iVES OENO One vine & wine

DOI: https://doi.org/10.20870/oeno-one.2024.58.3.8165

OPEN ACCESS

SHORT COMMUNICATION

‘Liking then CATA’ or ‘CATA then liking’? Impact of the hedonic question positioning on the wine sensory description and appreciation

Olivier Geffroy[1], Elie Maza[2], Georgia Lytra[3] and Christian Chervin[2]*

[1] PPGV, Université de Toulouse, INP - Purpan, 75 voie du TOEC, 31076 Toulouse Cedex 3, France
[2] AgroToulouse, Université de Toulouse, INP-ENSAT, BP 32607, 31326 Castanet-Tolosan, France
[3] Univ. Bordeaux, INRAE, Bordeaux INP, Bordeaux Sciences Agro, UMR 1366, OENO, ISVV, F-33882 Villenave d’Ornon, France

*correspondence: christian.chervin@toulouse-inp.fr

Associate editor: Gilles De Revel

Received: 31 May 2024

Accepted: 6 August 2024

Published: 6 September 2024

This article is published under the **Creative Commons licence** (CC BY 4.0).

Use of all or part of the content of this article must mention the authors, the year of publication, the title, the name of the journal, the volume, the pages and the DOI in compliance with the information given above.

ABSTRACT

The sequence of the questions asked in sensory testing is crucial and remains a topic of ongoing debate. This study examines the impact of question order in CATA (Check-All-That-Apply) analyses combined with hedonic evaluations, comparing two sequences: administering the liking question before the CATA question versus the reverse order. Conducted across four different times of the year, the study utilised four distinct panels and matrices of wines: two sets of red wines with various sweeteners, one set of rosé wines, and one set of white wines. The findings indicate that placing the CATA analysis before the liking question results in: (i) an enhanced product differentiation in correspondence analysis, (ii) a reduced number of panellists needed to achieve a strong dependence between products and attributes, and (iii) a greater number of significant differences in product liking scores as determined by ANOVA, compared to when the liking question precedes the CATA analysis.

In summary, our findings show that positioning CATA before the liking question yields more detailed and discriminative results, at least with wines under our test conditions. This indicates a need for further research to understand the influence of question order, especially regarding the dependency between food products and their attributes.

KEYWORDS: Sensory analysis, CATA, hedonic, wine

INTRODUCTION

The position of the liking question when combining hedonic and descriptive analyses has been and still is a matter of debate. A few decades ago, Earthy *et al*. (1997) tested various questionnaire orders for sensory assessment of chocolate mousses, asking for preference before rating attributes, or inversely. Globally, they found no clear difference between the two questionnaire orders. Moskowitz (2004) stated that the conception of the questionnaire is a key part when conducting sensory analysis studies. He described the potential critics regarding the placing of the liking question: "If the panellist rates all of the attributes first and only then rates overall liking, then the critic can argue that the panellist may be biased from the ratings assigned prior to the overall rating. If the panellist rates liking first, and subsequently rates all of the other attributes, then the critic can argue that the panellist will try to justify the overall rating by having all other attribute ratings, and especially attribute liking ratings, confirm the overall rating" (Moskowitz, 2004). CATA analyses that check the attributes corresponding to a product have become a popular descriptive method for sensory analyses with consumers (Ares and Jaeger, 2015). We looked at recent wine sensory studies and found many with a set-up starting with the liking question prior to the CATA characterisation (e.g., Alencar *et al.*, 2019; Hayward and McSweeney, 2020; Rinaldi *et al.*, 2021), which may have been influenced by the very pedagogic method paper written by Ares and Jaeger (2015). Before writing this method paper, these two scientists co-authored a joint research article in which the two orders were tested once (Jaeger *et al.*, 2013). This well-documented article gathers a set of nine studies carried out over a large range of foods (flavoured water, savoury dips, fresh fruit, beer, crackers, and tea). The main objective of this latter article was to test the impact of the presence of CATA analyses on hedonic ratings, more than testing the order of CATA and liking questions, which was tested only once. This was performed in the sub-study with beer samples, and there was no difference between liking scores when those were given prior to or after the CATA questions. The effect of placing hedonic or CATA first was not really further tested, the impact of hedonic onto CATA being mainly tested by comparing 'hedonic only' with 'hedonic followed by CATA' (Jaeger and Ares, 2014), without that any bias could be observed in this study.

This study aimed to formally test the impact of the two possible orders, 'liking then CATA' or inversely 'CATA then liking' on correspondence analysis and ANOVA, which are the main analyses performed with such data sets, and we tested these with four different panels and sets of wines.

