

# Comparability of Multiple Rank Order and Paired Comparison Methods

James B. Rounds, Jr., Thomas W. Miller, and René V. Dawis  
University of Minnesota

Two studies were conducted to compare multiple rank order and paired comparison methods in terms of psychometric characteristics and user reactions. For both studies, stimuli from the Minnesota Importance Questionnaire (MIQ) were cast in multiple rank order and paired comparison forms and were administered to subjects on two occasions (test-retest) in a counterbalanced design. For the multiple rank order form, item blocks of three stimuli were used in the first study ( $N = 158$ , retest after one week), and item blocks of five stimuli in the second study ( $N = 280$ , retest after two days). Individual and group item responses, preference counts, and Thurstone normal transform scale values obtained by the multiple rank order method were found to be very similar to those obtained by paired comparisons. Administration time decreased as number of stimuli in the item block increased. Two-thirds of the subjects preferred the multiple rank order method. The equivalence of the two methods is discussed, along with suggestions for further research.

The method of paired comparisons, conceived by Fechner (Boring, 1929; David, 1963) and developed by Thurstone (1927a, 1927b) as a psychophysical method, has been frequently applied in conjunction with the law of comparative judgment to scaling "psychological" variables. The seriousness of crimes (Coombs, 1967; Thurstone, 1927c), patients' mental health

(Karon & O'Grady, 1970), visual illusions (Howard, Wagner, & Mills, 1973), life goals (Gulliksen, 1964), food preferences (Bradley, 1953; Edwards & Thurstone, 1952), and conservatism (Ekman & Kuennapas, 1963) exemplify stimuli that have been scaled with this method. The paired comparison method has also been used in personality measurement of needs (Edwards Personal Preference Schedule, Edwards, 1959; Minnesota Importance Questionnaire, Gay, Weiss, Hendel, Dawis, & Lofquist, 1971) and sociometric measurement of peer preferences (Cohen & Tassel, 1978; Wytrol & Thompson, 1953).

These studies and many others (Davidson & Farquhar, 1976), demonstrate that the method of paired comparisons (Bradley, 1976; Kendall, 1970; Edwards, 1957) provides an excellent approach to the obtaining of judgment and preference data. However, as the number of stimuli increases, the method becomes exceedingly time-consuming and laborious for the subjects. The number of judgments increases very rapidly as the number of stimuli increases. For  $n$  stimuli, the number of pairs,  $k$ , is  $[n(n-1)]/2$ , e.g., for  $n = 10$ ,  $k = 45$ ; for  $n = 20$ ,  $k = 190$ ; and for  $n = 30$ ,  $k = 435$ . Subjects often complain that paired comparison questionnaires are long and repetitive. Most research using paired comparisons has typically been limited to 15 stimuli or less.

---

*APPLIED PSYCHOLOGICAL MEASUREMENT*  
Vol. 2, No. 3 Summer 1978 pp. 413-420  
© Copyright 1978 West Publishing Co.

With less labor, a subject can produce the same rank order obtained by paired comparisons through directly ranking the stimuli (Bock & Jones, 1968). As the number of stimuli increases, the rank order design becomes more economical and more practical than paired comparisons. However, there is also a limit to the number of stimuli that is practicable to rank, ordinarily no more than 25. Unlike paired comparisons, the rank order method does not provide a measure of consistency (transitivity) of the rankings.

An approach to scaling that provides a measure of consistency and yet allows for the ranking of numerous stimuli is the use of the incomplete block design (Bradley & Terry, 1952; Durbin, 1951) or the method of multiple rank orders (Bock & Jones, 1968; Gulliksen & Tucker, 1961). Gulliksen and Tucker (1961) describe the procedures for using multiple rank orders in place of paired comparisons. In constructing a multiple rank order instrument, stimuli are grouped in the same way that experimental treatments are grouped into blocks in a balanced incomplete block design, i.e., into subsets that are not mutually exclusive but that systematically include common stimuli such that each stimulus is "paired" (appears in a subset) with every other stimulus an equal number of times. Thus, all possible paired comparisons can be reproduced from the multiple rank orders. For example, when a subject ranks 5 stimuli within a block, implicit in the ranking of the 5 stimuli are 10 paired comparisons. With appropriate grouping, 21 blocks of 5 stimuli each can be used to obtain all possible paired comparisons (210 pairings) for 21 stimuli. Thus, the multiple rank order method reduces the number of judgments required by the paired comparison method, thereby reducing the length of the instrument and administration time. Furthermore, the multiple rank order method permits the determination of response consistency from circular triads in the subjects' responses.

