UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ADA138810

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR- 84-0126
2. GOVT ACCESSION NO.:
3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): QUANTIFICATION OF SUBJECTIVE RATINGS THROUGH CONJOINT MEASUREMENT ANALYSIS
5. TYPE OF REPORT & PERIOD COVERED: Final Report
6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Dr. David E. Greene
8. CONTRACT OR GRANT NUMBER(s): AFOSR-82-0220
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Mechanical Engineering Department, The University of Texas at Austin, Austin, TX 78712
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 2313/D9, 61102F
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research/NL, Bolling AFB, DC 20332
12. REPORT DATE: November 8, 1983
13. NUMBER OF PAGES: 37
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):
15. SECURITY CLASS. (of this report): UNCLASSIFIED
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:
16. DISTRIBUTION STATEMENT (of this Report): Approved for public release; distribution unlimited.
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):

18. SUPPLEMENTARY NOTES:
19. KEY WORDS (Continue on reverse side if necessary and identify by block number): Conjoint Measurement Theory, Interval Scales, Perturbations
20. ABSTRACT (Continue on reverse side if necessary and identify by block number):

Conjoint measurement theory is examined through a prototype example in which a fighter aircraft is subjectively rated on two factors. As a first step, a multifactor ordinal scale is developed. This ordinal scale provides a meaningful measure of aircraft quality. Interval scales of aircraft quality are produced by the basic analysis of variance model and two conjoint measurement methods: delta scaling and the computer algorithm MONANOVA. These methods produce interval scales that differ by constant factors, as guaranteed by the

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) theorem for additive conjoint measurement. The interval scale does not appear to be an improvement over the ordinal scale in the prototype example. There is no assurance that a specific conjoint measurement model can be used to "improve" the data. Major changes in the interval scales are caused by small perturbations in the rating matrix.

AFOSR-TR- 84-0126

FINAL REPORT

QUANTIFICATION OF SUBJECTIVE RATINGS THROUGH CONJOINT MEASUREMENT ANALYSIS

AFOSR-82-0220

Prepared for:

AFOSR/XOT (Mini Grant Program)
Bolling AFB
Washington, D.C.

Prepared by:

Dr. David E. Greene
Mechanical Engineering Department
THE UNIVERSITY OF TEXAS AT AUSTIN
Austin, Texas 78712

A-1

ABSTRACT

Conjoint measurement theory is examined through a prototype example in which a fighter aircraft is subjectively rated on two factors. As a first step, a multifactor ordinal scale is developed. This ordinal scale provides a meaningful measure of aircraft quality. Interval scales of aircraft quality are produced by the basic analysis of variance model and two conjoint measurement methods: delta scaling and the computer algorithm MONANOVA. These methods produce interval scales that differ by constant factors, as guaranteed by the theorem for additive conjoint measurement. The interval scale does not appear to be an improvement over the ordinal scale in the prototype example. There is no assurance that a specific conjoint measurement model can be used to "improve" the data. Major changes in the interval scales are caused by small perturbations in the rating matrix.

CONTENTS

INTRODUCTION

Subjective assessments of aircraft quality and pilot workload have been shown to be practical, minimally intrusive, and not disruptive of primary task performance (Eggemeier, 1980).

Sheridan and Simpson (1979) developed a prototype subjective rating method for Instrument Flight Rules pilot workload modeled closely after the Cooper-Harper (1969) rating system which was developed to evaluate aircraft handling qualities. Rehmann et.al. (1983) developed a technique by which participants subjectively rated workload every 60 s during the task, using a mechanical device, to overcome bias effects in subjective ratings at the conclusion of the task. Wierwille and Conner (1983) evaluated twenty workload measures using a psychomotor task in a moving-base aircraft simulator. They concluded that well-designed subjective rating scales were among the best techniques for evaluating psychomotor load.

Conjoint measurement theory, which might best be described as a general modeling and scaling approach, has been used to develop multidimensional subjective measures of workload. Reid et.al. (1981) used conjoint measurement theory to construct interval level workload scales from ordinal rankings of combinations of levels on three contributory workload scales from ordinal ranking of combinations of levels on three contributory dimensions. The dimensions were time load, mental effort load, and psychological stress load. They concluded that their subjective assessment technique was sensitive to differences in workload on a critically unstable tracking task.

Donnell and O'Connor (1978) and Donnell (1979) applied conjoint measurement theory to fighter aircraft to develop an interval scale measure of aircraft system operability. System operability was assumed to represent a combination of a number of factors, including the amount of workload required by the operator in performance of a task, and the degree of subsystem technical effectiveness demonstrated in accomplishment of the task. In order to develop an interval scale of systems operability, Donnell and O'Connor (1978) developed separate four-point ordinal rating scales for pilot workload and system technical effectiveness. Using conjoint measurement and related scaling procedures, independent ratings by pilots on each of these two dimensions were eventually combined and converted into an interval scale of systems operability. Donnell (1979) subsequently modified the pilot workload and technical effectiveness scales and applied them to a different aircraft. An interval scale of systems operability was also derived in this application with use of the conjoint measurement technique.

