2022 Issue 25(4)

Sample Selection for Method Comparisons

Benoît Rousseau and Daniel M. Ennis

Background: "Research on research" is a phrase often heard in consumer product companies with the most enlightened research programs. This research pays dividends when a company later studies the performance of their food, beverage, or personal care products. In methodological investigations where modified products are used, the choice of stimuli is of importance. For instance, with the introduction of the tetrad method, many companies are conducting internal research to answer the question of whether the tetrad is a suitable alternative to the triangle or duo-trio test when considering their own panel and product category. There is also an interest in comparing the sensitivities of panels with different levels of training, or in studying the performance of an internal panel with that of a representative group of consumers.

When conducting methods research, the stimuli play an essential role since the outcome of the research will depend on how well controlled the experimental variables are so that the conclusions can have clear and interpretable relevance in practice. A lack of careful selection can result in stimuli that are too similar or too different and defeat the purpose of the research. In this technical report, we describe how Torgerson's method of triads[1,2] can be a valuable tool to ensure the appropriate choice of test samples for methods research.

Scenario: You manage your company's sensory group and explore ways to improve the reliability of the information you collect using your internal panels. This sometimes means conducting research to increase methodological understanding and to exploit opportunities to increase management confidence in your conclusions and recommendations. Of particular interest is the consideration of potential improvements to your main discrimination testing program of highly flavored cookie products. Because these are relatively fatiguing products, you have been traditionally using the degree of difference method (DOD) on a 5-point scale (1= identical to 5= very different) which requires a comparison of just two samples in each trial. You use the tetrad method in some of your other product categories, and since it is theoretically more powerful than the DOD method[3], you would like to conduct research comparing the two techniques and to determine if the tetrad's four samples lower its statistical power. A small decrease might still make it a more powerful alternative to the DOD.

Preliminary Research: Preliminary comparative experiments are conducted and involve the same pair of samples evaluated using both the DOD and tetrad methods. You select five sample pairs based on sample availability involving prototypes as well as some competitor products. Each experiment involves a sample size of $n = 60$, that includes your base group of 20 panelists performing three replications for each pair and method. Upon completing the data collection, you analyze the results statistically and estimate the size of the underlying sensory difference (d' values.) Meaningful values of d' are usually in the range of 0.5 – 2.0.

The results showed that:

- Two of the pairs yielded non-significant differences ($p > 0.05$) and very small d' values (0.00 – 0.03) for both methods.
- The other three pairs resulted in highly significant differences for both methods ($p < 0.001$) and very large d' values (> 3.0) for both methods.

These results demonstrate a problem in sample selection leading to a loss of time and resources. Recognizing that this first investigation was not successful, you study how sample pairs could be chosen more systematically so that the comparative results will yield a decision on whether a switch to the tetrad method is advisable.

Torgerson's Method for Sample Pair Selection: Since the research outcome is driven by the choice of the sample pairs, careful selection is needed. Pairs with differences of different sizes are needed so that a result can be generalized to the range of relevant sensory discrimination testing differences (d' values of 0.5 – 2.0). An efficient approach involves the systematic variation of a stimulus along a sensory dimension of interest and the use of Torgerson's method of triads[1,2].

Torgerson's method resembles a duo-trio trial with a reference sample and two alternatives. However, unlike the duo-trio method, the three samples are putatively different, and the instructions require the subject to select the sample 'most similar' to the reference. For sample pair selection, a minimum of four variable levels are selected within the reasonable range of sensory differences. The 12 possible triads are evaluated by panelists and the pairwise d' values are estimated simultaneously using the corresponding Torgerson's Thurstonian model[2,4].

Application of Torgerson's Method: You apply Torgerson's method to the baking time of your flavored cookies. From this, a dose-response relationship is built and the level differences corresponding to d' values of 0.5, 1.0 and 1.5 can be estimated precisely. In this instance, you choose four levels summarized in Table 1.

Table 1. Four starting baking time levels for Torgerson's method of triads

Sample	Time
B_0	9mn
B_1	10mn
B_2	11mn
B_3	12mn

You use these four levels to create the triads for panel evaluation. Table 2 provides a subset of the 12 evaluated triads and the corresponding results for a total of 20 trials of each triad. Smaller or larger number of trials per triad can be collected, but a minimum of 12 trials is recommended to allow sufficient precision in the parameter estimates.