MATERIAL AND METHODS

Questionnaire and data collection

Four different studies were conducted: two studies with a series of 4 red wines, one study with a series of 4 rosé wines, and one study with a series of 4 white wines (see Data S1 for further details). For each study, nine attributes were chosen from a list generated by a subset of the same panellists with the same wine samples, a few days before the test with the full panel (see Data S2 for further details). We chose a restricted number of attributes, following examples of a recent study by Jaeger *et al.* (2023). In all studies, between-subjects experimental designs were used. The questionnaire and data acquisition were performed with FIZZ Nomad v2.7 (BioSystemes, France). For each study, half of the panellists got a questionnaire with the liking question first then the CATA second (see Data S3 for details according to the study), and half of the panellists got a questionnaire with the CATA first then the liking question second. The allocation of panellists to one question order was randomised upon their arrival, and the data acquisition was performed in a single tasting session. FIZZ generated random serving orders of the various wines, all coded with three-digit numbers. The volume of wine per serve and panellist was 20 mL, at 16 °C, monadically served in standard ISO wine glasses. The questionnaires/answers were presented/collected using the panellist mobile phones. In the CATA questionnaires, the attributes were randomly presented to each panellist. In all studies, at the start of each session, panellists provided informed consent and were told that their identities would remain anonymous. All panellists were recruited among AgroToulouse students in studies 1 to 3, with 64 % participating in all 3 studies, and among students and staff of the Bordeaux Sciences Agro in study 4. All panellist details are given in the following paragraphs and the associated supplemental data.

Study 1

It was conducted in October 2023 with 81 panellists (54 % female – 46 % male, 23 ± 2 years old) in Toulouse, France (see Data S2 for further details). The red wine was edulcorated with a commercial sweetener, PureVia® (Merisant, France), mainly composed of erythritol, at four dosages in g/L: 0 (named CONTROL), 10 (named PLUS10), 20 (named PLUS20) and 30 (named PLUS30). Note that erythritol is not authorised by the French wine regulatory system, and was just tested for research purposes.

Study 2

It was conducted in November 2023 with 103 panellists (56 % female – 44 % male, 23 ± 2 years old) in Toulouse, France (see Data S2 for further details). The red wine was edulcorated with a pure component of commercial sweeteners, rebaudioside-A-98 (Stevia Natura, France), at four dosages in mg/L: 0 (named CONTROL), 20 (named PLUS20), 40 (named PLUS40) and 60 (named PLUS60). Note that stevia glucosides are not authorised by the French wine regulatory system, and were just tested for research purposes.

Study 3

It was conducted in December 2023 with 101 panellists (60 % female – 40 % male, 23 ± 5 years old) in Toulouse, France (see Data S2 for further details). The rosé wines were two different blends, named Rose1 and Rose2, each wine was presented twice, named Rose1bis and Rose2bis, to reach

a total of 4 samples per questionnaire as in the two previous studies (see Data S1 for further details).

Study 4

It was conducted in February 2024 with 66 panellists (50 % female – 50 % male, 32 ± 12 years old) in Bordeaux, France (see Data S2 for further details). Four different white wines were presented, the same vintage, named according to their cultivar of origin (Muscat, Chardonnay, Riesling and Sauvignon) and different regions of origin (see Data S1 for further details). To assist the panellists in their choice of attributes, a slide showing associations of attributes and images was presented during the tasting session (see Data S4 for further details). A set of attributes was dedicated to plant aromas and we followed the methodology of the Bordeaux team, which illustrates the different attributes with images during tasting to limit the response variability.

Data analysis

All data were analysed with R software (R Core Team, 2024) using scripts developed previously (Nougarede *et al.*, 2023; Beaulieu *et al.*, 2022). All datasets and new versions of the scripts generated and used in the four studies of the present article are provided in Data S5. CATA citation scores were analysed by correspondence analyses, the dependence between attributes and wines, as a function of the panellist number was tested by a Chi-square test of independence, and the liking scores were analysed by two-way ANOVAs (panellist x product). Additionally, the liking scores were analysed by two-way ANOVAs (product x order), see Data S5.

