Scaling First-Last, MaxDiff and Best-Worst Data

Daniel M. Ennis

Background: Among market research practitioners, there has recently been interest in scaling product or category characteristics (such as possible benefits) based on responses indicating the items with the greatest and least magnitude among a subset of possible items. The basis for this choice could be liking, purchase interest, importance or even a sensory characteristic such as sweetness. For example, considering the characteristics of plug-in air care products, items to consider might include "*low cost*," "*does not fade over time*," "*has a use-up cue*," and "*has a fresh scent*." A respondent may be instructed to choose the attribute of most importance and the one of least importance in making a purchase decision. From a large collection of items, subsets of equal size are chosen and presented in a balanced design. The typical number of items used per respondent is four. The analytic task is to develop a scale on which each attribute can be placed so that scale values for all of the items from most to least can be obtained.

Scenario: Your company markets a variety of air care products including instant action air fresheners and plug-in products that last for as long as 60 days. In order to improve the new product development process by identifying key features of interest to your consumers, you seek to prioritize ten possible features of plug-in air fresheners. These features are presented in Table 1. Among a target group of interest, you obtain responses to questions as shown in Figure 1 concerning the importance of the features listed. Sets of four features are presented in a balanced design to each of 200 consumers and each consumer evaluates fifteen sets of four features.

Related Methods in Psychophysics: Although the use of this type of methodology in market research is fairly recent, it has a much longer history in psychophysics. In 1938, while investigating multidimensional scaling, Richardson[1] proposed a triadic method that required subjects to indicate the two most different and the two most alike members of the triad. This method was discussed by Torgerson[2] in the late 1950s. A unidimensional Thurstonian model for Richardson's method was published in the late 1980s that provides a technique to scale the magnitudes of the items tested[3,4]. The magnitudes can be interpreted as scaled utilities when the magnitudes are hedonic, such as liking or purchase interest. A component of this model includes the most different pair only. In Richardson's method, there is no information obtained from each respondent about attribute direction, which means that the respondent does not indicate which member of the 'most different' pair is greatest or least on the attribute of interest. In the description given earlier of the plug-in example, we have access to information about which pair is most different along with information on the individual items that are perceived to be greatest and least. For this reason, the method can be viewed as a first and last choice method. Thus the use of the term "MaxDiff" to describe this method, which implies that the subject simply chooses the most different pair, is inappropriate. Due to the fact that the method may be applied to non-hedonic characteristics (such as sweetness or sourness, for instance), it is also limiting to call the method "best-worst". Although both of the terms "MaxDiff" and "best-worst" have been used extensively to describe the method, the method is best described as "first-last", since it involves judgments of first and last choice.

Ten Plug-In Features	
Low cost	Comes in a wide variety of possible scents
Has a use-up cue	Has an attractive appearance
Does not fade over time	Fills the entire room with fragrance
Lasts at least 60 days	Made by an environmentally sensitive company
Has a fresh scent	Is the best selling brand

Table 1. Ten possible benefit features of plug-in air fresheners.

Of the following four features of plug-in products, which is the most and which is the least important to you when making a purchase?		
Most	**Feature**	**Least**
○	Low cost	●
●	Does not fade over time	○
○	Has a use-up cue	○
○	Has a fresh scent	○

Figure 1. Typical responses in a first-last choice task.

A first choice Thurstonian model was published by Ennis and O'Mahony[5] to account for sequential effects in product tests. Recently, a Thurstonian model for first-last choice has been generalized to include any number of items in the subset. The model is different from a commonly used alternative based on the logit[6], as the new model assumes that the items are distributed normally on the unidimensional utility or hedonic continuum. A great advantage of this approach is that it allows the prediction of results from other psychophysical tasks, including paired preference, rating and ranking tasks, using the already developed Thurstonian family of related models, and thus provides a basis for validation and comparison. This model also allows the determination of the variances and covariances of the parameter estimates at the aggregate level so that statistical comparisons can be made.

