Developing Multiple Paired Comparisons Model for Analysis of Variance Technique: Experimental Test

Abbood M. Jameel

Department of Public Administration, Cihan University-Erbil,
Kurdistan Region, Iraq

***Abstract*—A class of multiple paired comparison models can be developed for taste-testing experiments to evaluate several objects on a seven-point preference scale. This allows for calculating a preference score for each object using the two-way analysis of variance technique (ANOVAT). The primary objective of this research is to distinguish between different types of objects or items by calculating degrees of preference (scores), denoted as S j. These objects, represented by L, are evaluated through paired comparisons. Specifically, we consider six types of potato chips, labeled as A, B, C, D, E, and F. Each pair of chips is presented to two equal groups (2n) of judges or referees, where each group consists of *n* = 15 individuals. The goal is to determine which object or item is favored over the other. In these comparisons, judges or referees answer the question: Which do you prefer, Object O_j or Object O_k? Their responses are recorded on a seven-point preference scale. The data collected from these paired comparisons are organized using a tabulation method, as demonstrated in Table I. Various statistical methods, including the developed multiple paired comparison model and the two-way ANOVAT, were employed for analysis. The results, summarized in Table IV, indicate that potato chip type B is the most preferred, followed by types D and C, respectively. Type F was found to be the least preferred.**

***Keywords*—Multiple paired comparison, Ranking and hypotheses testing, Rating on a scale, Taste-testing experiment, Two-way analysis of variance.**

I. Introduction

Paired comparison methods can be applied in all cases where there are more than two objects to compare. If there is a set of L objects, each judge could make a total of L(L−1)/2 possible pairwise comparisons. Pairwise comparison is often the most suitable experimental procedure, such as when comparing various brands of potato chips, where two types are tasted at a time (Thurstone, 1927a). While more samples of each product are required, it is much easier for the judge to state their preference. The order of presentation may influence the judge's preference in taste-testing experiments (Davidson, 1970). This issue can be addressed by giving each judge both orderings of the two objects simultaneously. However, this approach can be challenging, especially when the differences between the objects are minimal. If a judge is asked to evaluate several highly similar objects, as in taste-testing experiments, their judgment is likely to be affected by the order of presentation (Bühlmann and Huber, 1963). Analysis of variance is a widely used statistical method, particularly among researchers and statisticians, for measuring and clarifying variance or differences between and among variables, whether independent or dependent, across various scientific disciplines (Massoudi *et al*., 2025). In this paper, we aim to develop a multiple paired comparison model to calculate a score for each object or item, making it suitable for two-way analysis of variance. We also employ the analysis of variance method to measure differences resulting from multiple paired comparisons, alongside other statistical techniques. The first application of the analysis of variance method for binary comparisons in this field was described in Scheffe (1952). The core idea of this research involves the evaluation of several objects or items (L) through a paired comparison method. This approach highlights the effect of the order in which objects are presented for evaluation and preference. Each pair is presented to n judges or evaluators in one order (j, k, k, j,) and to another group of n judges or evaluators in the reverse order (k j, j, k). The judges use a seven-point preference scale for their evaluation (Scheffé, 1952).

A. Seven-Point Scale of Preference

1. I slightly prefer object (O_j) over object (O_k): 1
2. I moderately prefer object (O_j) over object (O_k): 2

Cihan University-Erbil Journal of Humanities and Social Sciences (CUEJHSS)
Volume IX No. 1 (2025) 5 pages
DOI: 10.24086/cuejhss.vol9n1y2025.pp 183-187
Received: 21 January 2025; Accepted: 19 April 2025; Regular research paper: Published 10 June 2025
*Corresponding author's e-mail: abood.jameel@cihanuniversity.edu.iq

3. I strongly prefer object (O_j) over object (O_k): 3
4. I have no preference between the two objects: 0
5. I slightly prefer object (O_k) over object (O_j): −1
6. I moderately prefer object (O_k) over object (O_j): −2
7. I strongly prefer object (O_k) over object (O_j): −3.

The numbers in parentheses represent the degree of evaluation or preference assigned by the judges or referees when comparing the two objects (O_j and O_k). A positive score (1, 2, or 3) is assigned to the object presented first for evaluation, while a negative score (−1, −2, or −3) is assigned to the object presented second. In this way, we can collect data for such experiments, commonly referred to as taste-testing experiments (Mihálykó, 2024).

