

# Multiple Hierarchical Analysis

Richard J. Hofmann  
Miami University

A multivariate method for exploratory scalogram analysis of binary data is developed—Multiple Hierarchical Analysis (MHA). This method, based on certain factor analytic methodology, is demon-

strated using an illustrative data set. It is concluded that the MHA methodology will be functional in identifying underlying independent and nonindependent scales in binary data.

In the analysis of binary data one particular type of multiple-variable arrangement is the hierarchy or scale. A group of variables, heterogeneous with regard to empirically based difficulty, may be aligned along a difficulty continuum. The variables will be hierarchical from the point of view that given the total number of positive responses made by an individual, it is possible to accurately predict the specific variables to which the individual responded positively. The type of analysis that uses this logic is referred to as scalogram analysis.

The methodology for scalogram analysis has undergone little change since its introduction by Guttman (1944, 1947, 1950). There are a number of methodological problems associated with scalogram analysis, many of them associated with the assumption that a set of variables being analyzed must necessarily form a single scale. In attempting to analyze binary data for scales, it is necessary to specify the variables of the scale a priori. Including in an analysis just one variable that does not belong with the other variables in a scalar sense can be disastrous with regard to the final appraisal of scale reproducibility. Logically, two scales may be ordered relative to each other; but their variables, when pooled together, may not form a scale. These are problems associated with the analysis of data defining more than one scale.

Although Lingoes (1963) has developed a method for the analysis of multiple scales, the intention here is to present an alternative method for identifying multiple scales or hierarchies in binary data. In particular, the proposed method will be referred to as Multiple Hierarchical Analysis (MHA). The method is primarily probabilistic, borrowing considerably from factor analysis. Aside from its present restriction to binary data, MHA only determines cumulative Guttman scales. (It should also be noted that Christoffersson, 1975, has developed a procedure for studying the multiple factor structure of bi-

nary variables; however, Christoffersson's procedure has no demonstrated relevance for scalogram analysis.)

In the sections to follow, the basic procedure for computing a scalar correlation is described. Using a modification of a principal factor solution, an oblique scalar factor solution is developed. The process of translating the scalar factor solution to a determinate scalogram solution is briefly described. Finally, an illustrative data set is discussed.

### An Index of Scalar Correlation

Assume two binary variables having different difficulties, where the proportion of positive endorsements for a variable is its difficulty. Using the Goodenough (1944) technique to determine reproducibility, it is possible to define an asymmetric two-variable reproducibility index. This index will represent the proportion of (+ +), (- -), and (+ -) response patterns for a given pair of variables, assuming the variable order is fixed and (+) represents a positive endorsement and (-) a negative endorsement. Alternatively, it will represent the complement of the proportion of (- +) or nonscalar response patterns. The reproducibility index between two variables will be asymmetric, since its value is dependent upon the ordering of the two variables. This index will always have a possible maximum value of 1.0; however, its lower bound will vary as a function of the variable marginals. It is possible to compute a Pearson product-moment correlation between two binary variables, the phi coefficient; but the maximum unsigned value attainable for a phi coefficient is a function of the variable marginals, whereas its lower unsigned bound is always zero.

There is a compromise between these two indices, the ratio of phi to its maximum possible magnitude, phi-max. This index has been discussed at length elsewhere with varying degrees of enthusiasm (Carroll, 1961; Cureton, 1959; Johnson, 1945; Loevinger, 1948). It is the positive phi-max that has received the most attention, being referred to as a coefficient of homogeneity by Loevinger (1948). Considering the range of values of the signed phi-max (see Guilford, 1965) and the implications of the magnitude of such an index, it is possible to use it as an *index of scalar relationship* that ranges from +1.0 to -1.0. A value of +1.0 will imply a perfect two-variable scale. A value of zero will imply virtually no scalar relationship between the two variables. A value of -1.0 determined between two variables will imply a perfect two-variable scale if the responses to one of the two variables are reflected.

Aside from Guilford's (1965) discussion of negative phi-max, which is computed differently than positive phi-max, there seems to be very little discussion of this index. To avoid confusion, the use of both positive and negative phi-max to index the magnitude of scalar relationship between two variables will be referred to as a scalar correlation,  $s$ . Unlike a matrix of product-moment correlations, the analogous matrix of scalar correlation,  $S$ , is not simply derived using matrix algebra.

