Rapid indices for evaluating panel performance in fast paced product development environments

V.Mialon[1], C.Barnagaud[2], S. Ferris[2] | MMR Research Worldwide, United Kingdom[1] /USA[2]
v.mialon@mmr-research.com

Indices summarizing panellists' abilities to discriminate among products, consistency over replicated assessments and agreement with each other provide a simple and rapid tool for performance feedback and tracking

Sensory descriptive analysis is widely used across the food, beverage and personal care industries to compare the sensory characteristics of products and understand consumer responses to formulation parameters. This information is often used to assist in making critical product development and optimization decisions. Checking and monitoring panel performance to ensure the accuracy and reliability of sensory data is therefore essential.

Panel performance monitoring tools

Many statistical, mathematical and graphic methods have been discussed in order to establish the performance of a sensory panel and of its individual panellists [1, 2, 3] on a given set of data. A few specialist packages are available, and some practitioners have developed their own panel performance assessment procedures. Most tools capture, in various ways, panellists' abilities to discriminate among products, consistency over replicated assessments and agreement with each other.

MMR for example uses SAS/STAT® and an Excel based tool, developed in-house, which includes multiple outputs (Figs.1a, 1b & 1c):

- **Overall panel measures** such as 2-way and 3-way ANOVA and significance testing, discrimination ratio and PCA outputs

- **Individual assessor performance measures** with assessor x sample interaction charts, 1 way ANOVA p-values, discrimination ratio, range of scoring across replicates, assessor correlations and more

Fig. 1a - Example of interaction chart on a juice project, showing individual assessor scores for Sweetness. Assessors 2 and 7 scored two juices less sweet than the rest of the panel.
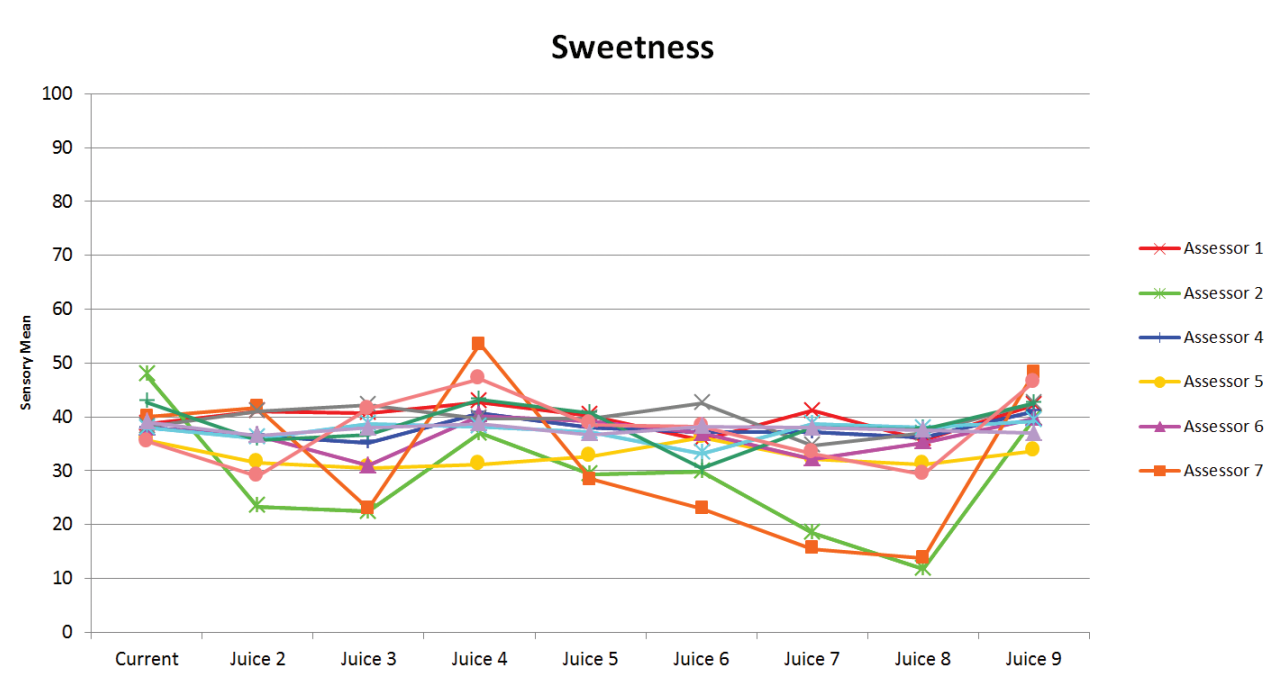