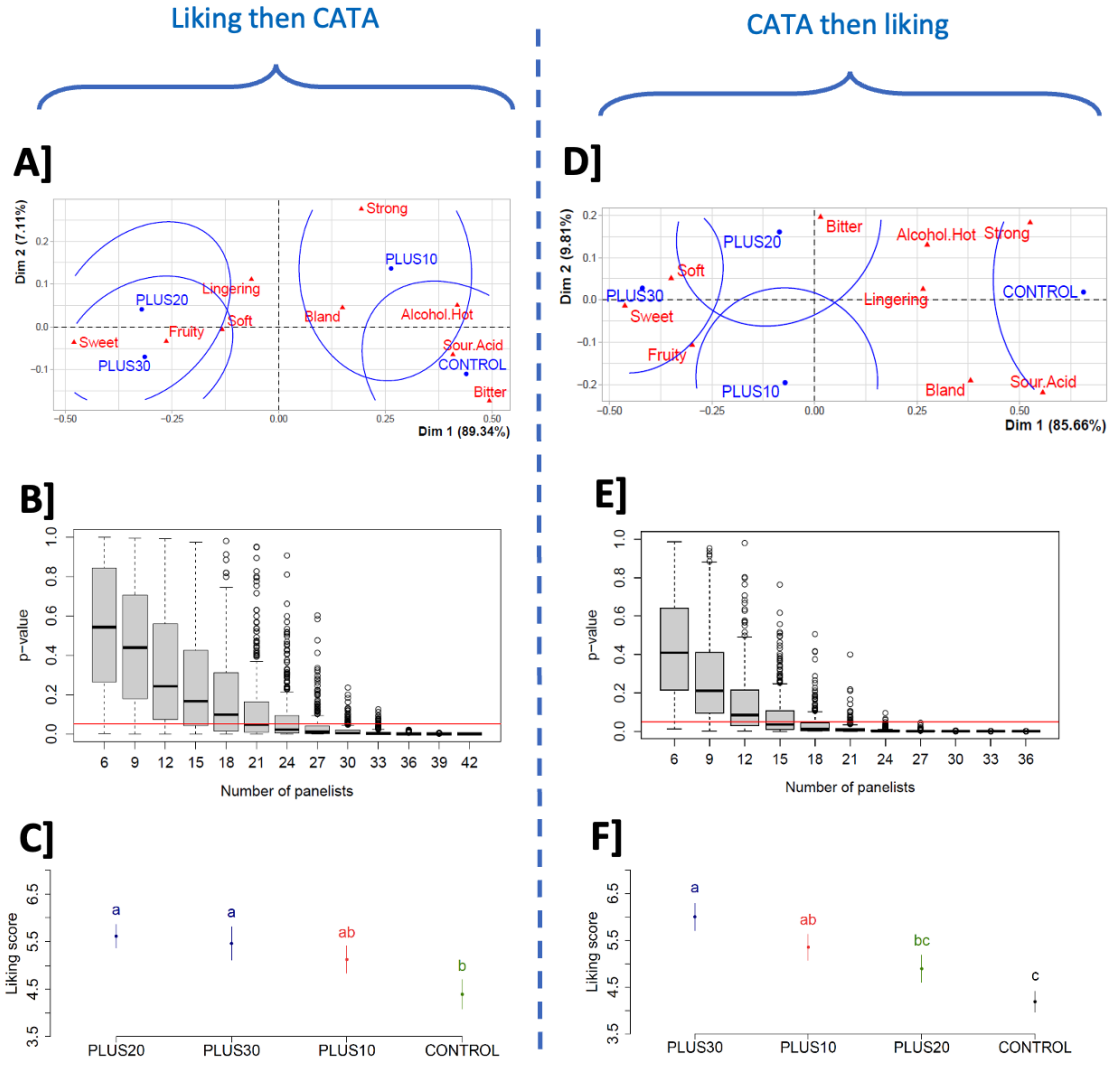


FIGURE 1. Results of the first study with red wine edulcorated with erythritol (e.g., PLUS10 stands for the addition of 10 g/L, see Material & Methods for all details), left panels show "Liking then CATA" results, right panels show "CATA then liking" results.

[A and D] Correspondence analyses; [B and E] p-value of the Chi-square test of independence between attributes and wines, as a function of the panellist number, the black line in the middle of the box is the median; [C and D] means of liking scores as a function of the wines, error bars represent the standard error of the mean, little letters show significant differences calculated with Tukey's test (*p-value* < 0.05) following ANOVA; n = 42 panellists and 39 panellists for "Liking then CATA" and "CATA then liking", respectively.

RESULTS

In the first study, with red wines edulcorated with erythritol, the correspondence analysis showed that Liking+CATA led to a clear dissociation of the sweetest and the least sweet wines (Figure 1A), without clear dissociation between the two sweetest (PLUS20 and PLUS30), and between the least sweet (CONTROL and PLUS10). The panel which performed CATA+Liking was able to better differentiate the CONTROL from the edulcorated wines, and to some extent differentiated the PLUS30 from the PLUS20 and PLUS10 (Figure 1D), however without differentiating PLUS20 and PLUS 10. In both graphs, the panellists logically associated the edulcorated wines with "sweet", "fruity" and "soft" attributes. In both graphs (Figure 1A and 1D), the sums of axis weight were very high, explaining more than 90 % of the variability. The Chi-square test of independence between wines and attributes, as a function of the panellist number, showed that the median of the p-value reached 0.05 with 21 panellists in the Liking+CATA test (Figure 1B). This highlights a strong dependence between attributes and wines, and somehow the robustness of the test, as shown before (Beaulieu *et al.*, 2022). All p-values, except outliers, fell under 0.05 from 30 panellists (Figure 1B). Whereas it took only 15 panellists to reach 0.05 with the CATA+Liking (Figure 1E), and all p-values, except outliers, fell under this threshold from 21 panellists. Finally, the ANOVA and Tukey's analyses show that Liking+CATA enabled to differentiate wines with two statistical classes (a and b) (Figure 1C), whereas it was possible to differentiate wines with three statistical classes (a, b and c) with the CATA+Liking test (Figure 1F). Looking for the attributes leading to the best liking scores, it is obvious that "sweet", "soft" and "fruity" were the preferred ones, comparing Figure 1A to Figure 1C, and Figure 1D to Figure 1F.