Whereas the interval scale produced by conjoint measurement theory may be sensitive to changes in workload or changes in aircraft quality, it has not been demonstrated that the derived interval scale is an improvement over the ordinal scale. Furthermore, the ordinal rating scale itself, upon which inferences and scale development is made, may not be sufficiently precise.

OBJECTIVES

The primary objective of this research is to investigate whether conjoint measurement theory can be effectively used to improve the quantification of subjective ratings of aircraft quality, pilot workload, etc. In particular, the interval scale produced by conjoint measurement theory will be examined to determine if it is a true improvement over an ordinal scale. The structure of conjoint measurement theory and the effectiveness of its representations are examined through a prototype example of aircraft quality.

CONJOINT MEASUREMENT THEORY

Conjoint measurement theory is developed in the following primary references: Shepard (1962a, 1962b), Kruskal (1964a, 1964b), Luce and Tukey (1964), Kratz and Tversky (1971), and Baird and Noma (1978). Parts of the following description of conjoint measurement theory and associated algorithms are based upon Nygren (1981).

Conjoint measurement can be defined as the procedure whereby we specify for a given combination rule, the conditions under which there exist measurement scales for the dependent and independent variables, such that the order of the joint effects of the independent variables in the data are preserved by the numerical composition rule. Conjoint analysis can be defined as the procedure whereby the numerical scale values for the joint effects and the levels of the independent variables are obtained. One attempts to find the appropriate combination rule and, assuming the rule is valid, finds numerical functions that best fit the order of the joint effects in the data according to the specified rule.

The following discussion primarily concerns additive conjoint measurement in two factors. Luce and Tukey (1964) showed that given:

(1) the set $A = A_1 \times A_2$, where A_1 and A_2 are non-empty sets,

(2) a binary relation $\gtrsim$ on $A_1 \times A_2$ (where "$\gtrsim$" might be a relation such as "is preferred over", and

(3) if $< A_1, A_2, \gtrsim >$ is an empirical relational system that satisfies the axiom structure of a weak order (i.e., $\gtrsim$ is connected and transiti·ve), solvability, cancellation, and the Arc midean property, then,

THEOREM: There exist real-valued functions ϕ on A, ϕ_1 on A_1, and ϕ_2 on A_2 such that for all (a_1, a_2) and (b_1, b_2) in A

(i) $(a_1, a_2) \gtrsim (b_1, b_2)$ if $\phi(a_1, a_2) \gtrsim \phi(b_1, b_2)$,

(ii) $\phi(a_1, a_2) = \phi_1(a_1) + \phi_2(a_2)$,

(iii) if ϕ', ϕ_1', and ϕ_2' are any other functions which satisfy (i) and (ii) above, then there exist real numbers $\alpha > 0$, β_1, and β_2 such that $\phi_1' = \alpha\phi_1 + \beta_1$, $\phi_2' = \alpha\phi_2 + \beta_2$, and $\phi' = \alpha\phi + \beta_1 + \beta_2$.

Thus, given four axioms that require only ordinal properties in the data for the binary relation $\gtrsim$, we arrive at a theorem which guarantees the existence of functions ϕ, ϕ_1, and ϕ_2 such that numerical scale values can be assigned to the stimulus objects in such a way that (1) the order among objects is preserved, (2) the levels of the factors on which the stimuli vary combine in an independent and additive fashion, and (3) the numerical scales have interval properties.

In addition to the additive model in (ii), there is the multiplicative model

$$\phi(a_1, a_2) = \phi_1(a_1)\phi_2(a_2)$$

and for three or more factors there is a distributive model and a dual-distributive model.

Krantz and Tversky (1971) determined ordinal properties that are sufficient for each of the four basic models to hold. A necessary condition for the existence of an additive model for two factors is the independence of A_1 and A_2 (that is, A_1 is independent of A_2 and A_2 is independent of A_1). A_1 is said to be independent of A_2 whenever

$$(a_1, a_2) \gtrsim (b_1, a_2) \text{ if and only if}$$
$$(a_1, b_2) \gtrsim (b_1, b_2)$$

Thus, the independence of A_1 implies that if $a_1 \gtrsim b_1$ for one level in A_2, then this relation will hold for any other level in A_2.

Given the set of necessary ordinal properties, it is possible to evaluate each of the four basic models for a set of observations obtained from a factorial design. It is sufficient to require each subject to present only rank-order estimates to each of the stimulus combinations generated by combining levels of the factors. In most applications of conjoint measurement theory, it is the additive model that is of interest. However, even for an additive model with two factors of several levels, both the testing procedures for the ordinal properties and the scaling procedure for obtaining the numerical scale values become tedious if not impractical without the aid of a computer-based algorithm.

CONJOINT MEASUREMENT ALGORITHM

There are computer algorithms which test whether the necessary ordinal properties hold. Two primary programs are CONJOINT (Holt and Wallsten, 1974) and PCJM2 (Ullrich and Cumming, 1973). These programs test for independence among the factors and count the number of violations of the necessary ordinal properties, specifying the data cells involved in violations.

