Assessing the Impact of Panel Size on Sensory Data Reliability

Shun Xie, Chloe McGurk, Ellie King, Valerie Mialon

Shun Xie, MMR Research Worldwide, Inc. (US) | Chloe McGurk - MMR Research Worldwide, Inc. (US), Ellie King - MMR Research Worldwide, Inc. (US) | Valerie Mialon - MMR Research Worldwide, Inc. (US)

BACKGROUND & OBJECTIVE

Determining the optimal number of trained panelists for sensory studies is critical for balancing data reliability with resource constraints. While larger panels enhance data robustness, they also raise costs, making it essential to identify a statistically sound minimum.

While previous studies recommend 8–12 panelists for descriptive analysis (Heymann et al., 2012; Campo et al., 2010; Tesfaye et al., 2010; Moussaoui & Varela, 2010), most focus exclusively on sensory descriptive analysis, without looking at additional aspects such as product clustering. Moreover, few have examined how panel size influences outcomes across varying levels of product similarity in real-world settings.

This study aims to empirically examine how varying panel sizes influence sensory outcomes across multiple dimensions, acknowledging the potential variability in product differentiation, panel location and product category. The study focused on trained panels with several years of experience in the relevant categories.

DATA

REGION	PRODUCT CATEGORY	# PANELISTS	# ATTRIBUTE	# PRODUCTS	PRODUCT DIFFERENCES
	Beverage	11	21	4	Low
	Household Care	12	66	16	Low
	Beverage	8	24	5	Medium
	Personal Care	12	41	4	Medium
	Dairy	14	46	4	Medium
	Personal Care	10	63	9	Medium
	Snacking	9	19	3	Medium
	Personal Care	10	35	6	High
	Dairy	9	24	10	High
	Dairy	11	30	4	High
	Beverage	8	67	12	High
	Beverage	10	62	10	High
	Beverage	10	91	10	High

METHOD

Datasets

All datasets were collected using MMR DA, 2 reps, from MMR's Sensory Science Centers across 4 regions with multiple product categories and with varying numbers of products and panelists. Product differences were categorized based on % of significantly different attributes (95% LOC).

Product Difference Definition

Low	Medium	High
% sig attribute < 50%	% sig attribute ≥ 50% & < 75%	% sig attribute ≥ 75%

Panel Size Impact Analysis

01 Start with full panel size (n)

02 Iteratively reduce the panel size to k panelists, where k ranges from n−1 down to 3, sampling up to 100 random combinations for each k.

03 Compare reduced data vs. full data on 3 statistical methods and metrics

Statistical methods and metrics to assess the effect of panel size

ANOVA: Proportional z test conducted on change in proportion of significant attributes based on 95% LOC

LSD: Change in % of product pairs exceeding 95% LSD

PCA Clustering: Using Rand Index to quantify changes in product clustering based on sensory PCA maps, with a value of 1 indicating perfect agreement with the full panel clustering.

REPEATEDLY REMOVE N-K PANELIST

Full Panelist Data (n) → ANOVA / LSD / PCA Clustering → Reduced Panelist Data (k)