The objectives for this study are: To develop a multiple paired comparison model for two-way analysis of variance. Also, to estimate the degrees of preference for each object submitted for evaluation. Identify the most preferred object among various types of objects. Determine whether there is a meaningful or significant difference between the degrees of preference. Finally, assess the stability, consistency, or homogeneity of the judges' or referees' opinions in this experiment.

II. Literature Review

Paired comparison methods are widely used in experiments that involve subjective evaluations by individuals of objects presented in pairs. These experiments typically require individuals to choose between objects based on several attributes, often resulting in multivariate analyses (Mihálykó, 2024). The concept of a subjective continuum has been modeled through paired comparisons, capturing inherent sensations where order can be discerned but physical measurements are not applicable (Thurstone, 1927b). In one study, it was assumed that the judges' responses (preferences) for object O_j, denoted as X_{jm}, and object O_m, denoted as X_{mj} (j, m = 1, 2,…, L), are independent and have equal variances on a subjective response scale. The distribution of any X, (X_{jm} or X_{mj}) is assumed to be normal, and hence the distribution of any (X_{jm}-X_{mj}) is normal. Guttman (1946) developed a method to quantify paired comparisons, assuming k objects are being compared. Each of n judges evaluates all possible pairs, making n sets of k(k−1)/2 comparisons, with no ties permitted. Mosteller (1951) provided a detailed analysis of Thurstone's Case V, which assumes equal correlations and variances across all variables (y). Under suitable scaling, each object is assigned a location on the continuum, representing the expected value of y_{jm} for T_j (j = 1, 2,…, k). When a judge compares T_j and T_m, they essentially report the order of sensations Y_j m and Y_{mj} which may be correlated and associated with T_j and T_m. A simple ranking method was proposed for paired comparison experiments, where each judge records "1" for the preferred object and "2" for the other Bradley (1953). Glenn and David (1960) extended the Thurstone-Mosteller method to allow for ties by introducing an interval length (2d) centered at the origin of the distribution of (T_j−T_m). Within this interval, judges are unable to distinguish between T_j and T_m and declare a tie. Rao and Kupper (1967) proposed a modification of the Bradley and Terry model (1952) to accommodate tied preferences, effectively generalizing the original model. Davidson and Bradley (1969) further developed a multivariate paired comparison model, extending the Bradley-Terry framework to handle comparisons across multiple attributes. Davidson (1970) proposed additional extensions to the Bradley-Terry model to better account for tied preferences, emphasizing that the modified model should meet certain conditions. Fienberg and Larntz (1976) introduced a log-linear representation for paired comparisons and multivariate paired comparison models. Their parameterized approach used log-linear models to provide analyses equivalent to the Davidson–Bradley method but relied on the odds ratio as a measure of association rather than the correlation coefficient. Fowlkes *et al*. (1988) discussed supplementary techniques to the basic log-linear/logistic methodology. Using three examples, they demonstrated methods for understanding large multidimensional contingency tables by comparing model fits and offering intuitively appealing ways to interpret results.

A. Hypotheses

$H1_0$: There is no real or substantial difference between the types of objects submitted for judgment.

$H1_1$: There is a real or substantial difference between the types of objects submitted for judgment.

$H2_0$: There is no homogeneity or consistency in the opinions of the judges when evaluating or preferring the objects submitted for judgment.

$H2_2$: There is homogeneity or consistency in the opinions of the judges.

$H3_0$: The order in which the objects are presented for judgment does not have a significant effect, regardless of the arrangement or order.

$H3_1$: The order in which the objects are presented for judgment has a real or significant effect.

$H4_0$: There is no difference between the original data from the community and the data obtained from the judges' opinions using the seven-point scale.

$H4_1$: There is a difference between the original data from the community and the data obtained from the judges' opinion.

III. Procedure of Developing Multiple Paired Comparisons Model

Multiple comparisons model conducts an analysis of all possible pairwise comparisons means or averages. With paired comparisons, judges choose the object, or item, in each pair that has the greater magnitude on the choice dimension they were instructed to use as we mentioned before (Nyarko, 2023). The simplest approach, which we will be used throughout, is to present all possible pairs of the objects or items to each judge. With L objects or items, the total number of pairs will be L (L–1)/2 pairs (Wieren's, 1974). The choices allow calculation of a set of scale values indicating the position of the items along the specified

10.24086/cuejhss.vol9n1y2025.pp183-187

dimension (Cowden, 1975). The choice dimension may be whatever fits the objects or items being evaluated. When presented with a pair of objects or items, judges are not offered an indifference option (Mihálykó, 2024). This practice is supported by the theory of stochastic preference, wherein the probability of true indifference at any one moment is assumed to be very small. The practice also has the practical benefit of maximizing the amount of information potentially learned about the judge's preferences. For each judge or respondent, the full set of choices yields a preference score for each object or item, which is the number of times the respondent preferred the object or item to other objects/items in the set (Nyarko, 2023). Preference scores are the simplest form of scale values for the objects/items.