There are, in addition, computer algorithms which use nonmetric conjoint scaling techniques to fit a given data structure to an additive, multiplicative, distributive or dual-distributive model. Two conjoint scaling approaches considered in the present research are the delta-scaling method (see COOMBS, 1970) and the computer algorithm MONANOVA (MONotonic ANalysis of VAriance) developed by Kruskal (1965).

The delta scaling method is an algorithm for converting an ordinal scale to a scale with interval properties satisfying the conditions of additive conjoint measurement. In this approach, an additive representation is sought which satisfies inequalities specified by the orderings of preference of the elements in the set $A = A_1 \times A_2$. The solution is found using linear programming methods. Scale values for levels of the factors are found. If the number of factors and the levels are not large, this approach is ammenable to "hand" calculation.

MONANOVA is used to fit data to an additive model by way of a nonmetric scaling procedure. In MONANOVA, a monotonic transformation of the data is found which leads to estimates of the scale values for levels of the factors. These scale values are such that when they are combined by way of an additive composition rule, their joint effects best fit the original data. This process is now described. First, an initial configuration (an initial set of scale values for the levels of the factors) is generated or read by the program. After this configuration is normalized, a set of distances, d_j, is calculated based upon the additive rule. The program next uses a monotone regression subroutine to create disparities, $\hat{d}_j$, which are modified distances subject to the constraint that they are monotonically ordered in the same way as their corresponding raw data values. To estimate how well the model fits the data, a goodness-of-fit measure called STRESS is computed. STRESS is defined by

$$\text{STRESS} = \sqrt{\frac{\sum_{j=1}^{n}(d_j-\hat{d}_j)^2}{\sum_{j=1}^{n}(d_j-\bar{d})^2}}$$

where n is the number of stimuli and $\bar{d}$ is the mean distance value. If the model were to fit the data perfectly, then the STRESS would be zero. MONANOVA uses the method of gradients in an iterative fashion to find the configuration that will minimize STRESS. MONANOVA prints the final configuration coordinates (the scale values of the levels of the factors) and the history of the iterative procedure.

PROTOTYPE EXAMPLE

Conojoint measurement theory is examined through a prototype example.

A new fighter aircraft has been developed. We would like to quantify a test pilot's subjective ratings in order to compare this aircraft with other classes of fighter aircraft and to identify strengths and weaknesses.

Suppose that the aircraft is to be subjectively rated on two different factors: factor A and factor B. Factor A might be workload and factor B might be subsystem effectiveness. Let factor A have verbal descriptive levels a_1, a_2, and a_3 and factor B have verbal descriptive levels b_1, b_2, and b_3.

A MULTIFACTOR ORDINAL SCALE

Our objective is to convert a subjective rating pair - a rating on factor A and a rating on factor B - into an ordinal number that represents the overall quality of the aircraft. We will develop an ordinal scale of aircraft quality.

Four basic steps are required to convert subjective ratings into meaningful ordinal numbers so that a comparison between aircraft can be made.

First, a rating standard must be established by a group of pilots to which the test pilot belongs. Each pilot in the group is asked to rank order the cells (a_i, b_j) in the rating matrix for factors A and B in Fig. 1, using the integers 1 through 9. (The row-column arrangement in the rating matrix, i is the row and j is the column, is for mathematical convenience.) Suppose, for the purpose of discussion, that all pilots in the group decide upon the rank ordering given by the matrix M

$$M = \begin{bmatrix} 1 & 3 & 6 \\ 2 & 5 & 8 \\ 4 & 7 & 9 \end{bmatrix} \quad (1)$$

Thus, (a_3, b_3) is the most preferred and (a_1, b_1) is the least preferred. It is noted that rank ordering is a weak mathematical description; while (a_3, b_3) is preferred over (a_2, b_3), it is not known by how much. Observe that in M the pilot is expressing a preference in which he weighs the

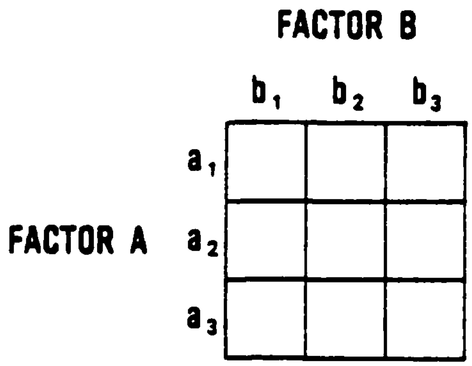

Fig. 1. Rating Matrix for Two Factor Ordering.

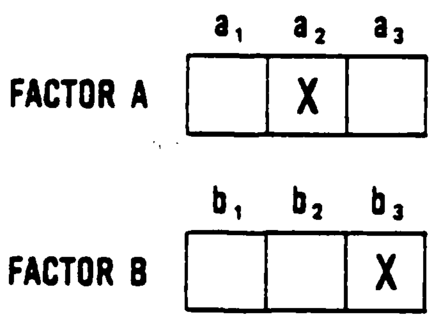

Fig. 2. Example Subjective Ratings for New Aircraft.

importance of one factor with another. Matrix M defines a transformation in which each verbal descriptive pair (a_i b_j) is mapped to a unique ordinal number; it produces an ordinal scale of aircraft quality.

