Italian wines:
Ratings & reputations

Eugenio Brentari

Department of Economics and Management, University of Brescia, Brescia, Italy.

1. Introduction

The objective of this paper is to compare two different types of wine evaluation and to measure the concordance between these two different ways of evaluating wine. In particular, reference will be made, firstly, to the Altroconsumo wine guide, where the evaluations of Italian wines are the result of various analyses and assessment made blindly by expert judges; and, secondly, to the Vivino app used by millions of consumers around the world, where the evaluation is made knowing all the characteristics of the wine tasted and also the evaluations of other consumers.

2. Altroconsumo wine guide and Vivino app

From 2006 to 2021, Altroconsumo, the most widespread consumer protection and defence association in Italy, published its own wine guide. This guide considers every year about 300 Italian wines including reds, whites, and rosés, and each year has a dossier dedicated to a particular type of wine (Prosecco, Franciacorta, Lambrusco, etc.).

They are all wines that are generally not very expensive (under 20 €) and chosen to represent all the various types of wines classified according to vineyard, producer, region of origin. Moreover, these wines are frequent in large scale retail trade and wine shops and, in recent years, can be found also online.

The ratings in the Altroconsumo guide are the result of a detailed analysis of the wine: The information contained on the labels (producer, varietal, vintage, origin, colour, alcohol degree), the chemical characteristics (free and total SO2, Ph, …), the sensory characteristics (obtained from a panel of experts) and the market price.

An important note is that the evaluation of the wines in this case is made blindly by the experts, i.e., without any information about the wine and using advanced experimental sensory analysis methods.

At last of this evaluation, there is a rating scale rating from 1 to 100, divided into four bands: a score less than 50 means that it is a poor quality wine; if the score is from 50 to 59, we have medium quality wine; from 60 to 69, a good quality wine; from 70 on, an excellent quality product (Figure 1).

The other wine evaluation tool considered in what follows is the Vivino app, the world's largest wine app for platform sales. Founded as a startup in 2010 and launched as an app in 2011 by Heini Zachariassen and Theis Søndergaard, two Danish entrepreneurs, this app is available in six languages (English, Italian, Deutsch, French, Spanish, and Portuguese) since 2014.

With over 50 million downloads, and more than 13 million wines from over 231,000 wineries, the app has millions of ratings and reviews. Vivino makes hundreds of thousands of wines available for purchase in 17 markets worldwide.

The Vivino app uses an ordinal categorical rating scale similar to Likert's, from one to five stars, and also allows half a point. In addition to the average price and average rating, Vivino reports consumer reviews, taste characteristics, scents of wine, even food pairings, and information on the winery that produced it.
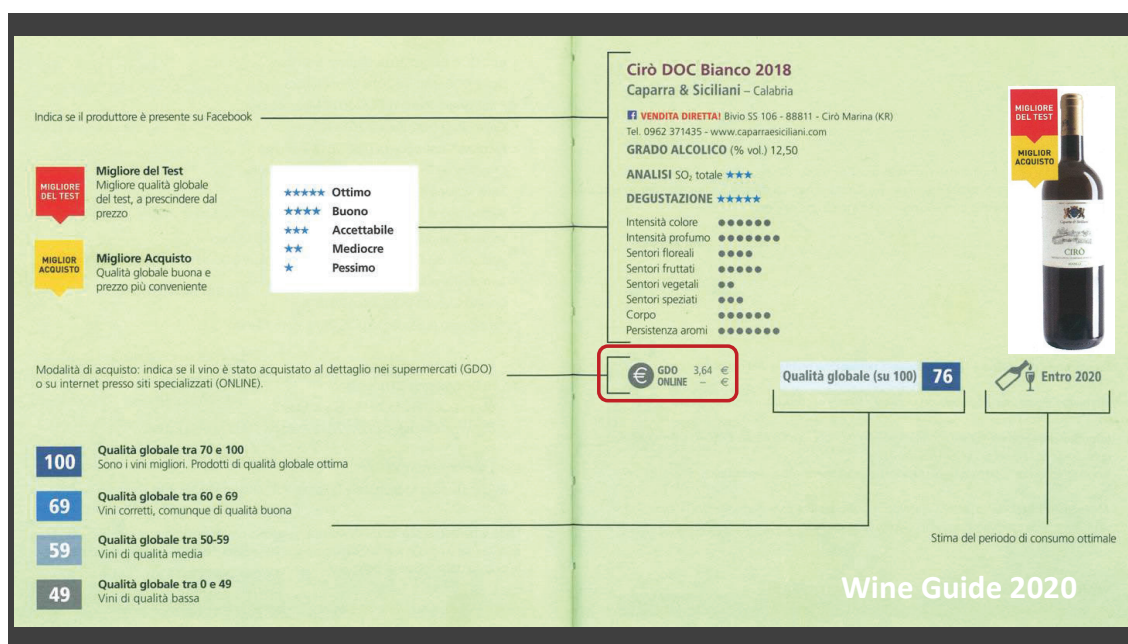
Figure 1: How to read a data sheet in the Altroconsumo guide (Wine Guide 2020)