Ref	S_1	S_2	Choice (S_1 more similar)	N
B_0	B_1	B_2	15	20
B_0	B_1	B_3	8	20
B_0	B_2	B_3	13	20
B_1	B_0	B_2	12	20
…	…	…	…	…
B_3	B_1	B_2	10	20

Table 2. Subset of Torgerson results involving four stimuli and 20 evaluations for each of the 12 triads

Using the data from Table 2, pairwise *d′* values are estimated using the Torgerson Thurstonian model (analysis conducted using the Tools version of the IFPrograms® software 9.0.5). The *d′* values are then plotted against the variable levels as illustrated in Fig. 1a.

From the graph (and associated regression line), you can establish the following (Fig. 1b):

d′ = 0.5 corresponds to baking times: 9′ vs. 9.4′

d′ = 1.0 corresponds to Baking times: 9′ vs. 10′

d′ = 1.5 corresponds to Baking times: 9′ vs. 11′

These butter cookie stimulus pairs can then be used for method or condition comparisons with the assurance that the underlying sensory differences are within discrimination testing -relevant ranges.

Satisfied with the information you collected in this preliminary phase, you proceed with the DOD and tetrad method comparisons. You are now confident that using these carefully selected sample pairs, your upcoming extensive investigation will deliver insights on whether the tetrad method is an advisable alternative to the DOD when using your panel and evaluating your company's products.

Conclusion: "Research on research" is used to increase the performance of sensory evaluation methods and to select the most appropriate method for a particular application. Through this research, valuable resources will be used more efficiently. For this type of research, careful stimulus selection is required. Torgerson's method of triads and its associated Thurstonian model offer a systematic approach that in turns paves the way for valuable experiments in method comparisons[5,6,7].

References

1. Torgerson, W. S. (1958). *Theory and Methods of Scaling*. New York: Wiley.
2. Ennis, D. M., Mullen, K., & Frijters, J. E. (1988). Variants of the method of triads: Unidimensional Thurstonian models. *British Journal of Mathematical and Statistical Psychology*, 41, 25-36.
3. Ennis, J. and Christensen, R. (2015). A Thurstonian comparison of the Tetrad and Degree of Difference tests. *Food Quality and Preference*, 40, 263-269.
4. Rousseau, B. (2007). Simultaneous estimations of multiple product similarities using a new discrimination protocol *Journal of Sensory Studies*, 22(5), 533–549.
5. Fu, Y., Rousseau, B., Ennis, D., Eberhardt, K, & Rossi, F. (2011). Using Thurstonian Predictions and Experimental Data to Investigate the Optimal Category Scale Structure for the Degree of Difference Protocol. *Rosemary Pang-born Symposium*, Toronto, Canada.
6. Kok, LL., Ullrich, N., Magnuson, L., Yglesias, R., Yurt-tas, N., & Ennis, J. (2015). Switching from Duo-Trio Test-ing to Tetrad Testing - A Case Study -. *Rosemary Pangborn Symposium*, Gothenburg, Sweden.
7. Rousseau, B., Ennis, J., Cress, C. Harmon, J., McIngvale, S., & Thompson, S. (2019). A Three-Step Approach to Tetrad and Triangle Comparisons. *Rosemary Pangborn Symposium*, Edinburgh, Scotland.

Fig. 1a

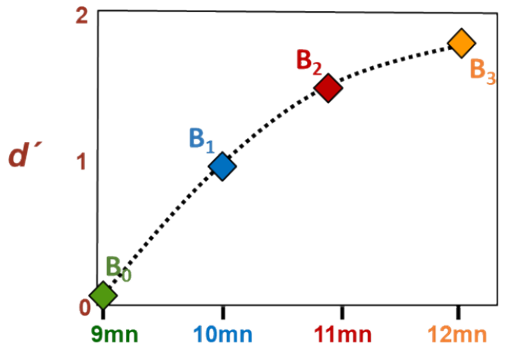
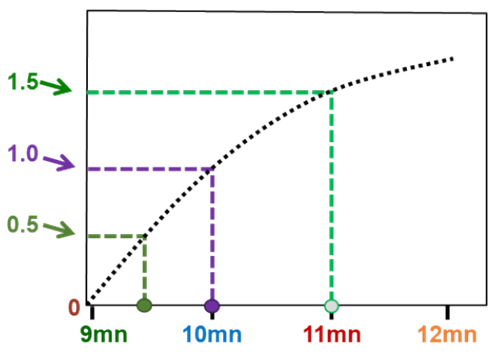


Fig. 1b



Figures 1a and 1b. The dose-response curve relating baking times to sensory differences (*d'*) (Fig 1a) and estimated baking times corresponding to *d'* values of 0.5, 1.0 and 1.5 (Fig. 1b).