Second, after having flown the new aircraft, the test pilot is asked to give subjective ratings for factors A and B. Suppose his ratings are those indicated by x in Fig. 2.

Third, the test pilot's subjective ratings are converted to a single ordinal number using the transformation determined in the first step. Thus, by the transformation M the aircraft quality is an 8.

Fourth, the ordinal number must be interpreted. This is simpler than in an independent factor by factor comparison. The new aircraft is "better" than another fighter aircraft with a 7, but "inferior" to one with a 9.

There are a number of disadvantages in this approach. First, it may be difficult to design a questionnaire such that there is general agreement in the pilot group on the rank ordering. Second, it may not be possible for the human to intelligently rank order several factors, each with a number of levels. Third, an aircraft may have a lower ordinal rating than another, but it is not immediately known which factor contributed most to the problem. Fourth, the ordinal numbers provide only a qualitative measure.

However, the composite multifactor approach, provided the questionnaire is properly designed and the transformation is generally agreed upon, can be expected to give a meaningful ordinal number by which aircraft can be simply and directly compared.

A MULTIFACTOR CONJOINT MEASUREMENT INTERVAL SCALE

We would like to convert the multifactor ordinal scale on factors A and B into numbers that have greater preciseness; we will use conjoint measurement theory to develop a scale of aircraft quality that has interval properties. In this analysis we will, in addition, prescribe numbers to verbal descriptions (that is, scale these descriptions) and model the aircraft quality as functions of the given factors.

Conjoint measurement theory has a subtle relationship with the analysis of variance. We begin by examining the underlying model for the analysis of variance.

We will assume in our initial analysis that the integers assigned to the cells of the rating matrix have precise mathematical meaning, rather than only expressing an ordering. Thus, the matrix M in (1) can be considered as data from a two-factor experiment.

We consider the underlying (deterministic) model for an analysis of variance of a two-factor experiment with three levels for each factor, no replication effects, and no random effects in the data. This model is given by

$$y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}, \qquad (2)$$

where y_{ij} is the observation taken at the ith level of factor A and the jth level of factor B, μ is the grand mean, α_i is the effect of the ith level of factor A, β_j is the effect of the jth level of factor B and $(\alpha\beta)_{ij}$ is the interaction of the ith level of factor A and the jth level of factor B. The model is further specified by

$$\sum_{i=1}^{3}\alpha_i = \sum_{j=1}^{3}\beta_j = \sum_{i=1}^{3}(\alpha\beta)_{ij} = \sum_{j=1}^{3}(\alpha\beta)_{ij} = 0. \qquad (3)$$

We identify y_{ij} with the elements of M and solve for the parameters in (2). The relationships in (2) and (3) provide seventeen equations that uniquely define the sixteen parameters. The parameters are

$$\mu = 5$$

$$(\alpha_1, \alpha_2, \alpha_3) = (-5/3, 0, 5/3) \qquad (4)$$

$$(\beta_1, \beta_2, \beta_3) = (-8/3, 0, 8/3)$$

$$(\alpha\beta)_{ij} = \begin{bmatrix} 1/3 & -1/3 & 0 \\ -1/3 & 0 & 1/3 \\ 0 & 1/3 & -1/3 \end{bmatrix}$$

We make two observations. First, according to the model there are interactive effects. Second, the model has assigned numbers α_i and β_j to the verbal descriptive levels a_i and b_j. These numbers give an indication of how "close" of how "far apart" the verbal descriptive levels are.

We illustrate what the model does. Suppose the interactive effects are identically zero, then the data would be represented by

$$y_{ij} = \mu + \alpha_i + \beta_j. \qquad (5)$$

This equation is the discrete version of the equation of a plane

$$y = \mu + \alpha + \beta \qquad (6)$$

defined on the α and β axes. Thus, in the absence of interactive effects, the model assigns numerical values to the levels of the factors A and B in such a way that the data values lie in the plane defined by (6). If there are interactive effects, then not all of the data values can be placed in this plane.

Suppose we drop the interactive terms in our representation of M in (2). We then have a new set of numbers y'_{ij}, which in general are different from the data values y_{ij}, defined by

$$y'_{ij} = \mu + \alpha_i + \beta_j, \quad . \tag{7}$$

with μ, α_i and β_j given in (4). We introduce the variables

$$\alpha'_i = \alpha_i + 5/3, \tag{8}$$

$$\beta'_j = \beta_j + 8/3, \tag{9}$$

$$y''_{ij} = y'_{ij} - 2/3, \tag{10}$$

into (7) to obtain

$$y''_{ij} = \alpha'_i + \beta'_j. \tag{11}$$

We denote the set of values y'_{ij} by T. Matrix T is given by

$$T = \begin{bmatrix} 0 & 2.67 & 5.33 \\ 1.67 & 4.33 & 7.00 \\ 3.33 & 6.00 & 8.67 \end{bmatrix} \tag{12}$$

(The elements of T are fractions; however, for the purpose of later comparisons, they are expressed as decimals.) Observe that while the matrix elements in T have different values than those in M, the rank ordering of the elements has been preserved.