Over the years, the distribution of Vivino scores was found to have a close-to-normal distribution, with mean and median scores both being 3.6. Surprisingly, 15% of the wines had a score of 4 (85th percentile).

Thanks to the large number of users of Vivino, the ratings of each wine on the app can be seen as a kind of reputation that the wine enjoys.

One question arises at this point: how do these two instruments behave with respect to the same wine, the two evaluation methodologies being quite different?

Or rather, is a wine reputation (resulting from consumers' judgments) a synonym for its quality (according to consumers)?

To answer these questions a simple statistical tool was considered in this paper: The analysis of correlation between Altroconsumo ratings and Vivino ratings.

3. The dataset and results

In the present study, a sample of 300 Italian wines (134 white wines, 160 red wines, and 6 rosés) was taken into consideration. The scores assigned to them by the Altroconsumo guide and those on the Vivino app were collected and compared.

As is well known, the Bravais-Pearson linear correlation coefficient r takes values from -1 to 1; positive values indicate agreement between the two assessments, while negative values indicate disagreement between the two assessments. Values close to zero indicate no relationship (Piccolo, 2010).

As can be seen in Figure 2, the assessments of Altroconsumo and Vivino are far apart: The value of r is close to zero (0.068).

An emblematic case is the first value on the left side of Figure 1, which appears as an outlier. It is a *Lambrusco Modena DOC Vino Frizzante Secco* (2015 vintage) from a well-known company that scored extremely low for Altroconsumo (21/100) while Vivino evaluations assign to it an average score of 3.3/5.

Inversely, the lowest rating given by Vivino was 2.4/5 for a *Sangiovese Merlot Rubicone IGP Assieme* red wine (2014 vintage), whereas the Altroconsumo's score for the same wine was 57/100.

The first discouraging conclusion is trivial: The Vivino evaluations are very distant from those

expressed in the Altroconsumo wine guide.

Even a separate analysis for white and red wines did not yield particularly interesting results: For white wines the value of the linear correlation coefficient r is 0.105, while for red wines it is 0.084. Only in the case of rosés the value of r turns out to be somewhat higher (0.33) but only six rosé wines were considered.
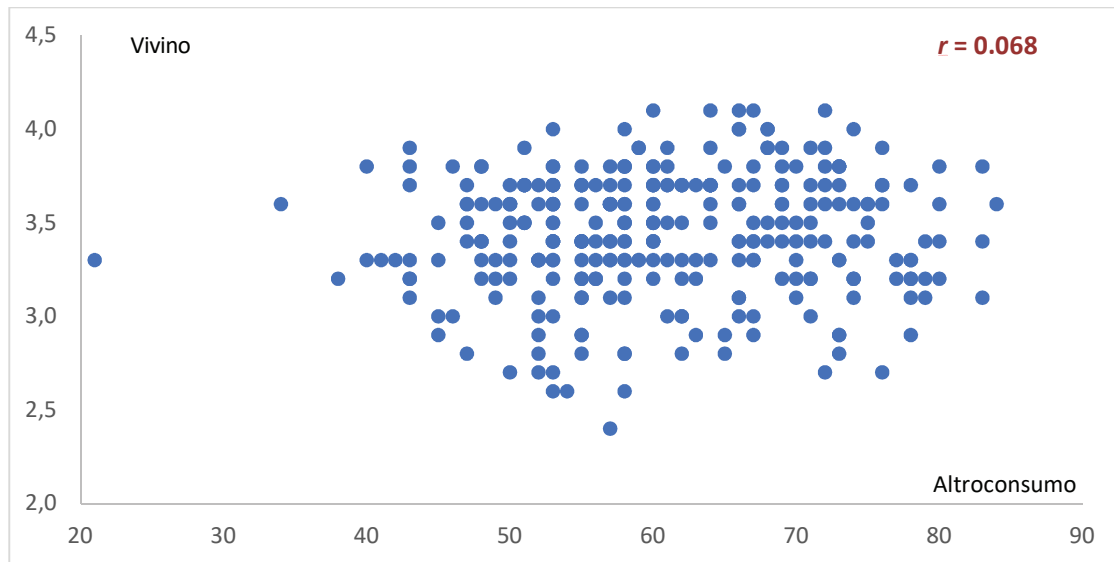