One might ask for empirical evidence that multiple rank orders and paired comparisons do indeed yield similar results. Does a multiple

rank order format reduce administration time, and if so, by how much? What proportion of subjects would prefer responding to a multiple rank orders format rather than to paired comparisons? Few studies have dealt with these questions. Slater (1965) examined the test-retest stability of complete paired comparisons, rank orders, and multiple rank orders with three and four stimuli in each item. In a series of experiments with British psychiatric patients, he found that the rank order of the patient's preferences for postcard and film-star pictures were unaffected by the method used.

The two studies reported below were designed to compare the multiple rank order method with the method of paired comparisons in terms of psychometric characteristics and user reactions. Since the paired comparison method is used to rank stimuli for groups as well as individuals, the analyses examined both group and individual response matrices. The subjects' reactions to the multiple rank order and paired comparison formats focused on preference for format, reasons for preference, and amount of time needed to complete the two formats.

Study I used a paired comparison (PC) format and a multiple rank order format with three stimuli in each block (MRO-3). Study II used a PC format and a multiple rank order format with five stimuli in each block (MRO-5). Both studies used stimuli from the Minnesota Importance Questionnaire (MIQ). Subjects' reactions to the different formats were investigated in Study II only.

## Study I

### Method

*Instruments.* Two questionnaires (PC and MRO-3) were constructed using the 20 stimulus statements of the MIQ (e.g., Ability Utilization: "I could make use of my abilities") and one additional stimulus statement (Autonomy: "I could plan my work with little supervision"). The scale names for the 21 stimuli are as follows: Ability Utilization, Achievement, Activity,

Advancement, Authority, Autonomy, Company Policies and Practices, Compensation, Co-workers, Creativity, Independence, Moral Values, Recognition, Responsibility, Security, Social Service, Social Status, Supervision-Human Relations, Supervision-Technical, Variety, and Working Conditions.

The PC form of the MIQ contained 210 paired comparison items, each stimulus statement being paired with each of the other 20 statements. For each item (or stimulus pair) the subject was asked to respond to the question, "Which statement is more important to you in your ideal job?"

The MRO-3 form contained 70 multiple rank order items (blocks) with three stimulus statements in each item. Each pair of stimulus statements was presented in one item only. For each item (or block of three statements) the subject was asked to respond to the question, "Which statement is the most important to you in your ideal job? Which statement is least important to you in your ideal job?" Both forms required a separate answer sheet.

*Procedure.* The subjects were 158 student volunteers from the introductory psychology course at the University of Minnesota. Each subject participated in two administrations, with a one-week interval between administrations. In a counterbalanced design, 70 subjects completed the PC form at both administrations; 41 subjects completed the MRO-3 form at both administrations; and 47 subjects completed the PC form at one administration and the MRO-3 form at the other (25 completing the PC form first and 22 completing the MRO-3 form first). These three groups will be referred to as the PC/PC, MRO-3/MRO-3, and PC/MRO-3 groups, respectively. Administration time in minutes was recorded at both administrations for subjects in the PC/PC and MRO-3/MRO-3 groups.

*Analysis.* Three different statistical analyses were conducted to investigate the stability of stimulus-pair choices, preference counts, and stimulus scale values resulting from the two methods. First, the stability of choices among

stimulus pairings across the two administrations was examined. The responses for each subject at the first administration were converted to a  $21 \times 21$  response matrix, with elements  $a_{ij}$ , where  $a_{ij} = 1$  if stimulus  $i$  was ranked over stimulus  $j$ ; otherwise,  $a_{ij} = 0$ . The 210 off-diagonal elements (the lower triangle of the matrix) were used in this analysis. When the PC form was used, each item response determined an off-diagonal element. When the MRO-3 form was used, each item response determined three off-diagonal elements. A similar matrix of elements  $b_{ij}$  was obtained for each subject at the second administration. The  $a_{ij}$  elements were compared with the corresponding  $b_{ij}$  elements. For each pairing of stimuli  $i$  and  $j$ , a comparison was in agreement if  $a_{ij} = b_{ij}$ . The proportion of comparisons in agreement was computed for each subject and served as a dependent variable in a one-way analysis of variance across the three groups.