Let me, now, introduce the essential elements in the suggested multiple paired comparisons model. Consider that:

n = Number of judges

L = Number of objects

j = 1, 2… L O j = Object j

= 1, 2,… L Ok = Object k k

Yi_{jk} = The response of the judge i who prefer the object O_j to the object O_{ki} = 1, 2, 3… n

X_{jk} = [n1I × (−1)] + [n2 × (−2)] + [n3 × (−3)] + [n4 × 0)] + [n5 × (1)] + [n6 × (2)] + [n7 × (3)] (1)

Where: $n_1 + n_2 + n_3 + n_4 + n_5 + n_6 + n_7 = n$.

That is: -$\sum X_{jk}$ = Equation (1) is equal to the sum of the multiplying of each of the seven degrees of the preference scale by the number of judges or evaluators corresponding to that degree when comparing the object (O_j) with the object (O_k), in the order ($_{j},_{k}$).

Hence

$$\bar{x}_{kj} = \sum X_{kj}/n \quad (2)$$

Mean preference

$$\bar{x}_{jk} = \Sigma X_{jk}/n$$

$$M_{jk} = \sum\sum(\bar{x}_{jk\,-}\,\bar{x}_{kj})/2 \quad (3)$$

Average mean preference

$$M_{kj} = \sum\sum(\bar{x}_{kj}-\bar{x}_{jk})/2$$

$$d_{jk} = (M_{jk} + M_{kj})/2 \quad (4)$$

Order effect

$$\bar{d} = \sum\sum d_{jk}/L\ (L-1) \quad (5)$$

Average order effect

Final preference score for each object, S_j and S_k

$$S_j = \sum M_{jk}/L$$

$$S_k = \sum M_{kj}/L \quad (6)$$

$$\gamma_{jk} = M_{jk} - (S_j - S_k) \quad (7)$$

Deviation from Subtraction.

Tc, c = 1,2.,7 vertical total sum of the number of Judges in each Cell of the 7- point scale.

Tc = Represents the total vertical total of the number of referees or evaluators who are granted a certain degree of preference from one of the 7° of the preference scale for both ranks (k, j), (j, k), and for all possible paired comparisons. Therefore, the Multiple Paired Comparison Model for two-way analysis of variance can be written as follows:

$$Y_{ijk\,=}(S_j–S_k) + (d_{jk}–\bar{d}) + \gamma_{jk} + E_{ijk} \quad (8)$$

This was developed to be used in the Two-way ANOVA Table I. After these basic definitions (Assumptions), the main concepts of the two-way analysis of variance technique can be formulated in Table II bellow (Scheffé, 1952).

IV. Results

We assume there are six types of potato chips, labeled as A, B, C, D, E, and F. These types were evaluated and ranked

TABLE I
All Possible Paired Comparisons and the Number of Judges at Each Level of the Preference Scale

Pair j, k k, j	Preference scale −3	−2	−1	0	1	2	3	Total score X_{jk}* X_{kj}	Mean preference $\bar{x}_{jk}$ $\bar{x}_{kj}$	M_{jk} M_{kj}
A, B	3	2	0	4	3	2	1	−3	−0.7000	−0.200
B, A	1	2	0	0	2	7	3	18	0.7000	1.200
A, C	3	5	2	1	2	1	1	−8	−0.7330	−0.933
C, A	1	2	2	1	3	4	2		0.7330	0.533
…				…				…	…	…
…				…				…	…	…
B, C	1	1	1	3	2	5	2	0.7335	0.800	12−10
C, B	3	3	3	3	1	2	1	−0.7335	-0.667	
…				…				…	…	…
…				…				…	…	…
C, D	3	3	2	2	2	2	1	−0.7330	−0.533	−8
D, C	1	1	2	1	3	3	4	0.7330	0.933	14
…				…				…	…	…
…				…				…	…	…
D, E	1	2	2	3	2	**2**	3	0.4665	0.400	6−8
E, D	3	2	2	4	2	**1**	1	−0.4665	−0.533	
…				…				…	…	…
…				…				…	…	…
E, F	2	1	1	4	2	**2**	3	0.5000	0.400	6−9
F, E	4	2	2	1	1	**2**	2	−0.5000	−0.600	

XAB = (3× −3) + (2× −2) + (0×−1) + (4×0) + (3×1) + (2×2) + (1×3) = −3.