We have in effect transformed the data in M to "new data" in T in such a way that the rank ordering is preserved and the new data satisfies an additive model. This transformation of data is illustrated in Fig. 3. Note that numbers have been assigned to the verbal descriptive levels (that is, they have been scaled) and each cell value in the new data matrix is the sum of the row and column scales.

What we have done is the essence of conjoint measurement; we have found a transformation of the data that preserves the rank ordering and such that a model (in this case, an additive model) holds. Through the transformation we have produced a scale of aircraft quality that has interval properties (the elements of T have precise numerical values while the elements of M, as originally defined, express only a rank ordering). We have also scaled the verbal descriptive levels of each factor. The model provides a simple law: the measure of aircraft quality is the sum of two scaled factors.

We transformed the data by neglecting the interaction terms in the deterministic model for the analysis of variance. The general approach in additive conjoint measurement, however, is to systematically search for the order preserving transformation such that the interactive effects are best reduced in the transformed data. Two conjoint measurement methods are now examined: delta-scaling and MONANOVA.

		FACTOR B		
		b_1	b_2	b_3
	a_1	1	3	6
FACTOR A	a_2	2	5	8
	a_3	4	7	9

→

		FACTOR B		
		0	2.67	5.33
	0	0	2.67	5.33
FACTOR A	1.67	1.67	4.33	7.00
	3.33	3.33	6.00	8.67

Fig. 3. "Old Data" Transformed into "New Scaled Data".

AN INTERVAL SCALE DERIVED BY DELTA SCALING

In the previous section we used the underlying model for the analysis of variance to transform the ordinal scale to an interval scale in such a way that an additive model held. We now make connections to the earlier presentation of conjoint measurement theory.

The assignment of preference to the cells (a_i, b_j) for factors A and B, given in the matrix M in (1), satisfies the axioms in the conjoint measurement theorem for an additive model. In particular, the factors A and B are independent. (The rows (columns) in M are all of the same rank order.) Thus, by the theorem, there exist real-valued functions ϕ on A x B, ϕ_1 on A, and ϕ_2 on B such that for all (a_i, b_j) in A, $\phi(a_i, b_j) = \phi_1(a_i) + \phi_2(b_j)$. Furthermore, any other functions ϕ', ϕ_1', and ϕ_2' which satisfy the theorem are related through linear transformations to ϕ, ϕ_1, and ϕ_2.

The basic theorem does not provide a procedure to obtain the functions ϕ, ϕ_1, and ϕ_2. However, in the previous section, we in effect constructed a set of functions that had the additive property. Thus, any other set of functions which we might determine by other methods must be related to the functions we have already found.

We now apply the delta scaling method to our subjective rating problem. The delta scaling method has been used in the assessment of fighter aircraft (Helm et.al., 1974). We use the "tally sheet" procedure in the cited reference, set all "d"s equal to unity, and obtain the matrix T'

$$T' = \begin{bmatrix} 0. & 7. & 14. \\ 5. & 12. & 19. \\ 9. & 16. & 23. \end{bmatrix} \tag{13}$$

As in T, each element in T' is the sum of the corresponding row and column scales. Each element in T' is a constant multiple of the corresponding element in T, as guaranteed by the theorem.

We have through conjoint measurement theory converted an ordinal scale to an interval scale in such a way that an additive model holds.

Delta scaling by hand computation is generally impractical. In further analysis the computer algorithm MONANOVA is used.

AN INTERVAL SCALE DERIVED BY MONANOVA

MONANOVA is an algorithm that seeks the monotone (hence, order preserving) transformation of the data such that the transformed data best fit an assumed linear model. It employs a gradient search. The degree to which the transformed data fit the model is indicated by the STRESS.

We apply MONANOVA to the ordinal matrix defined by M in (1). For an additive model, the MONANOVA computer algorithm produces, with zero STRESS, the matrix T" defined by

$$T'' = \begin{bmatrix} .00 & 1.47 & 2.94 \\ .92 & 2.39 & 3.86 \\ 1.84 & 3.32 & 4.77 \end{bmatrix} \qquad (14)$$

Scales for factor A and factor B are given in T" by the first column and first row, respectively. Each element in T" is the sum of the corresponding row and column scales.

As in T' in (13), each element in T" is a constant multiple of the corresponding element in T.

SENSITIVITY ANALYSIS

Through conjoint measurement we have converted an ordinal scale of aircraft quality to one that has interval properties. We will consider how the interval scale values change when there are small perturbations in the rank orderings.

We use as the unperturbed state the rank ordering in M_1

$$M_1 = \begin{bmatrix} 1 & 3 & 6 \\ 2 & 5 & 8 \\ 4 & 7 & 9 \end{bmatrix} \tag{15}$$

and as perturbed states M_1'

$$M_1' = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 6 & 8 \\ 4 & 7 & 9 \end{bmatrix} \tag{16}$$

and M_1''

$$M_1'' = \begin{bmatrix} 1 & 3 & 6 \\ 2 & 4 & 8 \\ 5 & 7 & 9 \end{bmatrix} . \tag{17}$$

Matrices M_1' and M_1'' are obtained from M_1 by interchanging two elements. The perturbed rank orderings are reasonable.

So that we can make comparisons, we "normalize" the computed interval scale so that the largest interval value is 100.