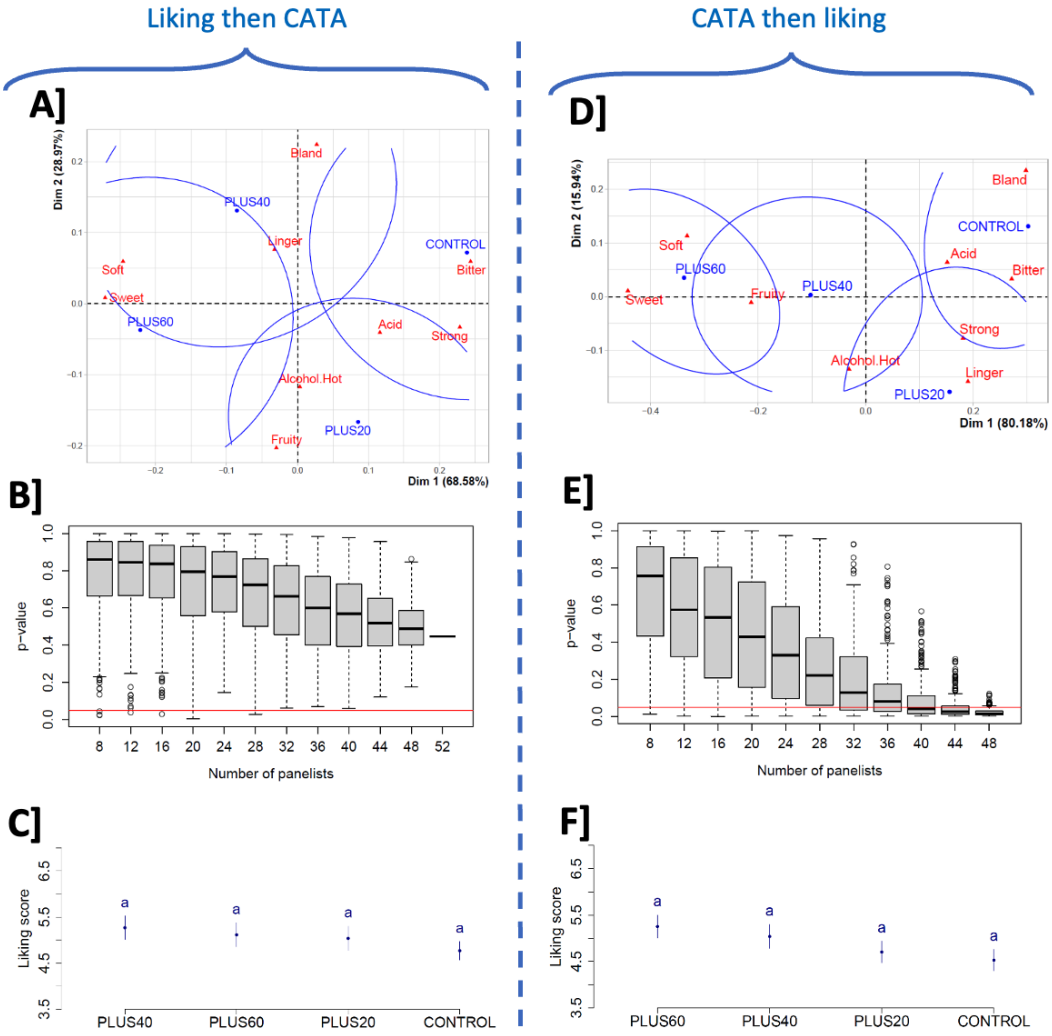


FIGURE 2. Results of the second study with red wine edulcorated with rebaudioside (e.g., PLUS20 means addition of 20 mg/L, see Material & Methods for all details), left panels show "Liking then CATA" results, right panels show "CATA then liking" results.

[A and D] Correspondence analyses; [B and E] p-value of the Chi-square test of independence between attributes and wines, as a function of the panellist number; [C and D] means of liking scores as a function of the wines, little letters show significant differences calculated with Tukey's test (*p-value* < 0.05) following ANOVA; n = 52 panellists and 51 panellists for "Liking then CATA" and "CATA then liking", respectively.

In the second study, with red wines edulcorated with pure rebaudioside, the correspondence analysis showed that Liking+CATA led to clear differentiation of the CONTROL and the "sweetest" wine (PLUS60), with a logical trend on horizontal axis from right to left, from "bitter" and "acid" to "sweet" and "soft" (Figure 2A); however, the confidence ellipses were quite large, showing some uncertainty that will be confirmed later. In the CATA+Liking part, the same trends occurred with a stronger horizontal axis and smaller confidence ellipses (Figure 1D). The Chi-square analysis revealed that in the Liking+CATA part, it would have needed more than 52 panellists for the p-value to drop under 0.05 (Figure 2B), whereas 40 panellists were sufficient for the median p-value to reach 0.05 in the CATA+Liking part (Figure 2E). The ANOVA and Tukey's tests showed a similar trend of preference for the sweetest wines whatever the questionnaire order, but no significant difference was outlined (Figure 2C and 2F). Looking for the attributes leading to the best liking scores, "sweet", "soft" and "fruity" were also the preferred ones, comparing Figure 2A to Figure 2C, and Figure 2D to Figure 2F, showing some consistency between the panels of studies 1 and 2, performed with similar red wines.