Figure 2 *Scatter plot of rankings of Altroconsumo Guide and Vivino App*

To understand the reason for this difference in ratings, the wines considered by Vivino as of medium quality (scores between 3 and 4) were studied. All the r-values are still extremely low: when all wines are considered, the r-value is 0.033, while if only white wines are considered, $r = 0.021$, and for red wines, the r-value increases slightly but only stands at 0.073.

The evaluation of medium-quality wines is particularly difficult if we consider the mid-range wines, i.e., those judged as such by both Altroconsumo and Vivino (for Altroconsumo scores from 50 to 70; for Vivino from 3 to 4), the value of the r coefficient further reduces from 0.033 to 0.030.

Hence, the panorama is certainly not exciting.

However, it should be noted that if one considers only the high-end wines and the low-end wines, in this case, the evaluations of Altroconsumo and Vivino are more in agreement with each other. In fact, if we consider the scores assigned by Vivino higher than or equal to 4 and lower than 3, the linear correlation coefficient r increases significantly, especially for white wines where it reaches a value of 0.501, while for red wines it stands at 0.348 and for all types of wine it is just under 0.3 (Figure 3).

Some improvement is obtained if the average wines of Altroconsumo are kept aside, but in this case, the residual number of wines is very small. For example, a very high r-value is obtained for red wines, but only three wines are considered.

Also, an attempt was also made to calculate the r-index jointly for all wines considered to be either very good or much less good for both Altroconsumo and Vivino. For example, by excluding all those wines within a circle with a center in the medians of the two distributions, the results change little and even in this case, are calculated on a small number of wines and therefore not very reliable for the present study.
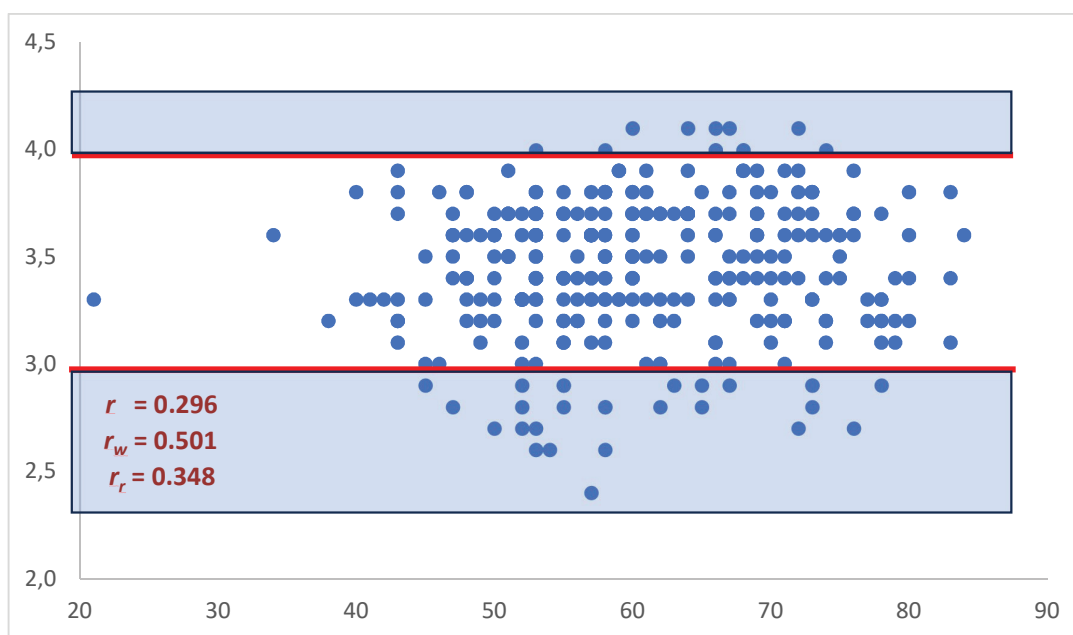
Figure 3 *Calculation of* r *for very good and bad wines of Vivino App*

An alternative tool for the analysis is the between-rank Spearman index ρ, which measures the degree of agreement or disagreement between two rankings.

