



Campden & Chorleywood Food
Research Association Group



www.esn-network.com

An Overview of Proficiency Testing in Sensory Analysis

Chantal Gilbert



Outline

- What is proficiency testing?
- Adapting proficiency testing to Sensory
- Measurement criteria for panel proficiency
 - Ranking
 - Descriptive Profiling
- Example - Proficiency test for Ranking
- Ongoing Research
- Proficiency Testing Rounds

Sensory Panel Performance

- Panel performance is an important part of data validation
- After executing a sensory test, can analyse **panel performance**:
 - Did my panel discriminate?
 - Could my panel replicate the results?
 - Are my panellists in agreement?
 - Has my panel produced a ‘sensible’ product space?

But...

- How does the panel compare to other similar panels?
- Does the panel produce reliable results?
- Are the results ‘externally valid’?

Need for Proficiency Testing

There is an increased need for laboratories to demonstrate their **reliability** and **competency** in performing sensory evaluation techniques.

What is proficiency testing?

- Proficiency testing is the use of inter-laboratory test comparisons to determine the performance of individual laboratories.
- Used to monitor **consistency** and **comparability** of a lab's test data with other similar labs using the same test materials.
- Proficiency tests are well established in many scientific disciplines - used in particular with instrumental methods (e.g. chemical analysis).
- International guidelines now exist for the design, administration, and interpretation of sensory proficiency tests:
 - *International guidelines for proficiency testing in sensory analysis. CCFRA Guidelines 35, 2001. Edited by DH Lyon.*

Issues unique to Sensory

- ‘Instrument’ = sensory **panel**
- Nature of data
 - response, not amount
 - relative, not absolute
- Absence of a ‘True value’
 - Instead, use of an ‘Expected value’
- Results dependent on method
- Results dependent on product

Special considerations

Test materials

- Must be safe, stable, and homogeneous.
- Suitable range of sensory differences between products (not too difficult, not too easy!)



Procedures

- All panels judged against a common standard.
- Performance is based on the results of the in-house methods normally adopted.



Measurement criteria for panel proficiency

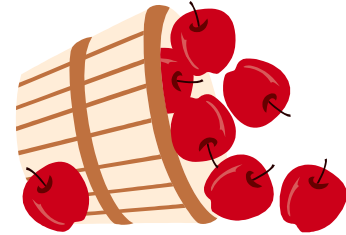
- Performance criteria - Stepwise procedure
- Exact criteria dependent on test: Ranking or Profiling
- **Expected results** are determined using results of ‘Validation’ panels:
 - The expected scores are determined for a variety of performance criteria (discrimination, panel agreement, correlation etc.)
 - Expected overall grade (range) is established
- **Trial results** are established: results of participating panels are graded against expected criteria

Example

Adapted from 'Profisens' Apple Juice Ranking

1. Expected results

Samples - Apple Juice



- Five samples of apple juice, spiked with different levels of glucose and fructose blends

Sample	Sugar Blend	
	Glucose	Fructose
1	75%	25%
2	63%	37%
3	50%	50%
4	37%	63%
5	25%	75%

Performance Criteria for Ranking

STEP 1

Calculate the expected mean rank for each sample and correlate with panel mean ranks

Correlation with expected mean rank (p-value)

STEP 2a

Calculate the significance level using Friedman rank test for sample differences

Expected significance level

Step 2b

Calculate assessors' agreement with respect to rank order

Expected concordance

STEP 3

Calculate number of significantly different pairs of samples

Expected sample differences

STEP 4

Establish the expected performance of the panel (sum of expected scores)

Validation Panels

- For the example shown, data from two validation panels are used to set the expected results.
- For each criteria, the performance of the validation panels was determined, and used to establish the ‘expected’ results.