The second analysis examined the stability of individual subject preference counts across the two administrations. For each subject, the preference count  $A_i$  for stimulus  $i$  at the first administration was the number of stimuli ranked lower than stimulus  $i$  and was obtained by summing  $a_{ij}$  across the  $j$  columns of the response matrix.

$$A_i = \sum_{j=1}^{21} a_{ij} \quad [1]$$

Similarly, the preference count  $B_i$  for stimulus  $i$  at the second administration was obtained for each subject by

$$B_i = \sum_{j=1}^{21} b_{ij} \quad [2]$$

Thus, for each subject the profile (vector) of preference counts  $A_i$  could be correlated with the profile of preference counts  $B_i$ . The profile correlation coefficients, transformed to Fisher's  $Z$  values (Hays, 1973), were then subjected to a

one-way analysis of variance across the three groups.

The third analysis examined the stability of group stimulus scale values for the three groups. Normal deviate transform scale values were computed following the standard methods described by Torgerson (1958, pp. 172-173) for each group at each administration. For the PC/PC and MRO-3/MRO-3 groups, the group stimulus scale values at the first administration were correlated with those for the second administration. For the PC/MRO-3 group, the group stimulus scale values for the PC form were correlated with those for the MRO-3 form.

### Results

Table 1 shows the group means and standard deviations for the proportion of stable stimulus-

pair choices. The mean differences across the three groups were not statistically significant [ $F(2,155) = 2.01, p > .05$ ]. These results show (1) a generally high level of stability, averaging better than 80%, for either form on retest after a short (one week) interval and (2) the equivalence of PC and MRO-3 forms in the results for the PC/MRO-3 group.

Table 1 also presents the results of correlating individual preference counts across the two administrations. Again, mean differences among the groups were nonsignificant [ $F(2,155) = 1.51, p > .05$ ]. Results for the PC/MRO-3 group show that preference counts for the PC form were equivalent to preference counts for the MRO-3 form.

Group stimulus scale values correlated across administrations were found to be highly stable over a one-week interval, none of the correla-

Table 1

Proportion of Stable Stimulus-Pair Choices, Individual Preference-Count Stability Correlations, Group Stimulus-Scale-Value Stability Correlations, and Administration Time, by Group (Study I)

Variable	PC/PC (N = 70)	MRO-3/MRO-3 (N = 41)	PC/MRO-3 (N = 47)
Proportion of stable stimulus-pair choices			
Mean	.808	.834	.807
SD	.092	.078	.076
Individual preference-count stability correlations			
Mean <sup>a</sup>	.894	.914	.887
SD <sup>a</sup>	.474	.258	.272
Group stimulus-scale-value stability correlations			
	.991	.986	.918
Mean administration time			
Test	29.6 min.	23.4 min.	
Retest	20.8 min.	18.0 min.	

<sup>a</sup>Calculated by using Fisher's Z transformation.

tions being less than .90. Again, results for the PC/MRO-3 group showed that the ranking of stimuli using the PC form was equivalent to ranking with the MRO-3 form.

The MRO-3 form required less time to complete than the PC form, over 6 minutes less on the first administration and almost 3 minutes less on the second administration.

## Study II

### Method

**Instruments.** Two questionnaires were used in this study. The PC form was the 20-stimulus MIQ (Gay et al., 1971), which contains 190 paired comparison items. The MRO-5 form used the same 21 stimuli employed in Study I. This form consisted of 21 multiple rank order items with five stimuli in each item block, each pairing of stimuli in a block occurring only once in the entire questionnaire. In responding to a multiple rank order item, the subject was asked to rank the stimuli in order of importance: "1 = most important to you in your ideal job, 2 = next most important," to "5 = least important to you in your ideal job." The PC form required a separate answer sheet; the MRO-5 form did not require a separate answer sheet.