One would hope that the perturbations in the rank orderings would cause only local effects in the interval scale values. But this is not the case, as may be seen in Fig. 4. The perturbation given by M_1' caused a shift of the interval scale values to the right and the perturbation given by M_1'' caused a shift of the interval scale values to the left.

These general shifts make it difficult to interpret the interval scale values and cast doubt upon the validity of the scales. Even though the cell with the ordinal 8 was not perturbed, it was mapped through conjoint measurement to 80.9, 84.6 and 73.1 for M_1, M_1' and M_1'', respectively. Suppose it was believed that 75. was a "passing" score. Then an 8 or 9 would be passing if M_1 was the standard; a 7, 8, or 9 would be passing if M_1' was the standard; and only 9 would be passing if M_1'' was the standard.

We examine the perturbation effects when the number of levels is increased. We assume four levels for each factor.

We define the unperturbed state by M_2

$$M_2 = \begin{bmatrix} 1 & 3 & 6 & 10 \\ 2 & 5 & 9 & 13 \\ 4 & 8 & 12 & 15 \\ 7 & 11 & 14 & 16 \end{bmatrix} \qquad (18)$$

And the perturbed states by M_2'

$$M_2' = \begin{bmatrix} 1 & 3 & 6 & 9 \\ 2 & 5 & 10 & 13 \\ 4 & 8 & 12 & 15 \\ 7 & 11 & 14 & 16 \end{bmatrix} \qquad (19)$$

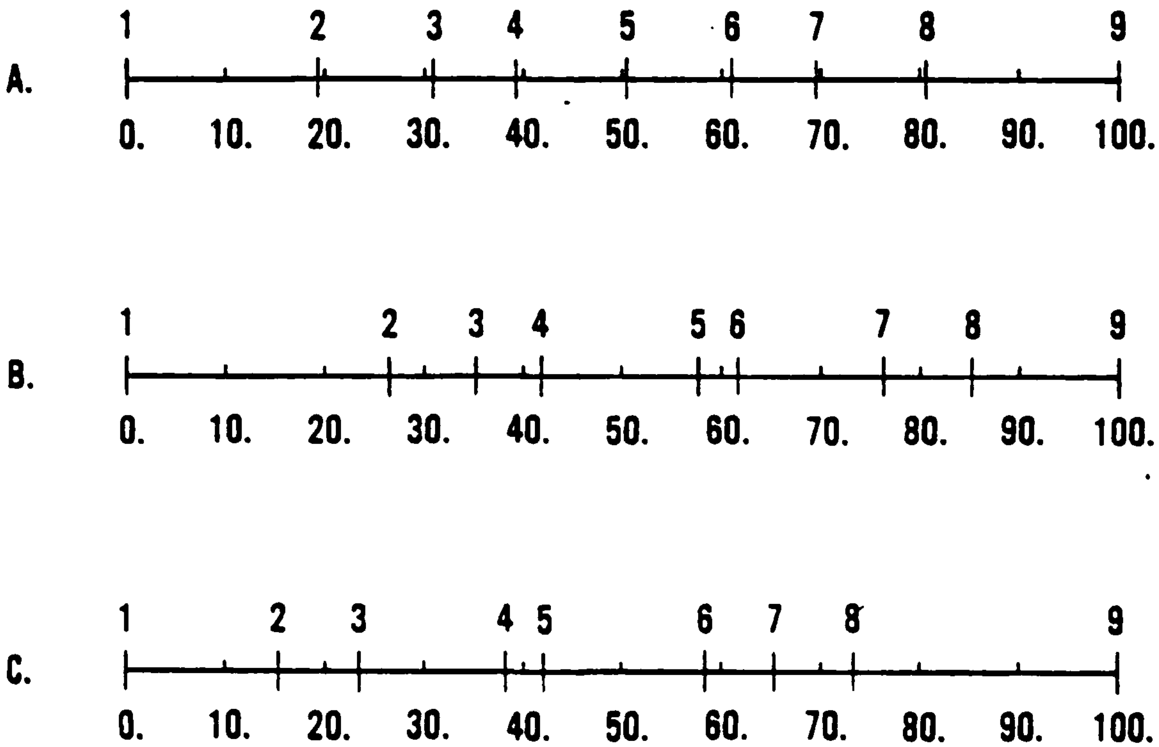

Fig. 4. Interval Scale Values Determined Through Conjoint Measurement for Ordinal Data in A. M_1, B. M_1', and C. M_1''.

and M_2''

$$M_2'' = \begin{bmatrix} 1 & 3 & 6 & 10 \\ 2 & 5 & 9 & 13 \\ 4 & 7 & 12 & 15 \\ 8 & 11 & 14 & 16 \end{bmatrix} \qquad (20)$$

The interval scale values determined through conjoint measurement for the ordinal data in M_2, M_2', and M_2'' are given in Fig. 5. While the perturbations still cause shifts in the interval scale values, they are not as pronounced as in the three-level case.

