Juries	Average (position 1)	Average (position n-1)	Student <i>t</i> -test
2012	19 (9)	79.78	82.99	0.0095**
2013	19 (8)	81.78	83.93	0.016*

The number of juries which were evaluated during the wine competition is indicated in bold and those with a rating average significantly different between the two positions are indicated in brackets. Significance threshold: 5%\*, 1%\*\*\*, 0.01\*\*\*\*.



**Figure 1.** Number of hypothetical medals for the same wine placed at the first (position 1) and the penultimate (position  $n-1$ ) positions of the tasting series on all jury tables for the 2012 and 2013 wine competition editions.

### 3.2. Consumer preferences

The wine consumer panel showed no significant preference in blind test for award winning wines over wines without awards (Table 6) for each separated pair. However when we pooled all the comparisons, the Chi-square test reached significance which suggest that globally, awarded wines in this specific competition tend to be preferred by consumers. These results confirm those obtained in recent studies, indicating that consumer preferences are not always consensual and correlated with expert recommendations [10,11].

## 4. Conclusion

The first study showed that wine rating homogeneity between juries improved over time. The variability between juries was more important when the same wine was tasted in the penultimate position of the tasting series. In addition, the chances of winning an award are multiplied by 1.4–1.9 when the wine was served in the penultimate position. This result is not incoherent because different wines were tasted by assessors between the two positions (wines were different for each jury). Thus, assessors might have affined their wine perception all along the tasting according to the different wines they have tasted.

For the 2013 edition, a very good rating homogeneity was observed between juries for the same wine at the first position, in comparison with the other editions. The three warming up wines tasted and commented by the organizers, prior to the official wine competition tasting, could explain this result.

Results from consumer preferences study showed no significant preference for award winning wines over wines

**Table 6.** Paired test sensory preference for the wine consumers.

Pairs of wines	Number of tests	Preferred wines		Chi-square, $p =$
		AW	NW	
AW1/NW1	35	20	15	0.40
AW2/NW2	37	21	16	0.44
AW3/NW3	43	26	17	0.17
AW1/NW2	43	23	20	0.60
AW1/NW3	37	24	13	0.07
AW2/NW1	43	21	22	0.87
AW2/NW3	35	20	15	0.40
AW3/NW1	37	18	19	0.86
AW3/NW2	35	18	17	0.86
<b>Total</b>	<b>345</b>	<b>191</b>	<b>154</b>	<b>0.046</b>

AW: awarded wine, NW: non-awarded wine.

Significance threshold of Chi-square tests: 5%\*, 1%\*\*\*, 0.01%\*\*\*.

without awards for specific comparison. However, when all the comparisons were pooled, awarded wines showed a significant preference. Recent studies showed there was no correlation between regular consumers and experts on quality perception. Nevertheless, a correlation was shown between different groups of experts [11]. It could be explained by a different knowledge, level of expertise and involvement with wines for the consumers compared to the experts [1,11]. In this way, awards provide expert advice for guiding consumers in their choice on quality perception among the numerous wines.

This work proposes a methodology to evaluate the reliability of sensory judgement during a wine competition. Such a method is all the more necessary because of the importance attributed to medals won in competition as a guarantee of quality.

Authors thank the organizing committee of the Trophée Lyon Beaujolais Nouveau for conducting these studies during three editions of the competition, the cooperative winery for providing the wine during three vintages and the wine consumers for having participated in tests.

## References

- [1] L. Lockshin *et al*, *Using simulations from discrete choice experiments to measure consumer sensitivity to brand, region, price, and awards in wine choice*, Food Quality and Preference **17**, 166–178 (2006)
- [2] M-P. Sàenz-Navajas *et al*, *Perception of wine quality according to extrinsic cues : the case of Burgundy wine consumers*, Food Quality and Preference **27**, 44–53 (2013)

- [3] S. Charters, S. Pettigrew, *The dimensions of wine quality*, Food Quality and Preference **18**, 997–1007 (2007)
- [4] République Française, Ministère de l'économie, de l'industrie et du numérique, DGCCRF – BOCCRF Publications, NOR : EINC1508606B, *Liste des concours vinicoles français dont les médailles peuvent figurer dans l'étiquetage des vins produits en France* (2015)
- [5] R.T. Hodgson, *An Analysis of the Concordance Among 13 U.S. Wine Competitions*, Journal of Wine Economics, **4**, Issue 1, Spring, Pages 1–9 (2009)
- [6] R.T. Hodgson, *An Examination of Judge Reliability at a major U.S. Wine Competition*, Journal of Wine Economics. **3**, Issue 2, Pages 105–113 (2008)
- [7] R. Gawel, P.W. Godden, *Evaluation of the consistency of wine quality assessments from expert wine tasters*, Australian Journal of Grape and Wine Research, **4**, Issue 1, pages 1–8 (2008)
- [8] P. Deneulin *et al.*, *Concours et préférences des consommateurs, Partie 1/2: les vins médaillés sont ils ceux préférés des consommateurs ? Cas d'un concours transfrontalier*. Revue des œnologues n°**144**. Pages 51–53. (2012)
- [9] Organisation Internationale de la Vigne et du Vin, *Norme OIV des concours internationaux de vins et de boissons spiritueuses d'origine vitivinicole (OIV-CONCOURS 332A–2009)*, (2009)
- [10] H. Hopfer, H. Heymann, *Judging wine quality: do we need experts, consumers or trained panelists?*, Food Quality and Preference **32**, 221–233. (2014)
- [11] M-P. Sàenz-Navajas, *Sensory drivers of intrinsic quality of red wines. Effect of culture and level of expertise*. Food Research International **54**. 1506–1518 (2013)