Fig. 1b - Example of repeatability check: table showing the number of times individual panellists' replications differ by more than 20% of the scale ("warning"). The 'cooked' aroma and flavor attributes show the most lack of consistency over replications.

Count of Warning	Assessor Number							
Attribute Number	1	2	3	5	7	10	11	Grand Total
Depth of color					1			1
Amount of brown tint	1	2			1			4
Amount of particles				1				1
Citrus aroma	1							1
Cooked aroma	2			2			3	7
Candy aroma						1	1	1
Zest aroma			1					1
Pith aroma			1					1
Fermenting aroma	2	2					2	6
Chemical aroma				1				1
Vegetable aroma				1				1
Citrus flavor	1				2			3
Cooked flavor	2	1		1	1		2	7
Sweet taste						1	1	2
Fermenting flavor	3	2						5

Fig. 1c - Example of discrimination check showing the level of discrimination from individual assessors on one attribute. Assessor 1 is the most discriminatory on 'Zesty' aroma.
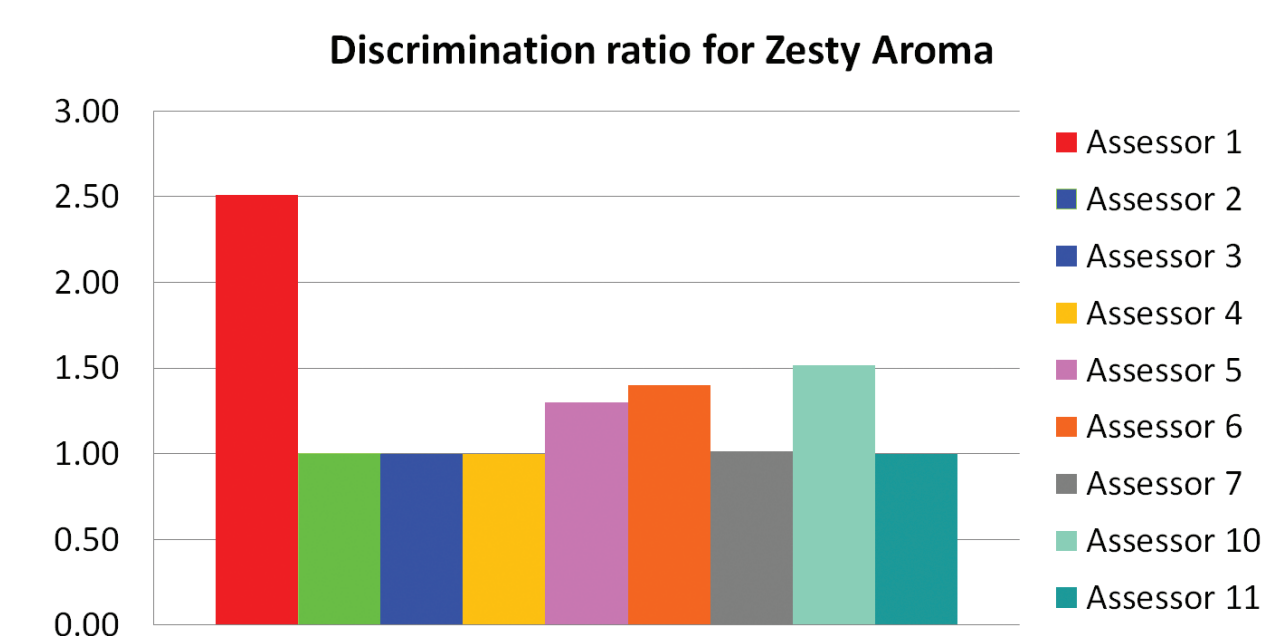
The need for summary indices

Comprehensive information is necessary to fully understand and validate panel performance and to highlight remedial actions, such as specific attributes to focus on. However, such a vast amount of information makes it difficult to have a rapid overall view of the panel and panellist performance. This poses several significant challenges:

- Large data sets make the **overall appreciation of panel performance** a lengthy process

- Too many variables make **tracking performance** difficult

- Outputs are often too complex for **timely and appropriate feedback to panellists**

Simple summary indices are therefore proposed [4, 5] to provide a useful rapid initial read of panel performance, a tool for tracking and comparing panellists and a simple resource to provide feedback.