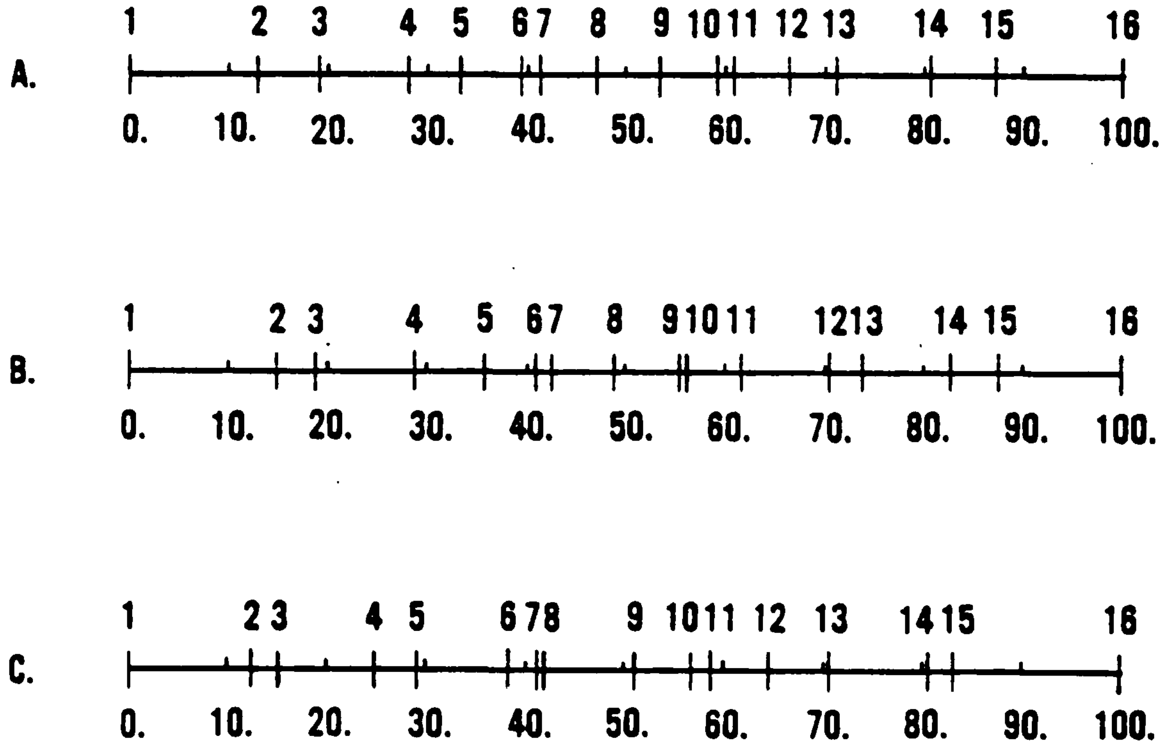

Fig. 5. Interval Scale Values Determined Through Conjoint Measurement for Ordinal Data in A. M_2, B. M_2', and C. M_2''.

SUMMARY

A subjective rating pair on factors A and B in the prototype example was converted into an ordinal number that represented the overall aircraft quality using four basic steps. First, a rating standard was established in which the factor pairs were rank ordered. Second, the pilot was asked to rate the new aircraft in the two factors. Third, the pilot's subjective ratings were converted into a single ordinal number using the transformation of the first step. Fourth, the ordinal number was interpreted, allowing comparisons between aircraft.

There are a number of disadvantages of this approach. First, there may not be general agreement in the pilot group on the rank ordering. Second, the human is limited in his capability to rank order several factors, each with a number of levels. Third, the role of a specific factor is disguised in the ordinal scale. Fourth, the ordinal numbers provide only a qualitative measure.

If, however, the questionnaire is properly designed and the rank ordering is generally agreed upon, this approach can be expected to produce a meaningful ordinal number by which aircraft can be simply and directly compared.

Conjoint measurement theory was introduced using the underlying model for the analysis of variance in a two-factor experiment. When the interactive terms in this model were dropped, the model produced an interval scale with additive properties. The verbal descriptive levels were scaled; that is, numbers were assigned to them.

The general approach in additive conjoint measurement theory is to systematically search for the order preserving transformation such that the interactive effects are best reduced in the transformed data. Two conjoint measurement methods were examined: delta scaling and MONANOVA.

The delta-scaling method produced an interval scale with additive properties. The associated transformation differed from that of the analysis of variance model by a constant factor as guaranteed by the conjoint measurement theorem for an additive model.

An interval scale was derived using the computer algorithm MONANOVA. MONANOVA is an algorithm that seeks the monotone transformation of data such that the transformed data best fit an assumed linear model. The associated transformation again differed by constant factors from those of the other two approaches.

A sensitivity analysis was made using MONANOVA. One might expect that "small" perturbations in the rank orderings would cause only local effects in the interval scale values. This was not the case, for when two elements were interchanged, large shifts occurred in the interval scales. When the number of levels was increased from three to four, the perturbation influence on the interval scale was reduced.

CONCLUSIONS AND RECOMMENDATIONS

The multifactor ordinal scale is a meaningful measure of aircraft quality. Like all subjective ratings, its validity depends upon whether there is a generally agreed upon rating standard and whether the test pilot's subjective ratings conform to the standard. These issues can be largely managed through the selection and training of test pilots and through the design of questionnaires.