TABLE II
Sum of Squares

Total sum square (SSt)	$=T_1(-3)^2+T_2(-2)^2+T_3(-1)^2+T_4 (0)^2+T_5 (1)^2+T_6 (2)^2+T_7 (3)^2$
Sum squares mean effect (S S s_j)	$= 2nL\sum S_j^2$
Sum squares of preference averages) S S $_{xjk}$)	$= n \sum\sum X^2_{jk}$
Sum squares of averages mean preference (S S $_{m\,jk}$)	$= 2n \sum\sum M^2_{jk}$
Sum squares of average ranking effect (S S_d)	$= S\ S\ x_{jk}–S\ S_{mjk}$
Sum square of co-effect (S S $_{\gamma jk}$)	$= S\ S_{mjk}–S\ Ss_j$
Sum square error (SSE)	$= S\ St–S\ Sx_j$

in an experiment referred to as the taste-testing paired comparison experiment (Bühlmann and Huber, 1963). This experiment was conducted at a food industry company but can also be replicated in any supermarket to determine which type is most preferred among the available options (Massoudi and Fatah, 2024). In this experiment, 30 judges, referees, or evaluators were selected from the company's employees, all of whom had prior experience participating in similar experiments. The judges were asked the following question: Which type of these two potato chips do you prefer: The first potato chips or the second potato chips? The judges were required to use a seven-point preference scale to indicate their choice. Since the total number of potato chip types is L = 6, the number of possible paired comparisons is calculated as follows (Castura *et al*., 2023) number of comparisons = L(L−1)/2=15. The analysis was conducted using the analysis of variance technique to test the following hypotheses: the hypotheses H0: $\mu_1 = \mu_2 = \ldots = \mu_L$ or H_0: $\mu_i = \mu_j$ for all pairs i, j. i, j = 1, 2,…, L (Scheffé, 1952). To ensure a balanced evaluation, the 30 judges were divided into two equal groups of 15 each. One group was presented with the pairs in the order (j, k), and the other group was presented with the same pairs in the reverse order (k, j). The collected responses were summarized in Table I below:

$\bar{x}$ AB = −3/15 = −0.2

Mjk = ∑ ∑ ($\bar{x}$ jk-$\bar{x}$kj)/2=−0.7000

Table I consists of the number of judges who tasted brand A and brand B together in the order (j, k) and order (k, j). [for example, the number of judges who expressed a strong preference for A to B is (1), and the number of judges who expressed a strong preference for B to A is (3). The other part of Table I concerns $M_{jk,}$ M_{kj} $\bar{x}_{jk}$ $\bar{x}_{jk}$ as it is calculated according to Equations 1, 2, and 3, respectively. Hence, conducting the analysis of all possible pairwise mean preferences and average mean preference is called multiple comparisons (Chung & Hwang, 1978). (We used Minitab, software package, in our data analysis). Now, one of the objectives of this research is to find an estimate of the final preference scores Sj. (Nyarko, 2023). From the information in Table I, the Score of preference (S_j) for each of the six types of Potato chips that were presented to this taste testing experiment for comparison and evaluation in the order (j k) can be calculated using Eq. 6 as follows (David, 1987).