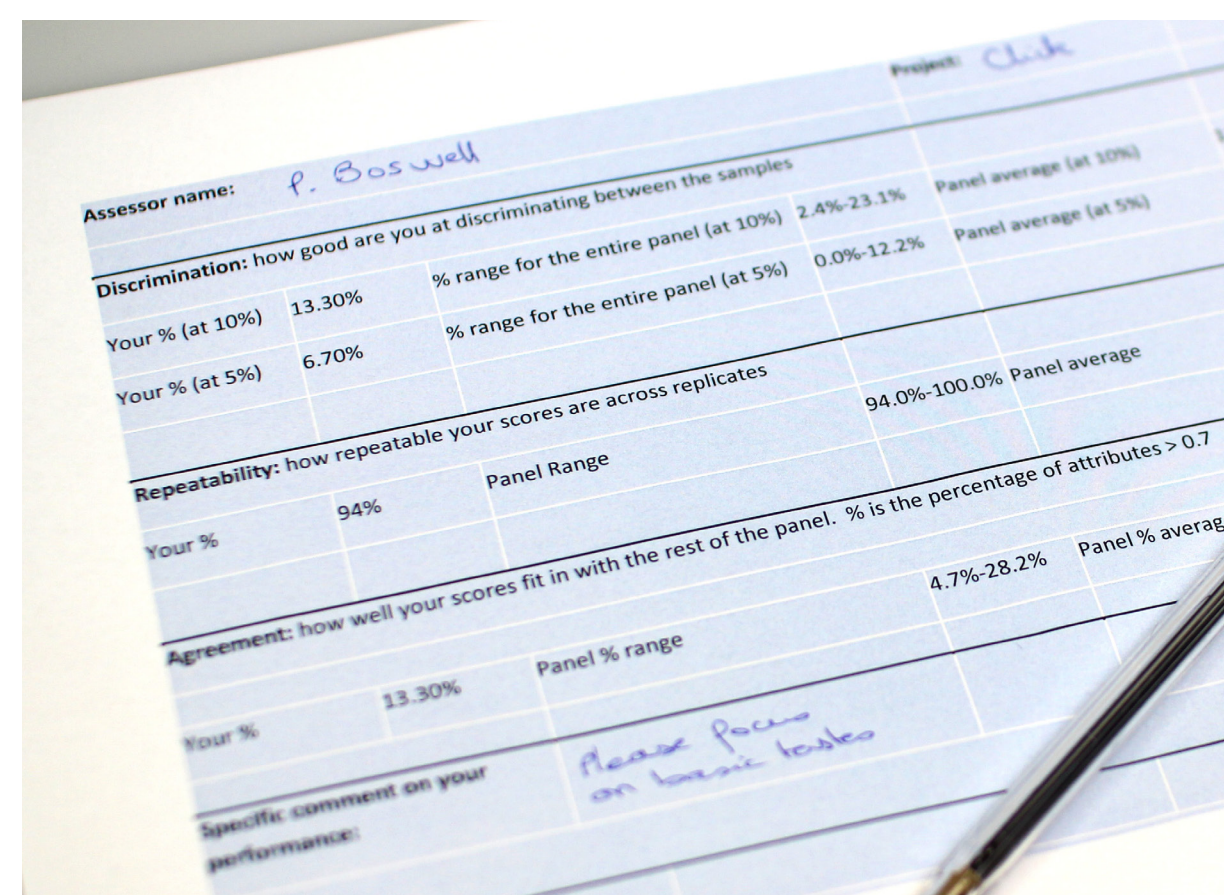### Summary indices of panel performance

In order to capture the key factors of panel performance, summary indices were developed that could be easily extracted from our panel performance tool (Table 1):