Consider a two-variable scale defined by binary variables,  $i$  and  $j$ , where the variables' difficulties are  $d_i$  and  $d_j$ , respectively, and ( $d_i > d_j$ ). Then, if the product-moment correlation between  $i$  and  $j$ ,  $\phi_{ij}$ , is positive; the maximum possible correlation,  $\phi_{\max_{ij}}$ , is given by the well-known equation for  $\phi_{\max}$ :

$$\phi_{ij} \geq 0 \rightarrow \phi_{\max_{ij}} = [(d_j - d_j d_i) / (d_i - d_i d_j)]^{1/2} \text{ for } d_i \geq d_j. \quad [1]$$

Alternatively, if the product-moment correlation between  $i$  and  $j$  is negative, the maximum magnitude of the negative correlation, still referred to here as  $\phi_{\max_{ij}}$ , is given by the following equation:

$$\phi_{ij} \leq 0 \rightarrow \phi_{\max_{ij}} = [(1 - d_i)(1 - d_j)/d_i d_j]^{\frac{1}{2}} \text{ for } d_i \geq d_j$$

$$\text{or } d_i \leq d_j. \tag{2}$$

Now the scalar correlation,  $s_{ij}$ , between two variables,  $i$  and  $j$ , may be defined as

$$s_{ij} = (\phi_{ij}/\phi_{\max_{ij}}). \tag{3}$$

Assuming an easy-difficult, two-variable hierarchy, let  $c$  represent the observed frequency of scalar disconfirmations,  $(- +)$  patterns. The term  $(d_j d_i - s_{ij})/n$  will represent the frequency of scalar disconfirmations that should be expected by chance if  $i$  and  $j$  are independent. With a bit of algebra it can be demonstrated that for a positive scalar correlation

$$s_{ij} = 1 - nc/(d_j d_i - d_j); \phi_{ij} \geq 0 \tag{4}$$

is simply an index of the proportionate reduction of the chance scalar disconfirmations achieved by assuming the easy-difficult hierarchy  $ij$ , with  $i$  and  $j$  forming a two-variable Guttman scale. If the phi-index is negative, a reordering or reflection of the  $(+)$  and  $(-)$  for the more difficult of the two variables is suggested. With a proper variable reflection the pattern that is  $(- -)$  would become  $(- +)$ , and a  $(+ +)$  pattern would become  $(+ -)$ . Let  $e$  represent the observed frequency of scalar disconfirmations associated with two negatively correlated but unreflected variables, a  $(- -)$  pattern. Assuming  $i$  and  $j$  to be independent, the term  $(1 - d_i)(1 - d_j)/n$  will represent the frequency of scalar disconfirmations that should be expected by chance. Following an algebraic derivation similar to that used to derive Equation 4, a negative scalar correlation

$$-s_{ij} = 1 - ne/[(1 - d_i)(1 - d_j)]; \phi_{ij} < 0 \tag{5}$$

is simply a negatively signed index of the proportionate reduction of the chance scalar disconfirmations achieved when the responses of the more difficult variables are reflected to form a new easy-difficult, two-variable Guttman scale,  $ji$ .

Those variables having a scale in common will have large scalar correlations. Those variables not having any scales in common will have near zero scalar correlations. Such patterns are similar to those expected in a Pearsonian correlation matrix whose intercorrelations are defined by more than one common factor.

### The Whittle Iteration

Although it would appear to be quite compelling to determine the analog of a principal factor solution of the scalar intercorrelation matrix,  $S$ , to identify "scalar factors," there are some undesirable properties associated with the scalar intercorrelation matrix. Variables with extreme marginals will tend to have high scalar correlations, approaching 1.0, with a number of other variables. Also, when a perfect or near perfect Guttman scale does exist within the data, many of the variables associated with the scale will have some scalar intercorrelations approaching 1.0. The consequence of

these high scalar intercorrelations is that a scalar intercorrelation matrix will almost always be non-Gramian.

