The reality of discrimination testing - a practical appraisal of Tetrad, Triangle and A not A

C V Barnagaud[1], N L Krog[1], P Mehring[1], V S Mialon[2], C L McLean[2]
MMR Research Worldwide Ltd[1], MMR Research Worldwide Inc[2]

AMER
For more information contact
Valerie Mialon
v.mialon@mmr-research.com

Introduction

In our ever increasing quest for relatively cheap and quick ways to find the right answer, discrimination testing is often an important tool for R&D functions. Discrimination testing can investigate if changes in formulation, manufacturing process or ingredient sources are likely to result in noticeable differences of the sensorial experience of products.

However, while a lot has been published about the theory involved and the statistical sensitivity and robustness of different methods, there isn't a lot of information available on the actual practicality of running these tests. Although overall the tests are considered quite simple to implement, the reality of conducting them correctly is not always easy.

Objectives

This study assessed three methods of discrimination testing: Tetrad, Triangle and A not A tests. The same number of assessors were used throughout to discover which approach may be more practical to implement, as well as assessing the relative complexity of the psychological task required.

Based on previous experience and initial trials with these discrimination methods, a few hypotheses were formed:

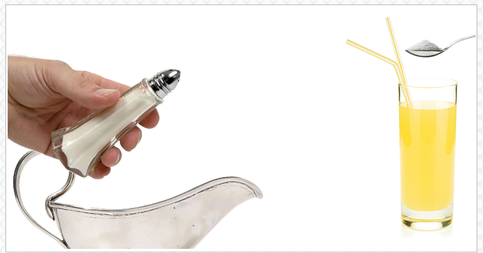
A not A (with repeated Reference)
Seemed straight forward to implement and most intuitive to panelists, as a direct comparison to a reference, but may include higher fatigue due to the high numbers of pairs presented

Triangle Test
Is widely used and easy to implement with the trained panelists who already had experience with the method (could be complex for naive consumers to perform), but may be less discriminative than the other approaches at equal sample size

Tetrad Test
Seemed most challenging initially for assessors and may therefore yield poor discrimination, though literature suggests otherwise [1].

Method

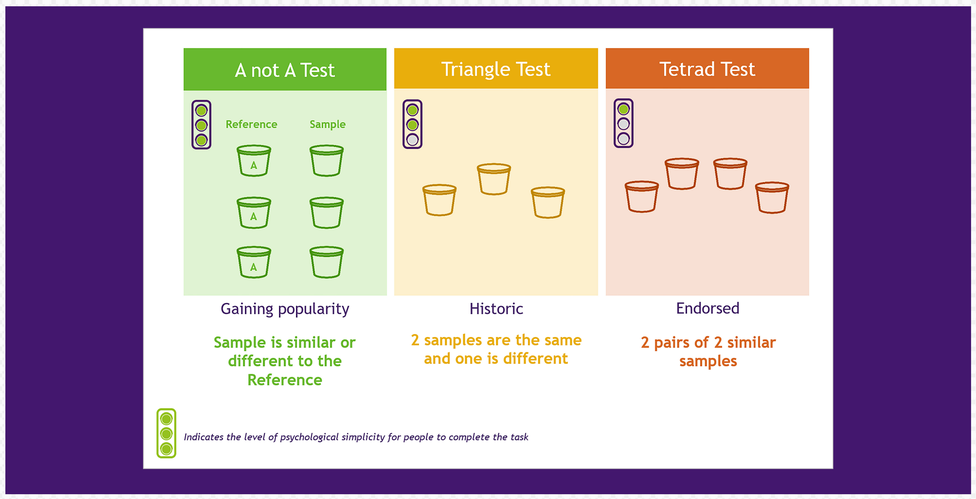
We used a pool of 22 highly trained panelists, from MMR's UK based descriptive panels, to evaluate two product categories:



- **Gravy with added salt**
  required careful temperature control and can be variable in texture
- **Orange 'squash' (i.e. dilutable drink) with added sugar**
  simple and easy to make up and serve

For both categories, differences were manufactured to create products that varied slightly, using an iterative process with the panel.

Each panelist conducted each test on each category, in duplicate (c.44 reads per test) over the course of a few months.



Findings

From a practical and psychological standpoint

The main practical challenge was found for the Triangle and Tetrad tests in terms of serving the gravy samples in quick enough succession to minimise the required memory effort, while leaving panelists with sufficient time to evaluate each sample in front of them.

In terms of the required task, the Tetrad approach was found to be the most difficult by panelists who reported low confidence. The method required more training and confidence building than the other two.

From a statistical standpoint

For each category we ran statistical analysis using Thurstonian modelling for Triangle and Tetrad, and R-index for A not A. We then calculated the P-value and d' as measures of discrimination.

For both categories, results were consistent and found to be highly significant on the Triangle and Tetrad tests. No significant differences were perceived however for the A not A tests (Tables 1 & 2).

Table 1. Orange Squash results

A not A Test	Triangle Test	Tetrad Test
Not significantly different	p=0.002 d'=1.36	p=0.006 d'=1.10

Table 2. Gravy results

A not A Test	Triangle Test	Tetrad Test
Not significantly different	p=0.024 d'=1.23	p=0.002 d'=0.98

The highly discriminating Tetrad results were in line with literature findings [1], despite the panelists' initial difficulties and apprehension, suggesting that training and familiarity with the method can lead to greater discrimination.

The Triangle results appeared more discriminative than expected (given the relatively low base size), however the panels were trained on the method suggesting again that familiarity with the test can lead to better results.

In terms of statistical robustness, A not A would be expected to outperform the Triangle test, but didn't in this study. The lack of discrimination for A not A could be due to the total number of samples tried. Indeed, the 'repeat with Reference' for A not A method was used to supposedly increase sensitivity. This method - in order to estimate whether a participant could discriminate above chance - required each assessor to try three different combinations of 'reference vs. reference' or 'reference vs. test' sample pairs (total of six samples i.e. the highest number of samples in all three tests). This may have led to assessor fatigue and negatively impacted the detection of small differences.

Although serving was a practical challenge found for the gravy product due to rapid temperature change, it had no impact on the test findings.
This suggests that with careful planning and trials, the same method can be successfully applied to various categories.

For all three tests, analysis was conducted to check for replication/learning effect. This was achieved by using a beta binomial model and applying a correction factor. In all instances, no significant replicate or learning effect was found.

A not A
Direct comparison may seem to be the easiest psychological task. However, in reality, sensory fatigue plays a big role!
Re-presenting the reference creates a design with the most samples tested

TRIANGLE
Although it is supposed to be underpowered relative to the other tests, it was as sensitive as the Tetrad test
This is perhaps because it is a simpler psychological task for trained panellists to complete!

TETRAD
With low base sizes, Tetrad is statistically very sensitive - as reported many times in conferences & publications
However, it requires some training to achieve these results
In reality it is a complex psychological task and the panel needed time to get used to the method

Conclusions

A not A (with repeated Reference)
We would not recommend using this method. It requires too many products to be tested in a sequence which can lead to sensory fatigue and showed the lowest overall results for sensitivity

Triangle Test
This would be our recommended consumer approach. It's easier to achieve the appropriate base size with consumers than with trained panelists, and the method is more intuitive for consumers to complete without training

Tetrad Test
Tetrad would be our gold standard approach when using trained panelists, but we would not recommend if using naive consumers. It is slightly more sensitive than Triangle, however some familiarization is required to achieve this level of sensitivity.

1) Ennis, John M. (2012) "Guiding The Switch From Triangle Testing To Tetrad Testing." *Journal of Sensory Studies 27.4: (pp.223-32)*