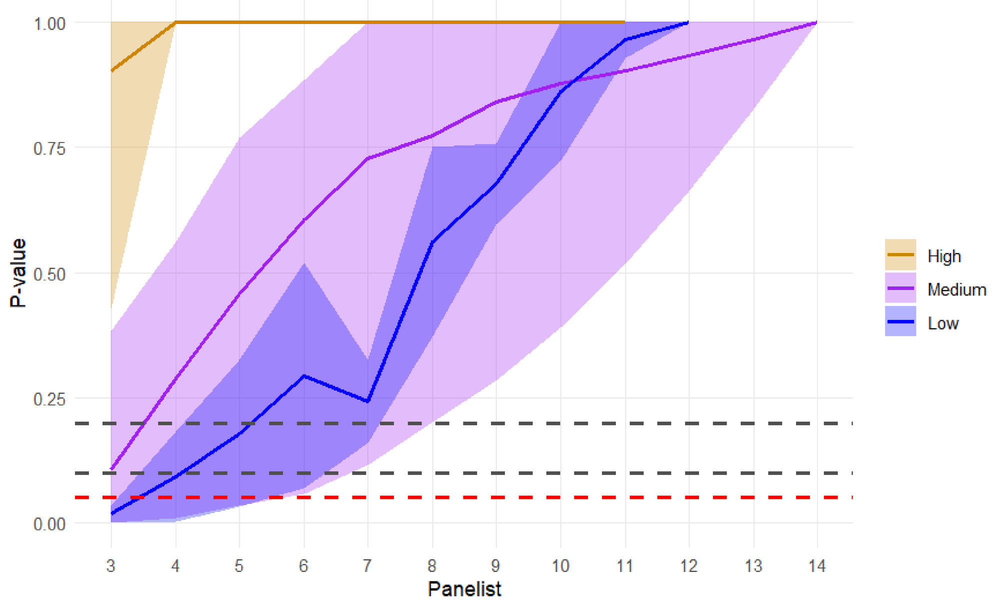
RESULT

ANOVA Chart

Proportional z test on change in proportion of significant attributes based on 95% LOC

Based on the lower interval band, the proportion of significant attributes changes significantly at p=0.2, when fewer than 8 panelists are used.

6 panelists represent a bare minimum threshold where significant changes are observed at p=0.05, based on the lower band.

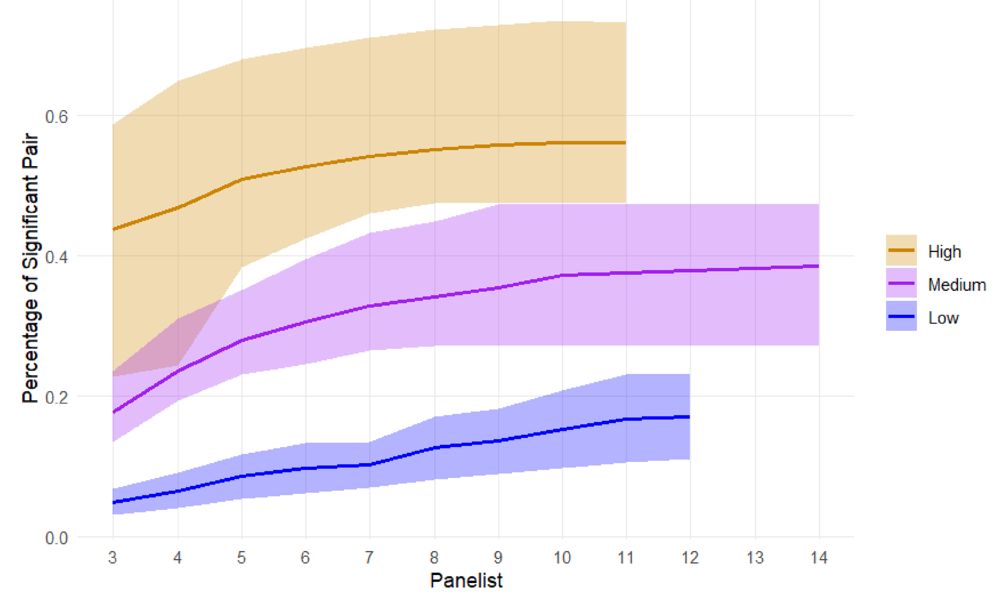


LSD Chart

Change in % of product pairs exceeding 95% LSD

Clear differentiation in significant product pairs is observed across product difference categories, with more pairs exceeding LSD in **Projects with High Product Difference** than in **Projects with Low Product Difference.**

Continuous decline in significant product pairs, with no distinct drop in significant pairs observed.