Modeling First-Last Data: In order to approach the development of a first-last Thurstonian model, we assume that the perceptual representations of the items can be modeled as univariate normal distributions. For simplicity, we assume these distributions have unit standard deviations. The difference between the means of the distributions are called δ values and the units of the δ values are perceptual standard deviations. These are the same assumptions that we used to model the results of difference tests, ratings and rankings in previous papers and technical reports[7,8,9] and these analyses can be conducted in *IFPrograms*™. In addition, we assume that a respondent considers the features listed in Figure 1 and uses the same perceptual values when making the first and last decisions. This last assumption is called *dependent* sampling. *Independent* sampling would occur if the respondent recorded their first choice and then considered a separate random sample to decide their last choice. We assume dependent sampling, although independent sampling may occur in other settings, such as when a respondent is re-tasting or re-smelling food items between their choice of the first and last items. A complicated issue that arises with independent sampling is that the same item may appear to be both first *and* last. Although here we assume dependent samples, this issue has been discussed previously for Richardson's method of triads[4]. The first-last choice model connects δ values with probabilities of choosing particular items first or last.

Results from Modeling the Plug-In Data: Figure 2 shows the *d*' values (estimates of δ values), or scale means, for the ten plug-in features. It can be seen that for this particular group, you conclude that the benefit "*does not fade over time*" is the most important item and "*made by an environmentally sensitive company*" is the least important. All of the benefits are scaled relative to the last one which is set at zero. Figure 2 also shows standard error bars for the scale means. Since the last benefit is assumed to have a δ value of zero, it has no error bar. Using these results, we can now predict the outcomes of other methods for which Thurstonian models have been developed.

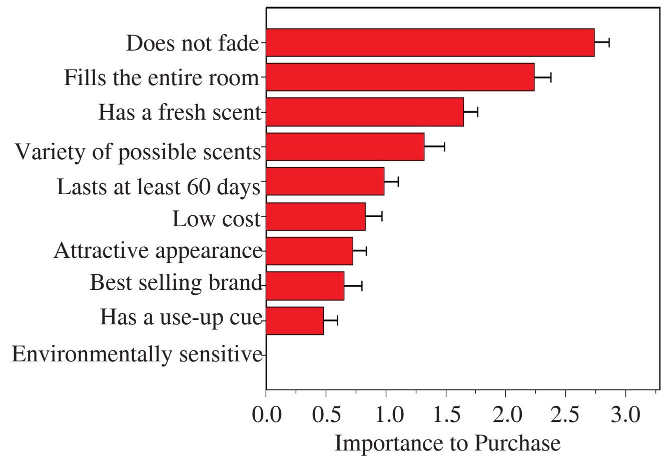

Figure 2. Scale means (*d'* values) and their standard errors for ten features.

Conclusion: A Thurstonian model with dependent sampling has been developed and, in the scenario, was applied to first-last data for plug-in fragrance products. The resulting δ value estimates were used to prioritize the benefits for this category. The psychophysical task involving first and last choices has been referred to historically as MaxDiff and Best-Worst testing. Neither of these descriptors is appropriate as the method involves more than finding the two items that exhibit maximum difference and is not limited to hedonic or utility continua.

References:

1. Richardson, M.W. (1938). Multidimensional psychophysics. *Psychological Bulletin*. **35**, 659-660.
2. Torgerson, W.S. (1958). *Theory and Methods of Scaling*. New York: Wiley.
3. Ennis, D. M., Mullen, K. and Fritjers, J. E. R. (1988). Variants of the method of triads: Unidimensional Thurstonian models. *British Journal of Mathematical and Statistical Psychology*, **41**, 25-36.
4. Ennis, D. M., Mullen, K., Frijters. J. E. R. and Tindall, J. (1989). Decision conflicts: Within-trial resampling in Richardson's method of triads. *British Journal of Mathematical and Statistical Psychology Society*, **42**, 265-269.
5. Ennis, D. M. and O'Mahony, M. (1995). Probabilistic models for sequential taste effects in triadic choice. *Journal of Experimental Psychology: Human Perception and Performance*, **21**(5), 1-10.
6. Marley, A.A.J. and Louviere, J.J. (2005). Some probabilistic models of best, worst, and best-worst choices. *Journal of Mathematical Psychology*, **49**, 464-480.
7. Ennis, D.M. (1999). Thurstonian models for intensity ratings. *IFPress*, **2**(3) 2-3.
8. Ennis, D.M. (2004). From ranks to intensities. *IFPress*, **7**(4) 2-3.
9. Ennis, D.M. (1998). Thurstonian scaling for difference tests. *IFPress*, **1**(3) 2-3.