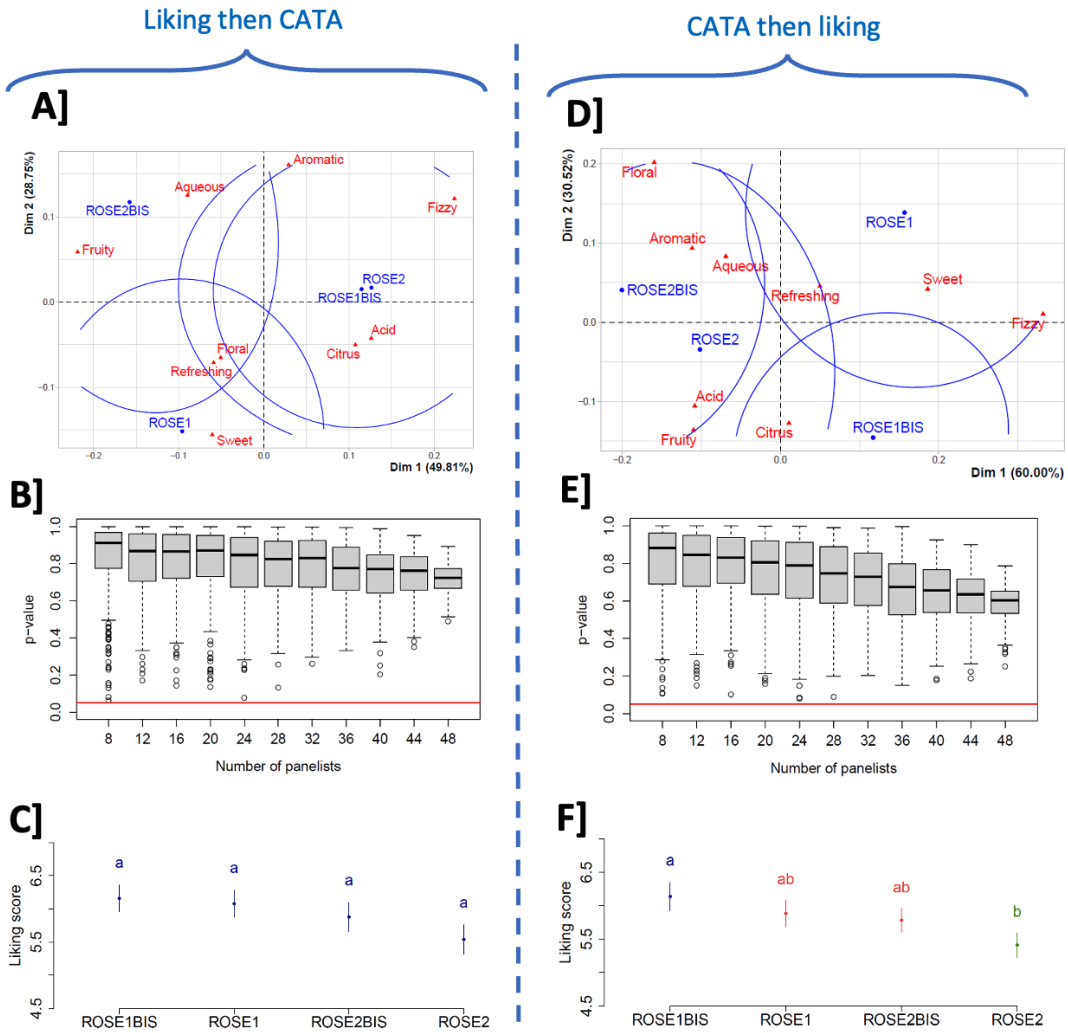


FIGURE 3. Results of the third study with two blends of rosé wines (e.g., ROSE1 and ROSE1bis stand for ROSE1 presented twice, see Material & Methods for all details), left panels show "Liking then CATA" results, right panels show "CATA then liking" results.

[A and D] Correspondence analyses; [B and E] p-value of the Chi-square test of independence between attributes and wines, as a function of the panellist number; [C and D] means of liking scores as a function of the wines, little letters show significant differences calculated with Tukey's test (*p-value* < 0.05) following ANOVA; n = 50 panellists and 51 panellists for "Liking then CATA" and "CATA then liking", respectively.