Non-Gramian matrices have negative eigenvalues and inflated positive eigenvalues. However, it is possible to adjust the inflated eigenvalues by using a modification of some methodology developed by Whittle (see Pruzek, 1977). Assume that  $\mathbf{S}$  is of order  $p$ . Then, the principal axis factoring of  $\mathbf{S}$  will determine  $p$  eigenvalues, the  $p$  diagonal entries of  $\mathbf{M}$ , associated with the  $p$  unit-length eigenvectors  $\mathbf{Q}$ . The sum of the eigenvalues will be  $p$ , and if there are negative eigenvalues, the sum of the positive eigenvalues will be greater than  $p$ . The modified Whittle procedure is an iterative procedure developed to adjust the positive eigenvalues such that their sum is precisely  $p$ . The difference between the sum of the positive eigenvalues and  $p$  represents the total inflation of the eigenvalues. This difference divided by the number of positive eigenvalues will represent the average inflation.

The average inflation may be subtracted from each positive eigenvalue to define a new set of adjusted eigenvalues. Usually several iterations will be sufficient to adjust the eigenvalues. At any iterative stage,  $k$ , there will be  $x$  positive eigenvalues. The number of positive eigenvalues may be reduced slightly as the number of iterations,  $k$ , increases. At each iterative stage the average eigenvalue inflation,  $\theta_{k+1}$ , is determined and then subtracted from the positive eigenvalues. These adjusted eigenvalues,  ${}_{k+1}\mathbf{M}^*$ , are then summed to determine the new average inflation,  $\theta_{k+2}$ :

$$\theta_{k+1} = \left( \sum_{i=1}^x k m_{ij}^* - p \right) / x \quad [6]$$

where

$${}_{k+1}m_{ij}^* = k m_{ij}^* - \theta_{k+1}. \quad [7]$$

The average inflation will quickly tend toward zero such that within two or three iterations, the average inflation is zero ( $\theta_i = 0.00$ ) and the positive adjusted eigenvalues,  $\mathbf{M}^*$ , will sum to  $p$ .

### A Scalar Factor Solution Following A Break-Point Analysis

In determining the number of scalar factors, the subjective rule that seems to work well in practice is based upon the scree test as defined by Cattell and Jaspers (1967). Cattell and Jaspers, in discussing the scree test, noted that when plotting the eigenvalues, the first values drop dramatically in size and quite suddenly the values taper off. The point of the tapering-off is referred to as a break in the curve defined by the eigenvalues as they are plotted in descending order. More specifically, this break is associated with the major point of inflection in the curve defined by the eigenvalues.

Frequently, it is possible to objectify the procedure for finding this major point of inflection. First, normalize the eigenvalues through division by  $p$ . Determine the difference between each set of consecutively decreasing normalized eigenvalues and call this the "first difference." Determine the difference between each set of consecutive first differences and call this the "second difference." For the type of curve typically associated with eigenvalues, the first second difference that approximates zero, or some epsilon criterion, say  $\epsilon = .05$ , will be associated with the major point of inflection. Algebraically, the number of scalar factors,  $r$ , can usually be determined by the following inequality:

$$r = j \rightarrow |m_j^*/p - 2m_{j+1}^*/p + m_{j+2}^*/p| \leq \epsilon \quad [8]$$

where  $j$  simply refers to the eigenvalue in descending order and ranges in possible value from 1.0 to  $(x - 2)$ , the number of positive eigenvalues minus two. For convenience of discussion any value determined by Equation 8 will be called a break point. When break point  $j$  is approximately zero, the point of inflection will be associated with eigenvalue  $j$ . Equation 8 seems to work quite well in practice, although it does on occasion "miss." Thus, the break point should always be checked with a visual inspection of the curve defined by the eigenvalues, especially with data defining a relatively small number of eigenvalues.

Because of the serious consequences of including a variable in a scale when it should be excluded from the scale, it is imperative that the scalar factor structure be quite clear, the analog of simple structure. Ordinarily, the clearest simple structure is obtained with an oblique solution. A  $p \times r$  scalar factor solution,  $F$ , is then defined as an oblique simple structure transformation of the principal axis scalar factor solution determined with the  $r$  adjusted eigenvalues, where  $r$  corresponds to the ordered number of the eigenvalue associated with the major point of inflection in the eigenvalue curve.

$$F = AM^{1/2}V, \tag{9}$$

where  $V$  is an oblique transformation matrix determined by the orthotran solution (Hofmann, 1978) or some other simple structure transformation procedure. The  $r \times r$  intercorrelation matrix of the scalar factors,  $L$ , is given by

$$L = D^{-1}(V'V)^{-1}D^{-1}, \tag{10}$$

where  $D^2$  is the extracted diagonal of  $(V'V)^{-1}$  with pre- and post-multiplication of  $(V'V)^{-1}$  by  $D^{-1}$ , resulting in a normalized matrix  $L$ .