**Procedure.** The subjects were 280 undergraduates at the University of Minnesota enrolled in an introductory psychology course. For participating in the study, all subjects received points toward their final course grade. Each subject participated in two administrations, with a two-day interval between administrations. In a counterbalanced design, a group of 73 subjects completed the PC form at both administrations; 73 subjects completed the MRO-5 form at both administrations; and 134 subjects completed the PC form at one administration and the MRO-5 form at the other (67 completed the PC form first, and 67 completed the MRO-5 form first). These three groups will be referred to as the PC/PC, MRO-5/MRO-5, and PC/MRO-5 groups, respectively. Administration time in minutes was recorded at the first administration

only. Subjects in the PC/MRO-5 group were asked which of the two forms they preferred and the reasons for their preference.

**Analysis.** The statistical analyses were the same as those performed for Study I, with the following exceptions: A  $20 \times 20$  rather than  $21 \times 21$  response matrix was used in the calculations. Responses to the PC form were recorded in a  $20 \times 20$  matrix. Responses to the MRO-5 form were first recorded in a  $21 \times 21$  matrix; then the 21st row and the 21st column (for the stimulus Autonomy) were eliminated from this matrix. Once in a  $20 \times 20$  matrix form, responses to the MRO-5 form could be compared with responses to the PC form. Also, omega-squared (Hays, 1973) and Scheffé pairwise contrasts were computed whenever significant *F*-ratios were obtained in the analysis of variance.

### Results

Table 2 presents the means and standard deviations of the proportion of stable stimulus-pair choices across testings for the three groups. Although mean differences among the three groups were statistically significant [ $F(2,277) = 15.08, p < .001$ ], the omega-squared value of .09 indicated that there was little association between group membership and the stability of stimulus-pair choices. Pairwise Scheffé tests showed that although the PC/PC and MRO-5/MRO-5 means did not differ significantly, the PC/MRO-5 mean was significantly different ( $p < .05$ ) from the PC/PC and MRO-5/MRO-5 means. Additional analyses indicated that these statistically significant mean differences may have been due to an order-of-administration effect. When the PC/MRO-5 data were examined from the standpoint of order of administration, analysis showed that the mean proportion of stable stimulus-pair choices was .797 ( $SD = .063$ ) for the PC-MRO (PC first, MRO second) group and .825 ( $SD = .053$ ) for the MRO-PC (MRO first, PC second) group. Furthermore, the Scheffé tests indicated that the MRO-PC and PC/PC means did not differ significantly, but the PC-MRO mean was significantly different ( $p$

< .05) from the PC/PC, MRO-5/MRO-5, and MRO-PC means.

Mean individual preference-count stability coefficients for the three groups are also shown in Table 2. A one-way analysis of variance indicated that the mean stability coefficients of the three groups differed significantly [ $F(2,277) = 17.06, p < .001$ ]; but again, the association was weak ( $w^2 = .10$ ). Pairwise Scheffé tests showed that the PC/PC and MRO-5/MRO-5 means did not differ significantly, but the PC/MRO-5 mean differed statistically ( $p < .05$ ) from both the PC/PC and MRO-5/MRO-5 means.

Analysis of the data for the PC/MRO-5 group showed that the preference counts for the PC forms were very similar to the preference counts for the MRO-5 forms. However, testing

for order effect showed that the PC-MRO mean stability coefficient (.852) differed from that for the MRO-PC group (.906). Again, pairwise Scheffé tests showed that the MRO-PC, PC/PC, and MRO-5/MRO-5 means did not differ significantly but that the PC-MRO mean was significantly different ( $p < .05$ ) from the three other means.

Table 2 also shows that group stimulus scale values for both forms were highly stable on two-day retest. Further, subjects completed the MRO-5 form in about half the time required to complete the PC form, but the range of completion times was wide for both forms.