A third study was conducted with very similar rosé wines, differing only by a small percent of fruity/aromatic/sweet wine added in the ROSE1 blend. The correspondence analysis showed the uncertainty of the panellists with rather large confidence ellipses in both Liking+CATA and CATA+Liking parts (Figure 3A and 3D, respectively). In the case of CATA+Liking the panellists associated logically similar wines, ROSE1 and ROSE1BIS on the right side of the horizontal axis, and ROSE2 and ROSE2BIS on the left side of this axis, but that was not the case with Liking+CATA panel. The p-value of the Chi-square tests never reached the 0.05 threshold with 48 panellists (Figure 3B and 3E),

with a slightly better result in the CATA+Liking test, to the right of the plot. Finally, both panels preferred the ROSE1 blend, the one with the addition of an aromatic fraction (Figure 3C, F), but there was a significant difference only in the CATA+Liking test, considered marginal as ROSE1 and ROSE2BIS mean scores were not significant (Figure 3F).

Finally, a fourth study was carried out with white wines, from different cultivars, whose names serve as labels in the graphs (Figure 4). The correspondence analysis for both panels gave rather good representations on the first two axes, with sums of eigenvalues of 84 % in Liking+CATA and 92 % in CATA+Liking (Figure 4A and 4D, respectively). The four different wines were well spread out on the graphs, but the confidence ellipses were larger in Liking+CATA than in CATA+Liking showing the uncertainty/variability of the panel when ticking the attributes. This variability was further illustrated by the absence of dependence between attributes and wines in the Liking+CATA panel (Figure 4B) compared to the robustness of the analyses by the CATA+Liking panel with a Chi-square p-value of 0.05 with 33 panellists (Figure 4E). Both panels generated similar hedonic analyses with a preference for the Muscat wine, which obtained significantly better scores than Chardonnay, using ANOVA and Tukey's tests (Figure 4C and 4F). Regarding the attributes leading to the best liking scores "sweet" and "floral" were the ones (Figure 4A vs. 4C, Figure 4D vs. 4F). In all studies, the two-way ANOVAs (product x order) revealed no difference in the order of the liking score (see Data S5).

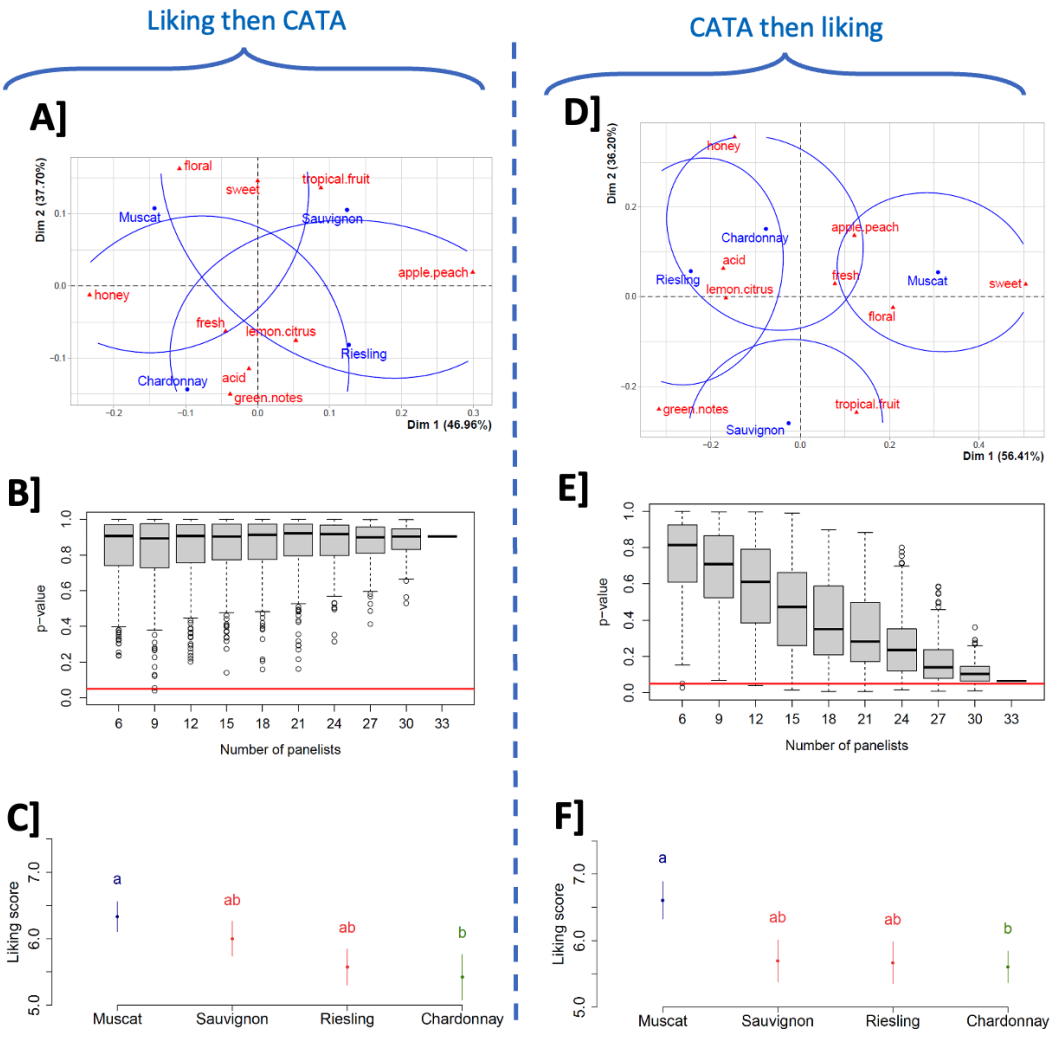


FIGURE 4. Results of the fourth study with four white wines (e.g., Muscat stands for dry Muscat wine, see Material & Methods for all details), left panels show "Liking then CATA" results, right panels show "CATA then liking" results.

