The Institute for Perception

For technical reports from The Institute for Perception or for information about short courses, please visit www.ifpress.com or email us at mail@ifpress.com

Reducing Costs with Tetrad Testing

John M. Ennis and Benoît Rousseau

Background: The Triangle test is one of the most popular discrimination methodologies used in consumer products companies. Beginning in 1941, Joseph E. Seagram and Sons were the first to employ it – since then it has been used in a variety of applications including product discrimination testing and panelist selection. An advantage of the Triangle test is that it does not require specification of the nature of the difference. Yet the Triangle test requires large sample sizes to be effective[1] and, recently, the Tetrad test has received interest due to its potential to provide increased power without specification of an attribute[2]. This greater power means that for the same sample size, an existing difference is less likely to be missed. Alternatively, greater power means that smaller sample sizes can be used to achieve the same performance as the Triangle - the sample sizes required by the Tetrad test are theoretically only one third that required by the Triangle test. Importantly, like the Triangle test, the Tetrad test does not require specification of a sensory attribute. The purpose of this technical report is to illustrate how sensory scientists can determine whether or not Tetrad testing is worth considering as a standard testing methodology within their discrimination testing programs.

Scenario: You work for a manufacturer of powdered sports drinks. As part of a recent cost reduction initiative, management would like to replace a flavoring agent with a lower cost alternative. From past work, you have determined that an effect size (δ) of 1 is the point at which differences become meaningful to your consumer population[3]. In order to be 80% sure of detecting such a difference, you consult a table specifying the sample sizes needed to obtain 80% power as a function of the methodology chosen and the relevant effect size[4]. A portion of this table is shown in Table 1.

δ	Triangle	Tetrad
0.80	488	140
0.85	389	113
0.90	318	94
0.95	262	78
1.00	220	65
1.05	184	57
1.10	154	47
1.15	135	42
1.20	116	39

Table 1. Sample size requirements for 80% power, Tetrad and Triangle tests.

From this table you conclude that Triangle testing would require 220 evaluations. But you only have 60 panelists and would prefer to not conduct 4 replications for each panelist. Then you notice that Tetrad testing requires 65 evaluations to detect an effect size of δ = 1 with 80% power. Looking through the table, you see that this reduction in sample size is consistent, with Tetrad testing only requiring about one third the sample sizes of Triangle testing.

The Tetrad Test: In the Tetrad test, four stimuli are presented to respondents. Two stimuli come from one group (A) while the other two come from a different group (B). Respondents are asked to group the samples into two groups of two based on similarity. Note that these instructions are different from asking the subjects to identify the two most similar samples. See Figure 1. The guessing probability for the Tetrad test is 1/3, as this is the likelihood of grouping the second 'A' sample with the first 'A' sample by chance. Thus, the Tetrad test is easily compared to the Triangle test – as long as the Tetrad test returns a higher proportion of correct responses than the Triangle test, the Tetrad test will be more sensitive.

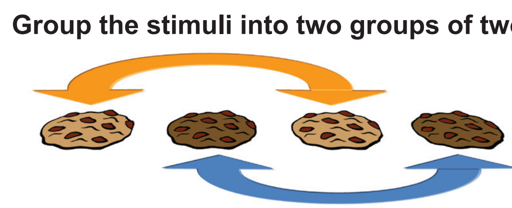

Figure 1. Instructions for Tetrad testing.

This higher proportion correct is predicted by Thurstonian theory[5] and is seen most easily by comparing the psychometric functions for the Tetrad and Triangle tests. See Figure 2. These functions show the predicted proportions of correct responses as the effect size (δ) increases. A higher proportion correct for the Tetrad test than the Triangle test has been observed in a variety of applications[6,7,8], yet the possibility exists that the addition of a fourth stimulus might cause a significant increase in perceptual noise[9]. If this increase in noise is large enough, the effect size of the Tetrad test could be reduced to the point that the theoretical power advantage of the Tetrad test might be lost. Since the amount of additional perceptual noise will depend of the product category (e.g., tasting of hot pepper samples will be more fatiguing then visual evaluations of baked goods), experimental comparison of the Tetrad and Triangle tests is recommended to ensure that a switch to the Tetrad test will provide more sensitive results in future testing.

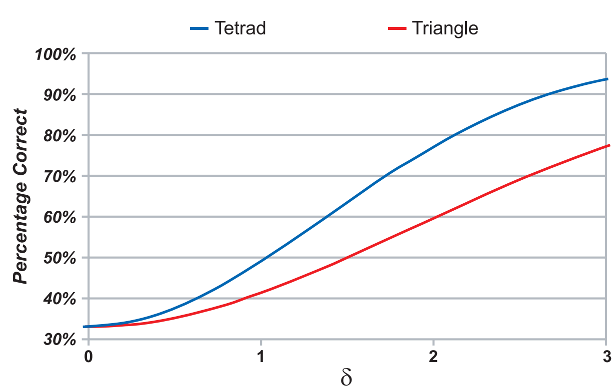

Figure 2. Psychometric functions of Tetrad and Triangle tests. Proportion correct is a function of effect size (δ).