Expected Results

	Laboratory A		Laboratory I1				
	Rep 1	Rep 2	Rep 1	Rep 2	Expected Results - Criteria		
Step 1: Corr p-value	0.000	0.000	0.002	0.026	Score 0 if $p > 0.10$		
					Score 1 if $p \leq 0.10$		'expected result'
Step 2a: Discrim p-value	0.000	0.000	0.000	0.002	Score 0 if $p > 0.05$		
					Score 1 if $p \leq 0.05$		
					Score 2 if $p \leq 0.01$		'expected result'
					Score 3 if $p \leq 0.001$		
Step 2b: Concordance	0.91	0.90	0.65	0.53	Score 0 if $W < 0.60$		
					Score 1 if $W \geq 0.60$		
					Score 2 if $W \geq 0.70$		
					Score 3 if $W \geq 0.80$		'expected result'
					Score 4 if $W \geq 0.90$		
Step 3: No. of sig. diff	10	9	6	6	Score 0 if 0 or 1 sig. diff		
					Score 1 if 2 or 3 sig. diff		
					Score 2 if 4 or 5 sig. diff		
					Score 3 if 6 or 7 sig. diff		
					Score 4 if 8 sig. diff		'expected result'
					Score 5 if 9 sig. diff		
					Score 6 if 10 sig. diff		

Expected Results

	Laboratory A		Laboratory I1				
	Rep 1	Rep 2	Rep 1	Rep 2	Expected Results - Criteria		
Step 1:	0.000	0.000	0.002	0.026	Score 0	if $p > 0.10$	
Corr p-value					Score 1	if $p \leq 0.10$	'expected result'
Step 2a:	0.000	0.000	0.000	0.002	Score 0	if $p > 0.05$	'expected result'
Discrim p-value					Score 1	if $p \leq 0.05$	
					Score 2	if $p \leq 0.01$	
					Score 3	if $p \leq 0.001$	
Step 2b:	0.91	0.90	0.65	0.53	Score 0	if $W < 0.60$	
Concordance					Score 1	if $W \geq 0.60$	
					Score 2	if $W \geq 0.70$	
					Score 3	if $W \geq 0.80$	'expected result'
					Score 4	if $W \geq 0.90$	
Step 3:	10	9	6	6	Score 0	if 0 or 1 sig. diff	
No. of sig. diff					Score 1	if 2 or 3 sig. diff	
					Score 2	if 4 or 5 sig. diff	
					Score 3	if 6 or 7 sig. diff	
					Score 4	if 8 sig. diff	'expected result'
					Score 5	if 9 sig. diff	
					Score 6	if 10 sig. diff	

Expected Results

	Laboratory A		Laboratory I1				
	Rep 1	Rep 2	Rep 1	Rep 2	Expected Results - Criteria		
Step 1:	0.000	0.000	0.002	0.026	Score 0	if $p > 0.10$	
Corr p-value					Score 1	if $p \leq 0.10$	'expected result'
Step 2a:	0.000	0.000	0.000	0.002	Score 0	if $p > 0.05$	
Discrim p-value					Score 1	if $p \leq 0.05$	
					Score 2	if $p \leq 0.01$	'expected result'
					Score 3	if $p \leq 0.001$	
Step 2b:	0.91	0.90	0.65	0.53	Score 0	if $W < 0.60$	
Concordance					Score 1	if $W \geq 0.60$	
					Score 2	if $W \geq 0.70$	
					Score 3	if $W \geq 0.80$	'expected result'
					Score 4	if $W \geq 0.90$	
Step 3:	10	9	6	6	Score 0	if 0 or 1 sig. diff	
No. of sig. diff					Score 1	if 2 or 3 sig. diff	
					Score 2	if 4 or 5 sig. diff	
					Score 3	if 6 or 7 sig. diff	
					Score 4	if 8 sig. diff	'expected result'
					Score 5	if 9 sig. diff	
					Score 6	if 10 sig. diff	