[A and D] Correspondence analyses; [B and E] p-value of the Chi-square test of independence between attributes and wines, as a function of the panellist number; [C and D] means of liking scores as a function of the wines, little letters show significant differences calculated with Tukey's test (*p-value* < 0.05) following ANOVA; n = 33 panellists and 33 panellists for "Liking then CATA" and "CATA then liking", respectively.

DISCUSSION

When looking at some statistical indicators, extracted from the Correspondence Analyses and ANOVAs of the four studies (Table 1), it is quite obvious that CATA+Liking gave better results, in green cell backgrounds than Liking+CATA.

The sums of eigenvalues for the first two axes, in the eight CAs, were rather high (sums ranging from 78.6 % to 97.6 %), whatever the question order, suggesting that the attributes were adapted to describe wine differences (van Dam *et al.* 2021, and refs herein). What is not shown in Table 1 is the uncertainty/variability outlined by confidence ellipses, which tended to be larger in Liking+CATA than in CATA+Liking (see Figure 2A vs. Figure 2D and Figure 4A vs. Figure 4D).

The Chi-square test of independence between attributes and wines as a function of the panellist number revealed the largest discrepancy between both orders of asking questions. One would expect to have a good dependence between attributes and wines to have a robust analysis (Mahieu *et al.*, 2020). The p-value of this Chi-square test should then be as low as possible. Hence it reached 0.05 only one time in the Liking+CATA studies, when it reached this threshold three times in the CATA+liking studies. The observations of the results led to many questions. One of them is: Was there a sufficient number of panellists? Probably not in the third study with Rosé wines, which were very similar products, and it is well known that highlighting tiny differences usually requires larger samples (Hough *et al.*, 2006). Whereas in studies 1, 2 and 4, the p-value of 0.05 was reached in the CATA+Liking studies, but not in the Liking+CATA studies. The main question for which we have no answer is: Why does the way of ordering questions impact the dependence between wines and attributes? In the particular case of Liking+CATA, one can say that the panellists may have ticked the attributes with less "accuracy", because of what Moskowitz (2004) wrote: "panellist will try to justify the overall rating by having all other attribute ratings confirm the overall rating". However, this was not a question we targeted and further research would be needed.

One can argue that our panels were not identical in Liking+CATA or in CATA+Liking, therefore, it is erroneous to compare the data. However, we randomly addressed each panellist entering the tasting room to a questionnaire to be tested (Liking+CATA or CATA+Liking), and there was very little chance that panels in four different studies would have been biased the same way. Moreover, the observation of jury details did not reveal any strong difference (Data S2). In each study, both panels were quite similar in age, gender, wine consumption frequency, and subjective knowledge about wine: self-estimation as a beginner, amateur or initiated.

TABLE 1. Statistical indicators for the comparison between Liking+CATA and CATA+Liking approaches.

		Sum of eigenvalues*	Number of panelists to reach 0.05**	Probability to have no effect***
Liking+CATA	Study 1	96.45	21	0.009
	Study 2	97.55		0.264
	Study 3	78.56		0.083
	Study 4	84.66		0.030
CATA+Liking	Study 1	95.47	15	0.000
	Study 2	96.12	40	0.047
	Study 3	90.52		0.041
	Study 4	92.61	33	0.026

*sum of the eigenvalues of the first two axes of the correspondence analyses (CA) between wines and attributes (the higher the better)

** 0.05 is the p-value of the Chi-square test of independence between wines and attributes as a function of the panelist number in the CA (the smaller the better), the written value, e.g. "21" means that 21 panelists were necessary to each 0.05, the absence of value indicates the required number of panelists to reach 0.05 was above the total number panelists in this part of the study

*** measured by ANOVA regarding the effect of wine on the liking score (the smaller the better)

Green backgrounds indicate rather good values, and red backgrounds indicate rather bad values, regarding the quality of the analysis

CONCLUSION

The use of consumers for descriptive analyses is now well accepted (Jaeger *et al.*, 2023) and combining such descriptive analyses with consumer hedonic appreciation has been a common practice over the last decade (Jaeger *et al.*, 2013; Jaeger and Ares, 2015; Alencar *et al.*, 2019; Hayward and McSweeney, 2020; Rinaldi *et al.*, 2021). Our results shed light on differences that may be due to the order of liking and CATA questions. In our cases, the CATA+Liking order led to more discriminant analyses, and this could be addressed in further studies, particularly how this order can impact the dependence between food products and attributes.