Let potato chips: A = 1 B = 2 C = 3 D = 4 E = 5 and F = 6, then

Preference score for A = S_1 = −0.23325
Preference score for B = S_2 = 0.37775
Preference score for C = S_3 = 0.12225
Preference score for D = S_4 = 0.24992
Preference score for E = S_5 = −0.22217
Preference score for F = S_6 = −0.29450

The Scores can be described also in Table III and Fig. 1 below:

TABLE III
Preference Score

Potato chips	Preference score
A	−0.23325
B	0.37775
C	0.12225
D	0.24992
E	−0.22217
F	−0.2945

TABLE IV
Ranks for Potato Chips Scores

Object/potato chips	Sj = score	Rank
B	0.37775	1
D	0.24992	2
C	0.12225	3
E	−0.22217	4
A	−0.23325	5
F	−0.29450	6

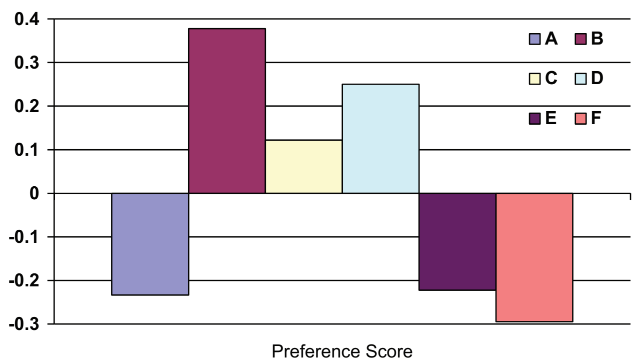

Fig. 1: Scores for potato chips.

The following Table IV describes the ranks of 6 brand of potato chips

Obviously, potato chips B is the most preferred variety (Ranked 1), followed by potato chips D, potato chips C, and so on. The least preferred type is potato chips F (Ranked 6) (Jameel, 2019). To find out whether there is a significant difference between the six preference Scores above (S_j) "as another goal" and another objective in this research, we resort to the method of Two-ways analysis of variance with through the calculation of the sums of squares that developed in Table II. as follows:

1. $SSt_c = 62(-3)^2+68(2)^2+53(1)^2+69(0)^2+60(1)^2+76(2)^2+61(3)^2 = 1805$
2. $SSs_j = 2(15)(6)[(-0.23325)^2+\ldots+(-0.2945)^2] = 73.9072$
3. $SSx_{jk} = 15[(0.200)^2+\ldots+(-0.600)^2] = 205.9417$
4. $SSm_{jk} = 2(15)[(-0.7)^2+\ldots+(0.5)^2] = 195.279$
5. $SS\gamma_{jk} = 195.279–73.9072 = 121.3718$
6. $SS_d = 205.9417–195.279 = 10.6627$
7. $SSE = 1805–205.9417 = 1599.058$

The sums of the above squares, their degrees of freedom, and the accompanying moral tests can be summarized in the following two-way variance analysis Table V.

10.24086/cuejhss.vol9n1y2025.pp183-187

TABLE V
Paired Comparison Two-Analysis of Variance Table

Source of variation	S.S=Sum square	D.F=Degrees of freedom	M.S=Mean square	F-ratio
Among objects (main effect)	S Ssj=73.9072	(L–1)=5	MSS=14.7814	MSS/MSE=3.8824**
Deviation from subtraction	S Sγ=121.3718	(L-1) (L-2)=20	M Sγ=6.0686	M S γ/MSE=1.5939*
Average mean preference	S S mjk=195.2710	L (L-1)/2=15	MSM=13.0186	MSM/MSE=3.4194**
Mean order effect	S Sd=10.6627	L (L-1)/2=15	MSD=0.7108	MSD/MSE=0.1867†
Mean Preference	S Sxjk=205.9417	L (L-1)=30	M S x=6.8647	M S x/MSE=1.8030**
Error	SSE	L (L-1) (n-1)=420	MSE=3.8073	
Total	SST	n L (L-1)=450		

*Significant **Highly Significant †Not Significant. D.F of Mean preference+D.F of Error=D.F Total=30+420=45

V. Conclusion

After reviewing the results in Tables IV and V above, and conducting the required statistical tests using the F-test at significance levels of $\alpha = 0.01$ and $\alpha = 0.05$, the following conclusions were drawn:

1. The most preferred potato chips among the six types of this brand are type B, followed by type D, type C, and so on. The least preferred type is F
2. The differences between the final preference scores (Sj) of the six types of potato chips are very significant. In other words, there are real differences between the scores, indicating that the potato chips vary notably in terms of taste
3. There are substantial differences between the mean preferences and the average mean preference
4. The average order effect (–d) represents, as expected, the mean difference in preference when potato chips are served in different orders. For example, when a pair (A, B) is presented to the judges, it is evaluated in both orders: (A, B) and (B, A)
5. The rank method was used to determine preference while accounting for serving order. It was found that the order of serving had no significant effect on the judges' decisions. This suggests consistency and stability in the judges' preferences
6. The interaction or co-effect between variables, as measured by the sum of squares, was found to be insignificant. This implies no meaningful difference between the original population data and the data collected from the judges using the seven-point scale
7. While there are very significant differences between the six types of potato chips, this does not determine which specific type is most preferred overall. Therefore, further analysis is needed to decide on future production quantities of the preferred type (Table IV).