The entries within the columns of  $F$  may be thought of as partial scalar correlations between the variables and scalar factors. The  $f_{ij}$  entry is the correlation between the  $i^{\text{th}}$  variable and the  $j^{\text{th}}$  scalar factor with the effects of the other  $(r - 1)$  scalar factors partialled out, the analog of a reference structure solution. Ideally, a factor solution where all substantive loadings are positive should be obtained, a positive manifold solution. If a variable has a substantial negative loading on a scalar factor, and all of the other definitive variables have positive loadings on the scalar factor, then the responses for that variable should be reflected when it is used in conjunction with the other variables associated with that particular scalar factor (see the discussion regarding Equation 5.) This single variable reflection does not preclude the use of a total scalar factor reflection analogous to a total factor reflection in factor analysis.

The scalar factor intercorrelation matrix is indicative of the degree of relationship between the scalar factors and is analogous to a primary intercorrelation matrix. Much of the rich factor analytic interpretation of the factor intercorrelation matrix does not readily generalize to the scalar factor intercorrelation matrix. The use of such concepts as factor scores and second-order analyses does not appear to be applicable at this time.

### Converting a Scalar Factor Solution to a Determinate Independent Scale Solution

The variables associated with a scalar factor may be subjectively identified as those variables with scalar factor loadings greater in magnitude than .30. These variables can then be grouped together for each scalar factor, noting variable reflections, and then analyzed from the raw data using an a

priori determinate scalogram analysis. Using each individual's determinate scale score, it is possible to compute the matrix of observed scale intercorrelations.

In factor analysis, variables are either factorially simple, variables defined by a single factor, or factorially complex, variables defined by more than one factor (Hofmann, 1977). A factor solution totally comprised of factorially simple variables is referred to as an independent cluster solution (see Harris & Kaiser, 1964, for an algebraic explication of such a solution). There seems to be no special name other than nonindependent cluster for a factor solution comprised of a mixture of factorially simple and factorially complex variables (Harris and Kaiser, 1964, discuss a solution referred to as "A' A proportional to L" that is an algebraic explication of the nonindependent cluster solution).

With regard to the scalar factor matrix, there will be variables that are factorially simple and there can be variables that are factorially complex. A scalar factor solution comprised totally of factorially simple variables will define independent scales such that the variable content of each scale is independent of the variable content of every other scale. A scalar factor solution comprised of factorially simple and factorially complex variables can define both independent and nonindependent scalar factors. Although an independent scale is defined by factorially simple variables, a nonindependent scale is defined primarily by variables that are factorially simple and several variables that are factorially complex. Nonindependent scales will have at least one, if not more than one, variable in common with some other nonindependent scale.

Following Hofmann (1977), the factor complexity of the  $i^{\text{th}}$  variable,  $c_i$ , defined by the scalar factor matrix,  $F$ , is

$$c_i = \frac{\left( \sum_{j=1}^r f_{ij}^2 \right)^2}{\sum_{j=1}^r f_{ij}^4} \quad [11]$$

The present technology associated with determinate scalogram analysis is not capable of providing a meaningful summary of nonindependent scales. Implicit, then, in most scalogram analyses is the assumption that all scales are independent. A factorially complex variable should not ordinarily be included in an independent scale, as its variable reproducibility will usually be low and it will adversely affect the other variable reproducibilities in the scale and, ultimately, the total scale reproducibility. An exception would be a variable with an extreme difficulty, as such variables usually will fit into any scale. Technically, a factorially complex variable should be excluded from any independent scale and should, perhaps, be refined in order to more properly portray the salient feature of the scalar factor on which it has its largest loading.

### Illustrative Example

Tucker (1964) has developed an artificial binary data set with a clearly established scalar structure. This structure appears not to be recoverable using traditional factor analytic procedures. The original Tucker data represents 30 observations on 9 binary variables reported as the first 9 data columns of Table 1. These 9 variables define 3 perfect independent but correlated scales. Thus, the data have an underlying dimensionality of 3. The 10th data column of Table 1 represents an additional variable purposely constructed for illustrating the use of the MHA methodology. In the discussion to follow, the 9 Tucker variables (original Tucker data) will first be analyzed, and then the 9 Tucker variables will be reanalyzed in conjunction with the 10th variable (modified Tucker data).