An improvement in the ordinal scale could be made by allowing a more quantitative ordering; for example, by allowing equal or great-than-integer ordering. Also, it may be appropriate to assign scalar values to the cells of the rating matrix.

The primary limitation of the multifactor ordinal scale is the ability of the human to rank order several factors with several levels. It may be possible to develop mathematical models to assist the human in the rank ordering. For example, there may be enough information to determine a complete rank ordering if the human rank orders the levels of the factors and also rank orders the factors themselves.

The interval scale determined by conjoint measurement theory does not appear to be an improvement over the ordinal scale in the prototype example. To begin with, imprecise data is generated through the imposition of the weakness of rank ordering. (The assignment of scalar values to the cells of the rating matrix could make the data more precise.) This data in effect is then transformed to fit a model. This approach would be acceptable if the model reflected the "true world" and it acted to "improve" the data. However, there is no assurance that a specific model will improve the data; and thus the interval scales have questionable validity. The major changes in the interval scales caused by small perturbations in the rating matrix illustrate the difficulty in interpreting the interval scales.

Conjoint measurement methods should be explored as a basic modeling and scaling approach - along with other methods - to seek the constitutive laws of subjective ratings.

BIBLIOGRAPHY

Baird, J.C. and Noma, E. Fundamentals of Scaling and Psychophysics, New York, New York, John Wiley and Sons, 1978.

Coombs, C.H., Dawes, R.M. and Tversky, A., Mathematical Psychology, Englewood Cliffs, New Jersey, Prentice-Hall, 1970.

Cooper, G.E. and Harper, R.P., Jr., "The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities", NASA, TN D-5153, 1969.

Donnel, M.L., "An Application of Decision Analytic Techniques to the Test and Evaluation Phase of a Major Air System: Phase III, McLean, Virginia, Decisions and Designs, Inc., TR-PR-79-6-91, May 1979.

Donnell, M.L. and O'Connor, M.F., "The Application of Decision Analytic Techniques to the Test and Evaluation Phase of the Acquisition of a Major Air System: Phase II, McLean, Virginia, Decisions and Designs, Inc., TR-78-3-25, April 1978.

Eggemeier, F.T., "Some Current Issues in Workload Measurement", Proceedings of Human Factors Society 24th Annual Meeting, Santa Monica, CA; Human Factors Society, pp. 669-673, 1980.

Holt, J.O. and Wallsten, T.S., "A User's Manual for CONJOINT: A Computer Program for Evaluating Certain Conjoint-Measurement Axioms", Technical Report No. 42, University of North Carolina, L.L. Thurstone Psychometric Laboratory, October 1974.

Krantz, D.H. and Tversky, A., "Conjoint Measurement Analysis of Composition Rules in Psychology", Psychological Review, Vol. 78, pp. 151-169, 1971.

Kruskal, J.B., "Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis", Psychometrica, Vol. 29, pp. 1-27, 1964a.

Kruskal, J.B., "Nonmetric Multidimensional Scaling: A Numerical Method", Psychometrica, Vol. 29, pp. 115-129, 1964b.

Kruskal, J.B., "Analysis of Factorial Experiments by Estimating Monotone Transformations of the Data," Journal of the Royal Statistical Society, Series B, Vol. 27, pp. 251-263, 1965.

Luce, R.D., and Tukey, J.W., "Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement", Journal of Mathematical Psychology, Vol. 1, pp. 1-27, 1964.

Nygren, T.E., "Development of a Manual of Use for Conjoint Scaling Techniques", SCEEE Summer Faculty Research Progrram, August 1981.

Rehnmann, J.T., Stein, E.S., and Rosenberg, B.L., "Subjective Pilot Workload Assessment", Human Factors, Vol. 25, pp. 297-307, 1983.

Reid, G.B., Shingledecker, C.A., and Eggemeir, F.T., "Application of Conjoint Measurement to Workload Scale Development", Proceedings of Human Factors Society 25th Annual Meeting, Santa Monica, CA: Human Factors Society, pp. 522-526, 1981.

Shepard, R.N., "The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function I", Psychometrica, Vol. 27, pp. 125-140, 1962a.

Shepard, R.N., "The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function II", Psychometrica, Vol. 27, pp. 219-246, 1962b.

Sheridan, T.B., and Simpson, R.W., "Toward the Definition of the Mental Workload of Transport Pilots", Technical Report No., DOT-OS-70055, Flight Transportation Man-Machine Laboratory, Massachusetts Institute of Technology, MA., 1979.

Sheridan, T.B. and Stassen, H.G., "Definitions, Models and Measures of Human Workload. In: N. Moray (Ed.), Mental Workload: Its Theory and Measurement. New York: Plenum Press, 1979.

Ullrich, J.R. and Cumming, D.E., "PCJM: A Program for Conjoint Measurement Analysis of Polynomial Composition Rules", Behavioral Science, Vol. 18, pp. 226-227, 1973.

Wierwille, W.W. and Connor, S.A., "Evaluation of 20 Workload Measures Using a Psychomotor Task in a Moving-Base Aircraft Simulator, Human Factors, Vol. 25, pp. 1-16, 1983.