Considering a Switch from Triangle to Tetrad Testing: In addition to the current proposed ingredient change, there are likely to be several other ingredient changes in the near future. Thus you begin to consider the possibility that switching from Triangle testing to Tetrad testing might lead to improved decision making for your company. One difficulty you notice, though, is that Tetrad testing requires evaluation of four stimuli instead of three. Because of the additional stimulus, you begin to wonder if the effect size for the Tetrad test might be smaller than the effect size for the Triangle test. For example, if in your situation you seek to detect an effect size of $\delta > 1$ in the Triangle test, it may be necessary to detect an effect size of 0.8 in the Tetrad test. But even if that were the case, you would still only require 140 evaluations from the Tetrad test instead of the 220 evaluations the Triangle test requires.

In order to quantify the decrease in effect size that the addition of a fourth stimulus may cause, you invest in a series of comparative experiments. In these experiments, you test the existing product against samples that contain 20%, 40%, 60%, 80%, and 100% of the new ingredient in place of the current ingredient. Since you require a stable comparison of the Triangle and Tetrad tests, you conduct 3 replications of each panelist for a total of 180 evaluations per condition. The data from these comparative experiments, together with the Thurstonian d' values, are shown in Table 2.

	Triangle		Tetrad	
New Ingredient	# Correct	d'	# Correct	d'
20%	61	0.247	61	0.174
40%	67	0.662	67	0.467
60%	73	0.918	85	0.921
80%	81	1.193	97	1.157
100%	100	1.751	112	1.437

Table 2. Data from comparative experiments ($N = 180$).

From these experiments, you see that the d' values for the Tetrad test are indeed less than the d' values for the Triangle test. Figure 3 shows a plot of the d' values for these two testing methods against each other. This plot shows a fairly linear relationship between effect sizes returned by the two methods. In particular, a regression line through the origin with slope 0.874 explains 95% of the variance in the results. Thus, instead of seeking to detect a Triangle effect size of $\delta = 1$, it is reasonable to seek to detect a Tetrad effect size of 0.874. To be conservative, you decide to detect a Tetrad effect size of 0.85. From Table 1, you see that by replicating your panelists twice, you can achieve more than the required 113 evaluations for 80% power assuming a homogeneous panel of assessors. Since 220 Triangle evaluations were required to obtain the same level of power, you find that by switching to the Tetrad test you can achieve higher quality information from two replications than you previously would have obtained from three or more replications. Over time, this efficiency gain will translate into significant cost savings. In addition, the business risk associated with possible changes in ingredient or processing will be more accurately quantified, allowing your company to make higher quality decisions.

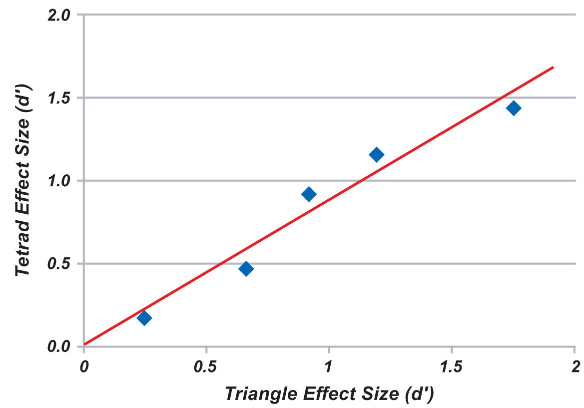

Figure 3. Comparison of Tetrad and Triangle effect sizes.

Conclusion: To stay competitive, businesses must reduce costs while increasing the quality of their decisions. Sensory science can play a central role in the achievement of these seemingly contradictory objectives by providing efficient testing methods that give more reliable information at lower costs. In particular, the Tetrad test holds promise as a viable replacement for the Triangle test in situations where the addition of a fourth stimulus does not cause so much additional perceptual noise as to overwhelm the theoretical advantage of Tetrad testing. The extent to which the addition of a fourth stimulus causes an increase in perceptual noise and a corresponding decrease in effect size is an experimental question that depends on the product category.

References and Notes

1. Ennis, D. M. (1993). The power of sensory discrimination methods. *Journal of Sensory Studies,* **8**(4), 353-370.
2. Ennis, J. M., and Jesionka, V. (2011). The power of sensory discrimination methods revisited. *Journal of Sensory Studies,* **26,** 371–382.
3. Rousseau, B. (2010). Action standards in a successful sensory discrimination program. *IFPress,* **13**(4), 2-3.
4. Such a table can be found in Ennis & Jesionka (2011).
5. Ennis, J. M., Ennis, D. M., Yip, D., and O'Mahony, M. (1998). Thurstonian models for variants of the method of tetrads. *British Journal of Mathematical and Statistical Psychology,* **51,** 205-215.
6. Masuoka, S., Hatjopoulos, D., and O'Mahony, M. (1995). Beer bitterness detection: Testing thurstonian and sequential sensitivity analysis models for triad and tetrad methods. *Journal of Sensory Studies,* **10,** 295-306.
7. Delwiche, J., and O'Mahony, M. (1996). Flavour discrimination - an extension of Thurstonian paradoxes to the tetrad method. *Food Quality and Preference,* **7,** 1-5.
8. Garcia, K., Ennis, J. M., and Prinyawiwatkul, W. (submitted). A large-scale experimental comparison of the tetrad and triangle tests in children. Submitted to *Journal of Sensory Studies*.
9. Ennis, J.M. (submitted). Guiding the switch from triangle testing to tetrad testing. Submitted to *Journal of Sensory Studies*.