- **Discrimination indices** = *% attributes found to discriminate significantly at 90% and 95% l.o.c.*
  Based on the sample effect from a 1-way analysis of variance for each panellist

- **Repeatability index** = *% times a panellist's scores across replicates varies by less than 20% of the scale*
  The number of occasions a panellist's scores differ between replicates by 20 % or more on a 100 point scale across all attributes and products, divided by the total number of product by attribute assessments - % reversed

- **Agreement indices** = *average correlation calculated by correlating each assessor with the mean of the remainder of the panel for each attribute and then calculating the average across attributes*
  % attributes > 0.7 is a second expression of this (counts the number of times each panellist has a correlation of or greater than 0.7)

For all measures, the higher the index the better the performance, making the measures easily interpretable for non-technical audiences, such as typical sensory panellists. In addition to providing an overall quick read of the whole panel and of individual panellists' performances, the panel's range and average measures also allow each panellist to compare their own performance with the rest of the panel.

Feedback forms (Fig.2), are created for each assessor automatically, enabling faster and more efficient feedback. These results, together with more specific feedback from the panel leader, provide a simple report to inform the panellists on their specific performance and future re-training needs.

Fig.2 - Automatic feedback form capturing an individual assessor performance indices against the panel ranges and averages. Space is allowed for the panel leader to add specific areas for improvement.

These indices are also used to compare the results of 2 panels enabling evaluation of the proficiency of a new vs. an established panel.

Simple indices that summarize panellists' abilities to discriminate, their repeatability and their agreement with the rest of the panel can be easily extracted from typical panel performance tools. They do not replace more comprehensive outputs but complement them.

The addition of such indices to comprehensive panel performance routines brings efficiency in the sensory evaluation process. These could be further improved by linking the summary results to the more detailed diagnostics, in an interactive framework, and providing more graphical outputs for added impact.

Table 1 - Panel performance indices on a beverage study. Assessor 10, who recently joined the panel shows a low level of discrimination and agreement with the rest of the panel.

	Assessor 1	Assessor 2	Assessor 3	Assessor 4	Assessor 5	Assessor 6	Assessor 7	Assessor 10	Assessor 11	Panel Average	Panel Range
Discrimination indices:											
% attributes significant at 90%	150%	9.8%	9.3%	10.9%	17.1%	23.1%	5.0%	2.4%	13.3%	11.8%	2.4% to 23.1%
% attributes significant at 95%	7.5%	7.3%	7.0%	8.7%	12.2%	7.7%	2.5%	0.0%	6.7%	6.6%	0.0% to 12.2%
Repeatability index:											
% Repeatability with no warning	95.0%	95.7%	97.9%	100.0%	98.6%	100.0%	97.9%	98.9%	94.0%	97.6%	94.0% to 100.0%
Agreement indices:											
Average Correlation to the rest of the panel	0.22	0.18	0.06	0.20	0.09	0.42	0.08	-0.13	0.13	0.14	-0.13 to 0.42
% attributes with correlation coefficient > 0.7	22.5%	24.4%	4.7%	23.9%	9.8%	28.2%	5.0%	4.8%	13.3%	15.2%	4.7% to 28.2%

References
1. Lea, P., Rødbotten, M. & Næs, T. (1995). Measuring validity in sensory analysis. Food Quality and Preference, 6, 321–326
2. McEwan, J.A. (2002). Proficiency testing for sensory profile panels: measuring panel performance. Food Quality and Preference, 13, 181–190
3. Kermit, M. and Lengard, V. (2006). Assessing the performance of a sensory panel – panellist monitoring and tracking. Journal of Chemometrics, 19, 154–161
4. Schlich, P., Pineau, N., Brajon, D., & Qannari, E.M. (2004). Multivariate control of assessor performances. In proceedings : 7th Sensometrics Meeting Davis, California, US
5. Tomic, O., Forde, C., Delahunty, C. & Næs, T. (2012). Performance indices in descriptive sensory analysis – a complimentary screening tool for assessor and panel performance. Food Quality and Preference. Accepted manuscript.