Panel and Panelist Performance Strategy in Discrimination Testing

Katie E. Osdoba, PhD

kosdoba@sensoryspectrum.com

sensoryspectrum

Sensometrics 2020 – Sense the Energy

www.sensometrics2020.com

What will you find in this presentation?

Discrimination programs in the CPG industry aid product **maintenance**, **quality control**, and **shelf life** initiatives, and superior results depend on a capable, high-performing panel.

On the following slides you will see key performance indicators for discrimination panels that participate in a variety of test types – **overall** and **attribute-specific** tests, difference and similarity tests.

The panel performance assessment techniques focus on **forced-choice methods** such as triangle, tetrad, and 2-AFC.

Techniques are designed to assess panels' and panelists' performance in terms of **repeatability**, **discriminability**, and **validity**.

These techniques are designed to assess panel performance **over a period of time** – e.g. monthly or quarterly.

How will it benefit your organization?

Panel performance monitoring can be applied to:

- Frequent data quality checks
- Establishment of performance criteria (cutoffs/action standards)
- Deeper understanding of panel capability
- Bringing panel back after a hiatus (e.g. COVID) – are they performing as normal?

Panel performance monitoring is only one piece of a Panel Quality Maintenance Program*, which may also include:

- Regular re-validation studies
- Training, orientation, & practice
- Community building & fun

**Email kosdoba@sensoryspectrum.com for a copy of a 2017 poster on Panel Quality Maintenance, presented at the 12th Pangborn Symposium.*

Key Performance Indicators

Overall Panel Performance

Repeatability

Questions

- When the panel completes multiple reps within a test, how often do the reps return the same results?
- Or, when the panel repeats a test, do they replicate their results? *Especially useful for assessing performance over time.*

Method

- Split test data by rep and calculate results (relax significance criteria to account for lower test power with smaller N, i.e. $p < 0.10$ instead of $p < 0.05$).
- Calculate the % of tests for which panel successfully replicates results (rep to rep or test to test).
- For attribute-specific methods, can assess performance by attribute or across all attributes.

Example

- The panel successfully replicates 50% of the tests in the table below.

	Test 1	Test 2	Test 3	Test 4
Rep 1 (p-value)	0.21	0.47	0.038	0.0097
Rep 2 (p-value)	0.067	0.74	0.075	0.29
Results Agree?	N	Y	Y	N

Discrimination & Validity

Questions

- Does the panel adequately differentiate products that are expected to be different?
- Does the panel give results that are aligned with expectations based on prior knowledge or other data types?
- *These measures are especially useful for attribute-specific data.*

Method

- Compare directional differences (in attributes, i.e. which sample is higher in X?) among data sources.
- Compare significant overall differences in discrimination tests to meaningful differences from other data sources.
- If available, compare perceptual effect sizes (DOD scores, scalar differences, deltas/d-primes, etc.) to assess alignment in magnitude of difference.

Example

- The panel is aligned with expected results in 75% of the tests below.

	Test 1	Test 2	Test 3	Test 4
Discrimination Result	Similar	Similar	Different	Different
Effect Size	Below Threshold	Threshold	Small	Large
Expected Result	Similar	Different	Different	Different
Expected Effect Size	Below Threshold	Small	Small	Large
Results Agree?	Y	N	Y	Y

Additional measure of overall panel validity:

- Can the panel correctly identify blind control pairs? When the panel evaluates a blind control pair (both samples are the same), they should consistently show effect sizes (d-prime) below a certain threshold for similarity, to make sure that the panel is not erroneously finding differences between identical samples. *This measure is especially important for panels that primarily conduct similarity tests.*

Key Performance Indicators

Individual Panelist Performance

	Repeatability	Discrimination & Validity
Questions	• Can the panelist give the same results across multiple replications? *(regardless of whether the panelist's result agrees with overall panel result)*	• When the overall panel successfully differentiates products, can the individual panelist as well? • Do individual panelists give results that are aligned with expectations? *This measure is especially useful for attribute-specific data.* • Is there evidence that an individual panelist has elevated sensitivity?
Method	• Calculate the % of tests for which a panelist successfully replicates results (2 out of 2 reps or 3 out of 3 reps). • For attribute-specific methods, can assess performance by attribute or across all attributes.	• Calculate % responses in agreement with overall panel results (when the panel showed a difference). • Determine if panelists are responding correctly at a rate greater than chance. • Compare overall or directional responses to expected results from other data sources. • For attribute-specific methods, can assess performance by attribute or across all attributes. • High Sensitivity: Calculate % responses in agreement with panel (at rate greater than chance) when panel concluded for similarity.
Example	• Five out of seven panelists successfully replicate at a rate higher than 50%.	• Four out of seven panelists are discriminating (in alignment with panel) at a rate greater than chance. One panelist has been identified as having high sensitivity.