Of the 134 subjects who completed both PC and MRO-5 forms, 44 (33%) preferred the PC form versus 90 (67%) who preferred the MRO-5

Table 2

Proportion of Stable Stimulus-Pair Choices, Individual Preference-Count Stability Correlations, Group Stimulus-Scale-Value Stability Correlations, and Administration Time, by Group (Study II)

Variable	PC/PC (N = 73)	MRO-5/MRO-5 (N = 73)	PC/MRO-5 (N = 134)
Proportion of stable stimulus-pair choices			
Mean	.842	.853	.811
SD	.064	.047	.058
Individual preference-count stability correlations			
Mean <sup>a</sup>	.931	.916	.883
SD <sup>a</sup>	.340	.330	.313
Group stimulus-scale-value stability correlations	.994	.995	.980
Administration time			
Mean	28.0	16.5	
SD	7.5	7.0	
Range	15 to 65 min.	8 to 31 min.	

<sup>a</sup>Calculated by using Fisher's Z transformation.

form. However, of those who preferred the PC form, it was preferred by proportionately more subjects (70%) when it was administered second than when it was administered first. Those who preferred the MRO-5 form were almost evenly divided in the reasons given for their preference. Fifty-seven percent stated that the task of ranking five need statements allowed for fuller expression of their preference; the other 43% chose the MRO-5 form because they found the PC form repetitive and boring. Those who preferred the PC form stated that judging between two stimulus statements was easier than ranking five stimulus statements.

### Conclusion

The results of the two studies described above show that paired comparison and multiple rank order are highly comparable methods. High stability of stimulus-pair choices and preference counts has been demonstrated on retest for paired comparison and multiple rank order forms alike. Similarly, multiple rank order profiles were as stable as paired comparison profiles over a one-week and two-day test-retest interval. Results for the PC/MRO-3 and PC/MRO-5 groups showed that the two methods yielded similar arrays of stimulus-pair choices and preference counts. Stimulus-pair choice and preference-count stability coefficients for the PC/MRO-3 and PC/MRO-5 groups approached the test-retest coefficients for their respective method groups. Finally, analysis of group response matrices showed that normal deviate transform scale values derived by the paired comparison and multiple rank order methods yielded equivalent rankings of the stimuli.

Other results of this study suggest several other conclusions. For the multiple rank order method, mean administration time decreased as the number of stimuli per item block increased. The MRO-5 form took less time for subjects to complete than the MRO-3 form; and the MRO-3 form, less time than the PC form. There is also

strong evidence that most subjects preferred multiple rank orders to paired comparisons, although an order-by-form effect was observed in Study II but not in Study I.

On the basis of the similarity of the group preference counts and Thurstone normal transform scale values, it can be concluded that at least for obtaining occupational reinforcer preferences, the two methods can be considered equivalent. For two types of stimuli, therefore, occupational reinforcer stimuli in this study and pictorial stimuli (Slater, 1965), the choice of methods can be decided by consideration of other matters, rather than by questions of equivalence of methods.

Before choosing multiple rank orders in place of paired comparisons, the following points should be considered. With some types of materials (e.g., taste or auditory stimuli), the task of ranking three or more stimuli may be more difficult than that of comparing two stimuli. The educational level or ability level of subjects is another important consideration. The multiple rank order format may be too difficult for some subjects, while the paired comparison format may be too simple or too repetitive for others. Finally, the internal validity of individual subject profiles is more easily assessed by using a paired comparison format. Circular triad information can be obtained from multiple rank orders, but this information is less complete than that obtained from a comparable paired comparison format. Further research (Gulliksen, 1975) is needed to investigate the comparability of total and stimulus circular triads generated by the multiple rank order and paired comparison methods.

Future studies comparing ranking methods or other methods of obtaining stimulus preferences should consider an alternative design to the counterbalanced one used in this study. The counterbalanced design assisted in controlling for order effects and provided baseline groups against which to evaluate results from the experimental groups. An alternative design could address the question of whether individual dif-

ferences in the experimental groups' dependent variables (e.g., stimulus preference counts across time) were due to the effects of individuals and/or methods, and/or to an individual-method interaction effect. Also, an individual's preference for type of method might have affected the stability of results over time. In retrospect, a repeated measures design with replications would be more appropriate for investigating the comparability of methods.