Expected Results

	Laboratory A		Laboratory I1				
	Rep 1	Rep 2	Rep 1	Rep 2	Expected Results - Criteria		
Step 1:	0.000	0.000	0.002	0.026	Score 0	if $p > 0.10$	
Corr p-value					Score 1	if $p \leq 0.10$	'expected result'
Step 2a:	0.000	0.000	0.000	0.002	Score 0	if $p > 0.05$	
Discrim p-value					Score 1	if $p \leq 0.05$	
					Score 2	if $p \leq 0.01$	'expected result'
					Score 3	if $p \leq 0.001$	
Step 2b:	0.91	0.90	0.65	0.53	Score 0	if $W < 0.60$	
Concordance					Score 1	if $W \geq 0.60$	
					Score 2	if $W \geq 0.70$	
					Score 3	if $W \geq 0.80$	'expected result'
					Score 4	if $W \geq 0.90$	
Step 3:	10	9	6	6	Score 0	if 0 or 1 sig. diff	'expected result'
No. of sig. diff					Score 1	if 2 or 3 sig. diff	
					Score 2	if 4 or 5 sig. diff	
					Score 3	if 6 or 7 sig. diff	
					Score 4	if 8 sig. diff	
					Score 5	if 9 sig. diff	
					Score 6	if 10 sig. diff	

Overall Performance Criteria

- Sum of expected scores
- Allowed to score 1 less on one of the criteria

Score 10.1 – 14	‘Better than expected’
Score 9 - 10	‘Expected result’
Score < 9	‘Less than expected’

Example con't

Adapted from 'Profisens' Apple Juice Ranking

2. Trial results

Participant Results

- 14 laboratories
- For each criteria (Steps 1 - 4), performance results calculated for each panel and compared to expected scores

Participant Results

Step 1: Correlation p-value

Example

Score 0	if $p > 0.10$	
Score 1	if $p \leq 0.10$	'expected result'

Panel C - Rep 1

Sample	Panel C mean rank
1	4.9
2	4.1
3	2.8
4	2.2
5	1

Sample	Expected mean rank
1	4.9
2	4
3	3
4	1.8
5	1.3



Correlation

p-value: < 0.001

Score = 1

Participant Results

Step 2a: Discrimination p-value

Example

Panel C - Rep 1 (Raw Data)

Score 0	if $p > 0.05$	
Score 1	if $p \leq 0.05$	
Score 2	if $p \leq 0.01$	'expected result'
Score 3	if $p \leq 0.001$	

Assessor	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
1	4	5	3	2	1
2	5	4	3	2	1
3	5	4	3	2	1
4	5	4	2	3	1
5	5	4	2	3	1
6	5	4	3	2	1
7	5	4	3	2	1
8	5	4	3	2	1



Friedman Analysis of Rank

p-value: < 0.001

Score = 3

Participant Results

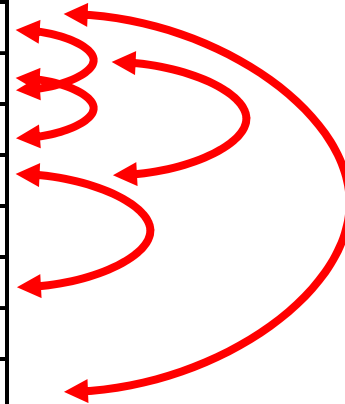
Step 2b: Concordance

Example

**Panel C - Rep 1
(Raw Data)**

Score 0	if $W < 0.60$	
Score 1	if $W \geq 0.60$	
Score 2	if $W \geq 0.70$	
Score 3	if $W \geq 0.80$	'expected result'
Score 4	if $W \geq 0.90$	

Assessor	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
1	4	5	3	2	1
2	5	4	3	2	1
3	5	4	3	2	1
4	5	4	2	3	1
5	5	4	2	3	1
6	5	4	3	2	1
7	5	4	3	2	1
8	5	4	3	2	1



Concordance - RV Coefficient

W: 0.95

Score = 4

Participant Results

Step 3: No. of significant pairs

Example

Panel C - Rep 1 (Raw Data)

Assessor	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
1	4	5	3	2	1
2	5	4	3	2	1
3	5	4	3	2	1
4	5	4	2	3	1
5	5	4	2	3	1
6	5	4	3	2	1
7	5	4	3	2	1
8	5	4	3	2	1