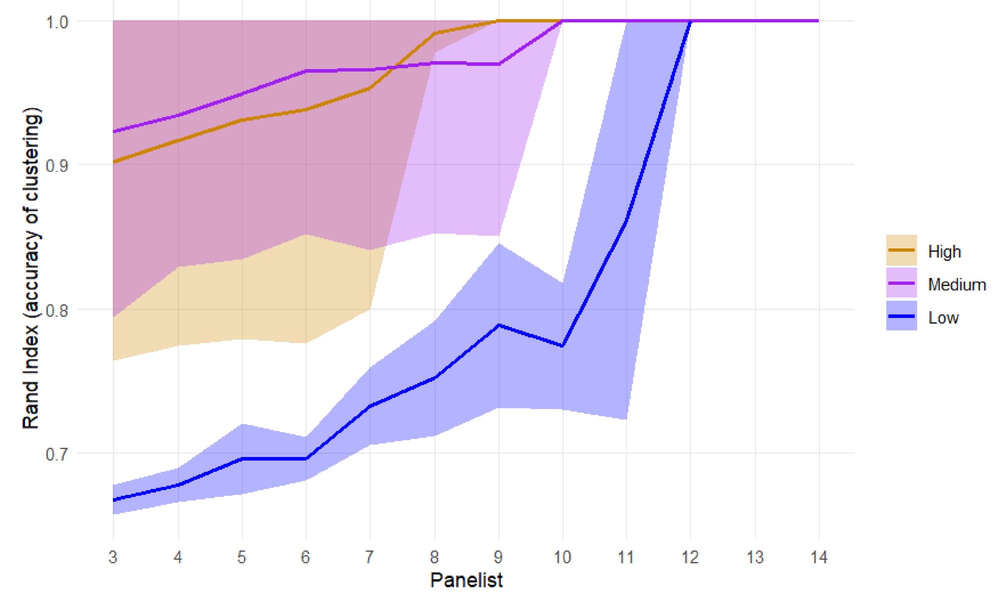


PCA Clustering Chart

Rand Index shows how similar the clusters are to the full panel (value of 1 = perfect agreement)

For **Projects with Low Product Difference**, removing a single panelist can change the clustering pattern observed on the first 2 dimensions of the PCA map.

A higher Rand Index indicates greater agreement between the reduced and full-panel clustering.



CONCLUSION

- In general, a minimum of 8 panelists ensures statistical rigor and produces results consistent with those obtained from the full panel across all projects.
  - The finding aligns with the proposed panelist range by Heymann (Heymann et al., 2012).
- Product differentiation strongly influences the impact of panel size on the sensory results:
  - If we have **High Product Difference**, there is opportunities to reduce the panel size
  - If there is **Low Product Difference**, clustering becomes unstable with even one panelist removed, indicating a need for a larger panel to ensure result consistency
- Data showed no meaningful differences across number of products and regions, aligning with prior research on cross-region panel consistency.

KEY TAKEAWAYS AND CONSIDERATIONS

- This study highlights key considerations for future applications, emphasizing that product differences have the greatest impact on appropriate panel size.
- The results provide general guidelines for a minimum number of panelists for general projects, while acknowledging the need and opportunity for flexibility in different research contexts:
  - Studies which demands high data robustness, such as claim substantiation do require a larger panel size (10-12)
  - Studies comparing product with **Low Product Difference**, such as data used to create precise predictive models (e.g. on liking) or reformulation, warrant using at least 8 panelists
  - Studies with **High Product Difference**, such as category sensory landscaping or benchmarking give the opportunity to reduce panel size to 6-8 panelists, and potentially even lower
- No single methods and metrics that we used offers a definitive answer on optimal panel size; all analyses provide complementary insights and should be considered together for robust conclusions.

REFERENCE

- Heymann, H., Machado, B., Torri, L. and Robinson, A.L., 2012. How many judges should one use for sensory descriptive analysis?. Journal of Sensory Studies, 27(2), pp.111-122.
- Campo, E., Ballester, J., Langlois, J., Dacremont, C. and Valentin, D., 2010. Comparison of conventional descriptive analysis and a citation frequency-based descriptive method for odor profiling: An application to Burgundy Pinot noir wines. Food quality and preference, 21(1), pp.44-55.
- Tesfaye, W., Morales, M.L., Callejón, R.M., Cerezo, A.B., González, A.G., Garcia-Parrilla, M.C. and Troncoso, A.M., 2010. Descriptive sensory analysis of wine vinegar: tasting procedure and reliability of new attributes. Journal of Sensory Studies, 25(2), pp.216-230.
- Moussaoui, K.A. and Varela, P., 2010. Exploring consumer product profiling techniques and their linkage to a quantitative descriptive analysis. Food Quality and Preference, 21(8), pp.1088-1099