### References

- Bock, R. D., & Jones, L. V. *The measurement and prediction of judgment and choice*. San Francisco: Holden-Day, 1968.
- Boring, E. G. *A history of experimental psychology*. New York: Appleton-Century, 1929.
- Bradley, R. A. Some statistical methods in taste testing and quality control. *Biometrics*, 1953, 9, 22-38.
- Bradley, R. A. Science, statistics, and paired comparisons. *Biometrics*, 1976, 32, 213-232.
- Bradley, R. A., & Terry, M. E. The rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 1952, 39, 324-345.
- Cohen, A. S., & Tassel, E. V. A comparison of partial and complete paired comparisons in sociometric measurement of preschool groups. *Applied Psychological Measurement*, 1978, 2, 31-40.
- Coombs, C. H. Thurstone's measurement of social values revisited forty years later. *Journal of Personality and Social Psychology*, 1967, 6, 85-91.
- David, H. A. *The method of paired comparisons*. London: Griffin, 1963.
- Davidson, R. R., & Farquhar, P. H. A bibliography on the method of paired comparisons. *Biometrics*, 1976, 32, 233-244.
- Durbin, J. Incomplete blocks in ranking experiments. *British Journal of Psychology*, 1951, 4, 85-90.
- Edwards, A. L. *Techniques of attitude scale construction*. New York: Appleton-Century-Crofts, 1957.
- Edwards, A. L. *Manual for the Edwards Personal Preference Schedule*. New York: The Psychological Corporation, 1959.
- Edwards, A. L., & Thurstone, L. L. An internal consistency check for the method of successive intervals and the method of graded dichotomies. *Psychometrika*, 1952, 17, 169-180.
- Ekman, G., & Kuennapas, T. Scales of conservatism. *Perceptual and Motor Skills*, 1963, 16, 329-334.
- Gay, E. G., Weiss, D. J., Hendel, D. D., Dawis, R. V., & Lofquist, L. H. Manual for the Minnesota Importance Questionnaire. *Minnesota Studies in Vocational Rehabilitation*, 28, 1971.
- Gulliksen, H. Intercultural studies of attitudes. In N. Frederiksen & H. Gulliksen (Eds.), *Contributions to mathematical psychology*. New York: Holt, Rinehart, & Winston, 1964.
- Gulliksen, H. Personal communication, December 11, 1975.
- Gulliksen, H., & Tucker, L. R. A general procedure for obtaining paired comparisons from multiple rank orders. *Psychometrika*, 1961, 26, 173-183.
- Hays, W. L. *Statistics*. New York: Holt, Rinehart, & Winston, 1973.
- Howard, R. B., Wagner, M., & Mills, R. C. The superiority of the pair-comparisons method for scaling visual illusions. *Perception and Psychophysics*, 1973, 13, 507-512.
- Karon, B. P., & O'Grady, P. Quantified judgments of mental health from the Rorschach, TAT, and Clinical Status Interview by means of a scaling technique. *Journal of Consulting and Clinical Psychology*, 1970, 34, 229-235.
- Kendall, M. G. *Rank correlation methods* (4th ed.). London: Griffin, 1970.
- Slater, P. The test-retest reliability of some methods of multiple comparisons. *The British Journal of Mathematical and Statistical Psychology*, 1965, 18, 227-242.
- Thurstone, L. L. A law of comparative judgment. *Psychological Review*, 1927, 34, 273-286. (a)
- Thurstone, L. L. Psychophysical analysis. *American Journal of Psychology*, 1927, 38, 368-389. (b)
- Thurstone, L. L. The method of paired comparisons for social values. *Journal of Abnormal and Social Psychology*, 1927, 21, 384-400. (c)
- Torgerson, W. S. *Theory and methods of scaling*. New York: John Wiley & Sons, 1958.
- Wytrou, S. L., & Thompson, G. G. A critical review of the stability of social acceptability scores obtained with the partial-rank-order and the paired comparison scales. *Genetic Psychology Monographs*, 1953, 48, 221-260.

### Acknowledgments

The study was supported, in part, by a research service grant from the Minnesota Division of Vocational Rehabilitation. A grant for computer time was provided by the University of Minnesota Computer Center.

### Author's Address

René V. Dawis, Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, MN 55455.