In the presence of rankings with peer units, as in the present case, the Spearman index can be easily computed. The interested reader can refer, among the others, to Leti (1983, pp. 808-812). By means of this index, the analyses done previously were traced back with quite similar results.

If we analyse all the 300 wines, the Spearman index takes on a value $\rho = 0.121$, which is certainly higher than the value of the linear correlation coefficient ($r = 0.068$), but not enough to state that the evaluations of Altroconsumo and Vivino agree.

For white wines alone, the index value slightly increases ($\rho = 0.118$, very similar to $r = 0.105$), while for red wines alone, it is $\rho = 0.082$, practically the same value obtained with the linear coefficient ($r = 0.084$).

Things somewhat change when only the best and worst wines in the two rankings are considered together. All the values of ρ calculated are significantly higher than those of r, a sign that there is some agreement between the two rankings on good and bad wines.

Considering white and red wines together, the Spearman index takes the value $\rho = 0.359$. The value of the index increases when only red wines are considered; in fact, for red wines, the index is 0.410. Among the 20 red wines considered, there is little agreement on the best wines, but the rankings agree on the worst ones.

A surprising agreement qualifies white wines: the index is close to one ($\rho = 0.977$). Both evaluation guides agree, for instance, in considering a *Lugana DOC* from 2018 as the best among the white wines considered; just as they agree in considering a *Falanghina Beneventano IGP* (2014 vintage) as the worst.

Even with the Spearman's index, if one considers only the high-end wines and the low-end wines, the evaluations of Altroconsumo and Vivino are more in agreement with each other.

It should be emphasized that knowledge of wine market characteristics may heavily condition wine evaluations. In particular, knowledge of the price is positively correlated with judges' scores and this may be the case with Vivino's ratings (Table 1). In fact, while Altroconsumo's evaluations, obtained in cases the judges do not know the price of the wine, have a very low (indeed negative) correlation with the price. The same can be noticed regarding the alcoholic degree: the knowledge

of the alcoholic degree while evaluating a wine leads to a linear correlation coefficient $r = 0.429$; instead, a much lower correlation ($r = 0.241$) is shown between alcoholic degree and the Altroconsumo's evaluations.

	Alcoholic grade	Price	Altroconsumo quality	Vivino quality
Alcoholic grade	1			
Price	0,194	1		
Altroconsumo quality	**0,241**	**-0,056**	1	
Vivino quality	**0,429**	**0,313**	0,068	1

Table 1: Estimates of r between wine characteristics

4. Conclusion

In conclusion, it seems very difficult to give an objective assessment of wine without established scientific criteria, such as those used for example in sensory analysis.

A striking example comes from the evaluation of the wine considered as a "best buy" and "best test" in the Altroconsumo wine guide of 2021. In that year, the best wine was a 2018 Cirò DOC white wine from the *Calabrian company Caparra & Siciliani*, which scored 76/100 (the highest score in that guide) and the price was only € 3.64 (Figure 1). The same wine for Vivino consumers was rated 2.5/5 stars, highlighting once again the difference in judgement between the Altroconsumo and Vivino approaches.

I like to recall the advice given by Giovanni (Gianfranco) Manzoni (1923-2010), director for many years of the San Michele all'Adige Agricultural Institute (Trento, Italy), who answered the question "*How do you judge if a wine is good?*" simply by saying "*If you like it, it is good!*".

References

Altroconsumo (2006 - 2021). *Wines Guide*. Altroconsumo Edizioni S.r.l., Milano
Leti, G. (1983). *Statistica descrittiva*, Il Mulino, Bologna
Piccolo, D. (2010). *Statistica*, Il Mulino, Bologna
Vivino (2023). https://www.vivino.com/app