Downloaded from the Digital Conservancy at the University of Minnesota, <http://purl.umn.edu/93227>.

May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, <http://www.copyright.com/>

Table 1  
 Illustrative Data Set: Raw Binary Data Matrix Illustrative  
 of Multiple Correlated Scales (Adapted from Tucker, 1964)

Respondent	Variable									
	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	0	0	0	0	1
2	1	0	0	0	0	0	0	0	0	1
3	1	0	0	0	0	0	0	0	0	1
4	1	1	0	0	0	0	1	0	0	1
5	1	1	1	0	0	0	0	0	0	1
6	1	1	1	0	0	0	1	0	0	1
7	1	1	1	0	0	0	1	1	0	1
8	1	1	1	0	0	0	1	1	1	1
9	0	0	0	1	0	0	0	0	0	1
10	1	0	0	1	0	0	0	0	0	1
11	1	0	0	1	0	0	1	0	0	1
12	1	1	0	1	0	0	1	1	0	1
13	1	1	0	1	0	0	1	1	0	1
14	1	1	1	1	0	0	1	1	0	1
15	1	1	1	1	0	0	1	1	1	1
16	0	0	0	1	1	0	0	0	0	0
17	0	0	0	1	1	0	1	0	0	0
18	1	0	0	1	1	0	1	0	0	0
19	1	0	0	1	1	0	1	1	0	0
20	1	1	0	1	1	0	1	1	0	0
21	1	1	0	1	1	0	1	1	1	0
22	1	1	1	1	1	0	1	1	1	0
23	0	0	0	1	1	1	0	0	0	0
24	0	0	0	1	1	1	1	0	0	0
25	0	0	0	1	1	1	1	1	0	0
26	0	0	0	1	1	1	1	1	1	0
27	1	0	0	1	1	1	1	1	0	0
28	1	0	0	1	1	1	1	1	1	0
29	1	1	0	1	1	1	1	1	1	0
30	1	1	1	1	1	1	1	1	1	0

In Table 2 the product-moment intercorrelations of the variables are reported in the upper right and the scalar correlations are reported in the lower left triangular portion of the matrix.

**Solution for Original Tucker Data**

The first 9 variables were initially treated as a single data set. The eigenvalues and adjusted eigenvalues of the 9 × 9 scalar intercorrelation matrix are reported in Table 3, along with the break points.

Table 2  
Pearson Intercorrelations of Tucker Data (Upper Right Triangular Portion) and Scalar Intercorrelations (Lower Left Triangular Portion)

Variable	Variable									
	1	2	3	4	5	6	7	8	9	10
1		.60	.36	-.19	-.30	-.32	.32	.30	.19	.30
2	1.00		.60	-.30	-.33	-.30	.30	.33	.30	.33
3	1.00	1.00		-.32	-.30	-.19	.19	.30	.32	.30
4	-.53	-.50	-.32		.60	.36	.32	.30	.19	-.60
5	-.50	-.33	-.18	1.00		.60	.30	.33	.30	-1.00
6	-.32	-.18	-.07	1.00	1.00		.19	.30	.32	.60
7	.32	.50	.53	.32	.50	.53		.60	.36	-.30
8	.50	.33	.50	.50	.33	.50	1.00		.60	-.33
9	.53	.50	.32	.53	.50	.32	1.00	1.00		-.30
10	.50	.33	.50	1.00	-1.00	-.36	-.50	-.33	-.18	

The third break point is equal to an epsilon of .05, suggesting that a maximum 3 scalar factors be retained.

Following Hofmann's (1978) orthotran solution, an oblique scalar structure solution was determined (Table 4). When arbitrarily retaining all variables with loadings of .30 or greater as salient, the variable groupings were quite distinct: 1, 2, 3; 4, 5, 6; 7, 8, 9. These groupings are the precise groupings built into the data by Tucker (1964).