References

Bradley, R.A. (1953), Some statistical methods in taste-testing and quality evaluation. *Biometrics*, 9(1), 22-38.

Bradley, R.A., & Terry, M.E. (1952). Rank analysis of incomplete block designs: The method of paired comparisons. *Biometrika*, 39(3/4), 324-345.

Bradley, R.A., & Terry, M.E. (1952). The rank analysis of incomplete block designs. *Biometrika,* 39(3/4), 324-345.

Bühlmann, H., & Huber, P.J. (1963). Pairwise comparison and ranking in tournaments. *Annals of Mathematical Statistics*, 34(2), 501-510.

Castura, J.C., Varela, P., & Næs, T. (2023). Investigating paired comparisons after principal component analysis. *Food Quality and Preference*, 106(3), 104814.

Chung, F.R.K., & Hwang, F.K. (1978). Do stronger players win more knock-out tournaments? *Journal of the American Statistical Association*, 73(363), 593-596.

Cowden, D.J. (1975). A method of evaluating contestants. *The American Statistician*, 29(3), 82-83.

David, H.A. (1987). Ranking from unbalanced paired comparison data. *Biometrika*, 74(2), 432-436.

Davidson, R.R. & Bradley, R.A. (1969). Multivariate paired comparisons: The extension of a univariate model and associated estimation and test procedures. *Biometrika*, 56(1), 81-95.

Davidson, R.R. (1970). On extending the Bradley-Terry model to accommodate ties in paired comparison experiments. Journal of the American Statistical Association, 65(329), 317-328.

Fienberg, S.E.& Larntz, K. (1976). Log linear representation for paired and multiple comparisons models. *Biometrika*, 63(2), 245-254.

Fowlkes, *et al*. (1988). The Fowlkes–Mallows statistic and the comparison of two independently determined dendrograms. *Canadian Journal of Fisheries and Aquatic Sciences*, 45(6), 971-975.

Glenn, W.A., & David, H.A. (1960). Ties in paired comparison using modified Thurstone-Mosteller model. *Biometrics*, 16(1), 86-97.

Guttman, L. (1946). An approach for quantifying paired comparisons and rank order. *Annals of Mathematical Statistics*, 17, 144-163.

Jameel, A.M. (2019). Proposed statistical model for scoring and ranking sport tournaments (Racquetball, Squash, and Badminton). Cihan University-Erbil Journal of Humanities and Social Sciences, 3(1), 15-19.

Massoudi, A., & Jalal Fatah, S. (2024). Deceptive point-of-sale marketing tactics impact on consumer purchase intentions with an attitude as a mediator. *Inquietud Empresarial*, 24(2), 1-27.

Massoudi, A.H., Agha, A.M., & Zamoum, K. (2025). Bridging AI Marketing and customer loyalty. *Cihan University-Erbil Journal of Humanities and Social Sciences*, 9(1), 106-113.

Mihálykó, É.O. (2024). Evaluating the capacity of paired comparison methods to aggregate rankings of separate groups. *Central European Journal of Operations Research*, 32, 109-129.

Mosteller, F. (1951). An experimental measurement of utility. *Journal of Political Economy*, 59(5), 371-380.

Nyarko, E. (2023). On the design of paired comparison experiments with application. *Applied and Interdisciplinary Mathematics*, 10(1), 2180873.

Rao, P.V. and Kupper, L.L. (1967). Ties in paired-comparison experiments: A generalization of the Bradley-Terry model. *Journal of the American Statistical Association*, 62, 194-204.

Scheffé, H. (1952). An analysis of variance for paired comparisons. *Journal of the American Statistical Association*, 47(259), 381-400.

Thurstone, L.L. (1927a). A law of comparative judgment. *Psychological Review*, 34, 273-286.

Thurstone, L.L. (1927b). The method of paired comparisons for social values. *Journal of Abnormal and Social Psychology*, 21, 384-400.

Wieren's, B. (1974). Paired comparison product testing-when individual preferences are stochastic. *Applied Statistics*, 23(3), 384-390.