

Results of the Sensory Scale Comparison Experiment May/June 2007

Aim of the experiment:

The aim of the sensory scale comparison experiment was to determine if traditional hedonic rating or Best-Worst Scaling (BWS) is better suited to measure consumers' sensory wine preferences. Best-Worst Scaling is a new technique gaining much favour in consumer research in Marketing and Economics. We compared the methods regarding their ability to discriminate consumer preferences for different wines as well as their practicality.

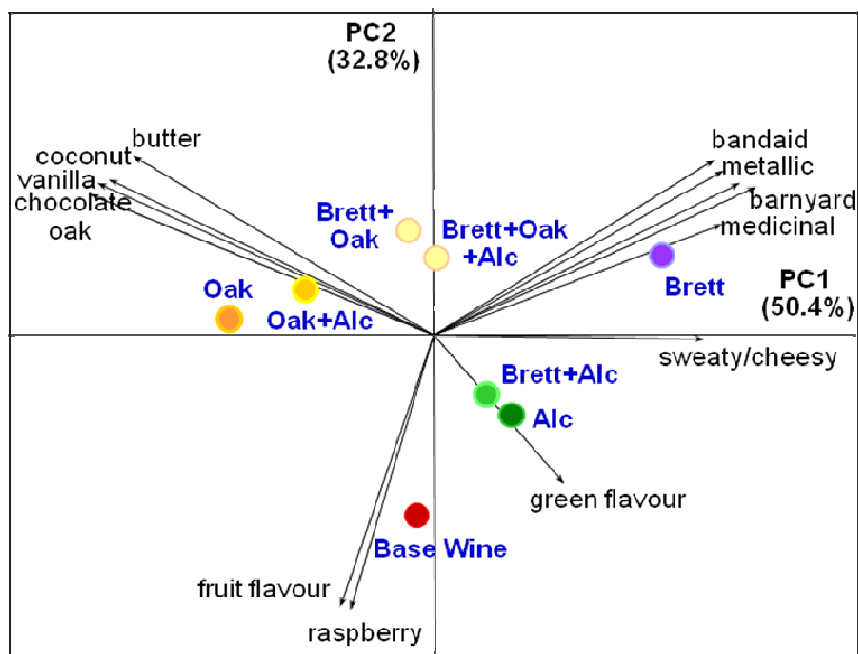
Set-up of the experiment:

1) Capacity pre-test

Because Best-Worst has not been applied in sensory science before, a pre-test was necessary to determine how many wine comparisons an average wine consumer could reliably accomplish in a Best-Worst task. For three choice sets of four red wines per respondent we measured perceived confidence in the decision, perceived ease of the task, time necessary to make a decision, choice consistency, necessary volume of wine tasted and blood alcohol for twenty four regular wine consumers. Results showed that consumers were overstrained by three Best-Worst choice tasks of four wines each. Therefore we had to limit the design of the main scale comparison to two Best-Worst choice tasks of four wines.

2) Scale comparison test

From a neutral fruity Australian multi-region Cabernet Sauvignon base wine we designed eight different wines by adding three components: oak flavour, brettanomyces and alcohol in a full factorial design. A descriptive analysis by the AWRI sensory panel confirmed that most of the eight wines were clearly distinctive in their aroma profiles.



The same 112 regular wine consumers evaluated the eight wines with hedonic rating and Best-Worst on two different days (intra-subject design). For hedonic rating we used a structured nine point liking scale. Because of the tasting capacity constraints we had to split the Best-Worst design into fourteen choice tasks with each respondent choosing the most and least liked wine in two sets of four wines. Thus, one replication of the BW design took seven respondents. Half of the respondents used hedonic rating before Best-Worst and vice versa to control for order effects.

Results:

Over the total sample of all respondents both methods measured different consumer wine preferences. For instance, the base wine was highest liked with hedonic rating but achieved only the fourth highest B-W score in Best-Worst Scaling. The correlation of ranks of wines evaluated with both methods is low ($Rho=0.29$) and statistically insignificant. An analysis of variance (ANOVA) showed a greater ability to discriminate the preferences of the eight wines using hedonic rating. This seems to be due mainly to the problem of each person only being able to make Best-Worst choices on two sets of 8 wines. This limits the amount of data, which can be collected using Best-Worst compared to hedonic rating.

Components	Hedonic rating average liking	ranking HR	B-W score	ranking B-W
Base	5.85	1	2	4
Brett + Oak	5.80	2	9	1
Brett + Oak + Alc	5.74	3	-8	7
Alc.	5.63	4	9	1
Oak + Alc	5.60	5	-5	6
Brett + Alc	5.50	6	-12	8
Oak	5.42	7	4	3
Brett	5.21	8	1	5

A different issue is that not all consumers preferred the same wines; this is referred to as heterogeneity. We used a clustering method to group consumers, who had similar taste preferences. For hedonic rating we found two distinct clusters, which differed in their liking of the eight wines. By taking consumer heterogeneity into account we could differentiate wine preferences much more than was possible over the whole sample. Furthermore, a cluster-wise linear regression revealed that six of the sensory components and their interactions had a significant impact on consumer liking. This shows that different consumers were influenced by their liking of different components in the wines. For example, people in cluster 1 preferred some brett if oak was existent but refused higher alcohol levels when brett or oak were present. Contrary, cluster 2 people preferred the base wine and liked higher levels of alcohol when brett or oak were prevalent.

Because we had to split up the Best-Worst design into seven respondents for each replication, we could not cluster the consumer preferences with Best-Worst preference scores. Best Worst can still be used for attitude and preference measurement, where there is no issue of alcohol consumption.

Wine	Components	Cluster 1 (n=67)	Cluster 2 (n=45)
		Average HR	Average HR
1	Base	4.99	7.13
2	Brett	5.16	5.29
3	Brett + Alc	5.01	6.22
4	Brett + Oak	5.91	5.64
5	Brett + Oak + Alc	4.93	6.96
6	Oak	4.61	6.62
7	Oak + Alc	5.30	6.04
8	Alc.	6.13	4.87

Implications for Future Research/Project Sub-projects:

Based on hedonic rating's higher discriminatory power and its ability to consider respondents' heterogeneity, future tasting experiments will utilise hedonic rating instead of Best-Worst Scaling as the sensory measurement method.

Adelaide, 9 November 2007