	Percentage of Tests Replicated
Panelist 1	81%
Panelist 2	55%
Panelist 3	65%
Panelist 4	25%
Panelist 5	56%
Panelist 6	59%
Panelist 7	50%

	Percentage of Tests Successfully Discriminated	High Sensitivity?
Panelist 1	30%	N
Panelist 2	57%	N
Panelist 3	36%	N
Panelist 4	47%	N
Panelist 5	48%	N
Panelist 6	56%	Y
Panelist 7	45%	N

Example: Panel Performance Results

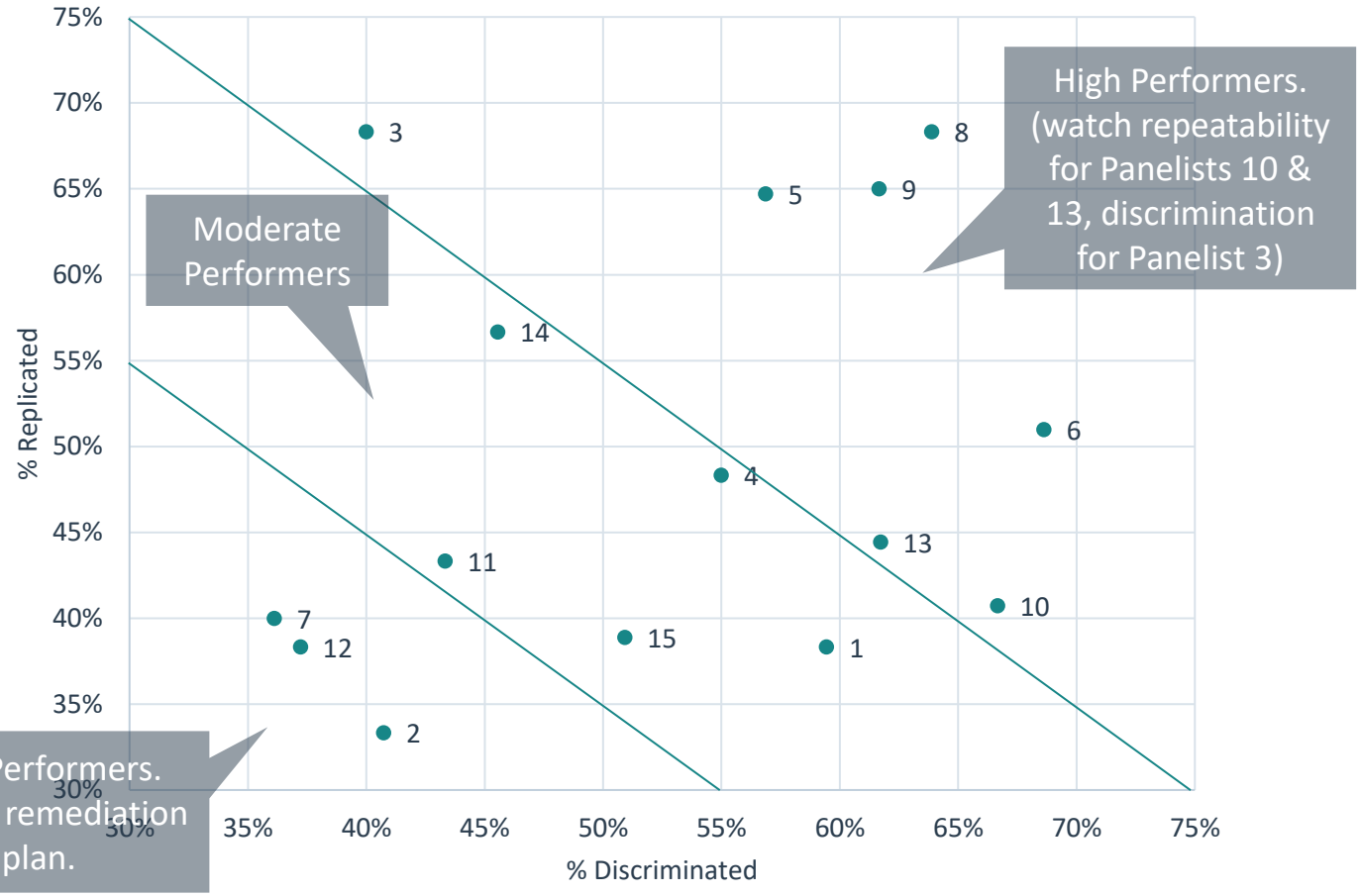
Overall Panel Repeatability (by attribute)

- 2-AFC methodology, 11 attributes evaluated per test
- N = 15 panelists x 2-3 reps per test
- 12 Tests

	Percentage of Tests Replicated
Attribute A	75%
Attribute B	67%
Attribute C	83%
Attribute D	17%
Attribute E	50%
Attribute F	50%
Attribute G	83%
Attribute H	67%
Attribute I	58%
Attribute J	100%
Attribute K	92%

Panel may need additional training/ practice with Attribute D

Individual Panelist Performance



For this panel, no measure of discrimination/validity available.