Sample	Mean Rank	Groups
1	4.9	a
2	4.1	b
3	2.8	c
4	2.2	d
5	1	e
MC - 5%	0.34	

Number of Significantly Different Pairs: 10

Score = 6

Score 0	if 0 or 1 sig. diff	
Score 1	if 2 or 3 sig. diff	
Score 2	if 4 or 5 sig. diff	
Score 3	if 6 or 7 sig. diff	
Score 4	if 8 sig. diff	'expected result'
Score 5	if 9 sig. diff	
Score 6	if 10 sig. diff	

Participant Results

Step 4: Overall Grade

Rep 1 - Performance scores by panel

Panel	Step 1	Step 2a	Step 2b	Step 3	Step 4 - Overall
A	1	3	4	6	14
B	1	3	0	2	6
C	1	3	4	6	14
D	1	3	1	3	8
E	1	3	0	3	7
F	1	2	0	2	5
G	0	0	0	0	0
H	1	2	2	3	8
I	1	3	0	2	6
J	1	3	4	6	14
F1	1	2	0	2	5
I1	1	3	1	3	8
J1	1	3	0	3	7
K1	1	3	0	3	7

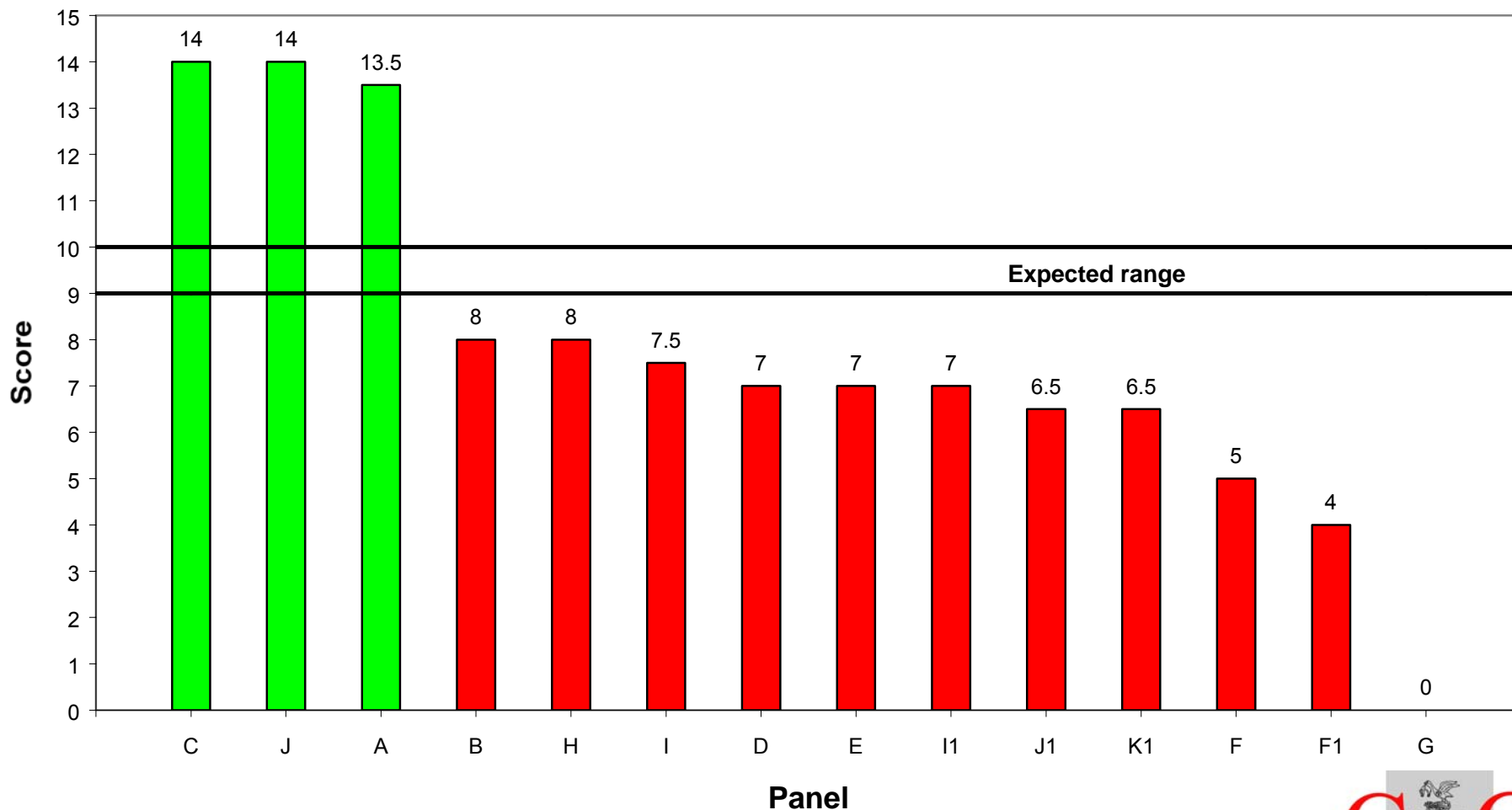
Participant Results

Step 4: Overall Grade

Panel	Replicate 1	Replicate 2	Average	Grade *
A	14	13	13.5	> expected
B	6	10	8.0	< expected
C	14	14	14.0	> expected
D	8	6	7.0	< expected
E	7	7	7.0	< expected
F	5	5	5.0	< expected
G	0	0	0.0	< expected
H	8	8	8.0	< expected
I	6	9	7.5	< expected
J	14	14	14.0	> expected
F1	5	3	4.0	< expected
I1	8	6	7.0	< expected
J1	7	6	6.5	< expected
K1	7	6	6.5	< expected
Expected			9-10	

Participant Results - Overall

Performance Results - Apple Juice RANKING (Profisens)



* Expected result: score 9-10

Further Analysis

- During the trial, participants fill out a **checklist** to record information about procedures used, deviations from instructions, and information about the panel.
- Overall performance results are correlated with information provided in the checklist.
- This helps to establish areas for improvement or further training - Participants receive positive feedback from results of trial.



Performance criteria for Descriptive Profiling

- Similar step-wise approach
- Performance criteria