ACKNOWLEDGEMENTS

Acknowledgements to all agronomy students from Toulouse and Bordeaux Sciences Agro, who participated to the set-up and panels. Special thanks to Toulouse students: Loïse Vilain, Margaux Renard, Maylis Poifol, Mathias Ourliac, Marie Codromaz and Dauphine Bellanger who helped to develop the seminal set of experiments. Thanks to the Bordeaux staff of the School of Agronomy who participated in the panel of the study 4. Finally, thanks to two anonymous reviewers for their time and comments improving the quality of our manuscript.

REFERENCES

Alencar, N. M. M., Ribeiro, T. G., Barone, B., Barros, A. P. A., Marques, A. T. B., & Behrens, J. H. (2019). Sensory profile and check-all-that-apply (cata) as tools for evaluating and characterizing syrah wines aged with oak chips. *Food Research International*, *124*, 156–164. https://doi.org/10.1016/j.foodres.2018.07.052

Ares, G., & Jaeger, S. (2015). Check-all-that-apply (CATA) questions with consumers in practice: experimental considerations and impact on outcome. In *Elsevier eBooks* (pp. 227–245). https://doi.org/10.1533/9781782422587.2.227

Beaulieu, A., Giraud, V., Magro, P., Nougarede, S., Maza, E., Samson, A., Geffroy, O., & Chervin, C. (2022). Development of the Adapted Pivot Test method for descriptive sensory analyses with young untrained students. *Journal of Sensory Studies, 37*, e12779. https://doi.org/10.1111/joss.12779

Earthy, P. J., MacFie, H. J. H., & Hedderley, D. (1997). Effect of question order on sensory perception and preference in central location trials. *Journal of Sensory Studies, 12*, 215-237. https://doi.org/10.1111/j.1745-459X.1997.tb00064.x

Hayward, L., & McSweeney, M. B. (2020). Investigating caloric values and consumers' perceptions of Nova Scotia rosé wines. Food Research International, 127, 108761. https://doi.org/10.1016/j.foodres.2019.108761

Hough, G., Wakeling, I., Mucci, A., Chambers, E., Gallardo, I. M., & Alves, L. R. (2006). Number of consumers necessary for sensory acceptability tests. *Food Quality and Preference*, *17*(6), 522–526. https://doi.org/10.1016/j.foodqual.2005.07.002

Jaeger, S. R., & Ares, G. (2014). Lack of evidence that concurrent sensory product characterisation using CATA questions bias hedonic scores. *Food Quality and Preference, 35*, 1–5. https://doi.org/10.1016/j.foodqual.2014.01.001

Jaeger, S. R., Chheang, S. L., Jin, D., Ryan, G. S., & Ares, G. (2023). How do CATA questions work? Relationship between likelihood of selecting a term and perceived attribute intensity. *Journal of Sensory Studies, 38*(4). https://doi.org/10.1111/joss.12833

Jaeger, S. R., Giacalone, D., Roigard, C. M., Pineau, B., Vidal, L., Giménez, A., Frøst, M. B., & Ares, G. (2013). Investigation of bias of hedonic scores when co-eliciting product attribute information using CATA questions. *Food Quality and Preference, 30*, 242–249. https://dx.doi.org/10.1016/j.foodqual.2013.06.001

Mahieu, B., Visalli, M., Thomas, A., & Schlich, P. (2020). Free-comment outperformed check-all-that-apply in the sensory characterisation of wines with consumers at home. *Food Quality and Preference, 84*, 103937. https://doi.org/10.1016/j.foodqual.2020.103937

Moskowitz, H. R. (2004). Questionnaire Design. In *Viewpoints and Controversies in Sensory Science and Consumer Product Testing* (pp. 191–208). Food & Nutrition Press, Inc. https://doi.org/10.1002/9780470385128.ch11

Nougarede, S., Diot, A., Maza, E., Samson, A., Olivier-Salvagnac, V., Caillé, S., Geffroy ,O., & Chervin, C. (2023). Comparison of Check-All-That-Apply and Adapted-Pivot-Test methods for wine descriptive analyses with a panel of untrained students. *Journal of Sensory Studies, 38*, e12862 https://doi.org/10.1111/joss.12862

R Core Team (2024). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. https://www.R-project.org/.

Rinaldi, A., Vecchio, R. & Moio, L. (2021). Differences in Astringency Subqualities Evaluated by Consumers and Trained Assessors on Sangiovese Wine Using Check-All-That-Apply (CATA). *Foods, 10*, 218. https://doi.org/10.3390/foods10020218

van Dam, A., Dekker, M., Morales-Castilla, I., Rodríguez, M. A., Wichmann, D., & Baudena, M. (2021). Correspondence analysis, spectral clustering and graph embedding: applications to ecology and economic complexity. *Scientific Reports, 11*, 8926 https://doi.org/10.1038/s41598-021-87971-9.