 - Number of significant sensory dimensions (GPA)
 - Number of significantly different pairs
 - Agreement with expected sensory map
 - Agreement between assessors

Ongoing Research

- Development of samples for proficiency testing
 - Variety of product categories
 - ‘Engineered’ products
- Refining of performance criteria - incorporating ‘known’ information about product attributes into performance criteria

Benefits of proficiency testing

- Helps to demonstrate the **reliability** of sensory panel results, which is beneficial towards **accreditation** (e.g. through UKAS & ISO).
- Participants are shown their performance for each of the performance criteria to help pinpoint where improvement or further training is needed.
- Participation can foster **confidence** in the results of a sensory panel, for management and clients.
- Proficiency Testing helps to **encourage improvement**.

Proficiency Testing Rounds

- Clear benefits of participation
- Two separate proficiency tests available
 - Ranking
 - Descriptive profiling
- Costs
 - Dependent on the type of proficiency test, and the number of participants
 - Usually 2000-3000 Euro
- If you would like to take part, or would like to receive information about future proficiency tests, please contact...
Chantal Gilbert (c.gilbert@campden.co.uk)

Acknowledgements

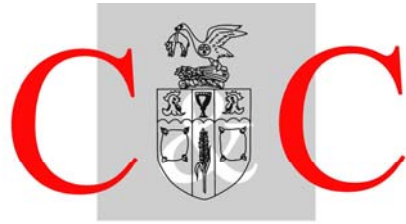
- Research originally undertaken within EU-funded project '**ProfiSens**' (SMT-4-CL98-2227)
- Ongoing research into the area of Proficiency testing, and regular proficiency testing schemes, offered by the **European Sensory Network**



www.esn-network.com



Campden & Chorleywood Food
Research Association Group



Campden & Chorleywood Food
Research Association Group

Thank you!



Questions?