Also apparent in Table 4 are the relatively low factor complexities (Hofmann, 1977). The largest complexity is 1.46, which suggests that the most complex variables, 1 and 4, were defined at most by 1.46 scalar factors. Using the three distinct factorially simple variable groupings defined by the scalar

Table 3  
Original Tucker Data: Observed Positive Eigenvalues, Adjusted Positive Eigenvalues, and Break-points ( $\epsilon = .05$ )

	Positive eigenvalues		Break-points
	Observed	Adjusted	
1	4.48	4.36	.32
2	4.00	3.88	.32
3	0.67	0.55	.05
4	0.26	0.13	
5	0.21	0.09	
$\theta$	.12	0.00	



Table 4  
Solution of Original Tucker Data: Oblique Scalar Structure Solution

Variable	Scalar factors			Variable difficulty	Factor complexity
	1	2	3		
1	.50	-.19	.14	.73	1.46
2	.63	.00	.00	.50	1.00
3	.74	.18	-.11	.27	1.16
4	-.19	.50	.14	.73	1.46
5	.00	.63	.00	.50	1.00
6	.18	.74	-.11	.27	1.16
7	.11	.11	.49	.73	1.20
8	.00	.00	.60	.50	1.00
9	-.07	-.07	.67	.27	1.04

factor solution, three perfect Guttman scales can be defined (reproducibility = 1.00 for each scale). In practice, such perfection is not ordinarily obtained. The intercorrelations of the scalar factors and the intercorrelations of the real approximations to the scalar factors are reported as the lower and upper triangular portions, respectively, of the matrix reported in Table 5. It is interesting to note the pattern of similarity between the intercorrelations of the scalar factors and the real approximations to the scalar factors. Even for data defining perfect scales, the two sets of intercorrelations are not identical. The similarity between these two sets of intercorrelations will become even less apparent as the data define scale reproducibilities that depart from perfection.

**Solution for Modified Tucker Data**

The original 9 variables were reanalyzed with the 10th variable included. Unlike the other 9 variables, the 10th variable was factorially complex. The eigenvalues and adjusted eigenvalues along with the break points are reported in Table 6. The second break point was considerably greater than an

Table 5  
Original Tucker Data: Intercorrelations of Scalar Factors (Lower Left Triangular Portion) and Real Approximations to Scalar Factors (Upper Right Triangular Portion)

Scale	Scale		
	1	2	3
1		-.42	.42
2	-.35		.42
3	.45	.45	

Table 6  
 Modified Tucker Data: Observed Positive Eigenvalues,  
 Adjusted Positive Eigenvalues, and Break-points  
 ( $\epsilon = .05$ )

	Positive eigenvalues		Break-points
	Observed	Adjusted	
1	4.90	4.74	.31
2	4.34	4.19	.34
3	.73	.58	.01
4	.53	.38	.03
5	.23	.08	
6	.18	.03	
$\theta$	.15	0.00	

epsilon criterion of .05, while the third break point was much less than .05, thereby suggesting that three scalar factors be retained.

Again following Hofmann's (1978) orthotran solution, an oblique scalar structure solution was determined (Table 7). Unlike the 9-variable solution, the factor complexities of some variables were quite large. Variables 1 and 4 were more complex than they were in the 9-variable solution and the complexity of Variable 6 increased from 1.16 to 1.61. Variable 10, purposely constructed to be complex, had a large complexity of 1.98. When determining the salient variables in a fashion similar to that used with the 9-variable solution, the groupings were not so distinct. Variables 1, 2, and 3 grouped together along with Variable 6 and Variable 10. Variables 4, 5, and 6 grouped together as

Table 7  
 Solution of Modified Tucker Data: Oblique Scalar  
 Structure Solution

Variable	Scalar factor			Variable difficulty	Factor complexity
	1	2	3		
1	.39	-.12	.16	.73	1.54
2	.43	-.03	.11	.50	1.15
3	.62	.20	-.08	.27	1.24
4	-.16	.35	.11	.73	1.61
5	-.06	.42	.05	.50	1.07
6	.33	.72	-.28	.27	1.74
7	.00	.00	.49	.73	1.00
8	.01	.00	.48	.50	1.02
9	-.05	.08	.55	.27	1.07
10	.52	.07	-.43	.50	1.98

Table 8  
 Modified Tucker Data: Intercorrelations of Scalar Factors (Lower Left Triangular Portion) and Real Approximations to Scalar Factors (Upper Right Triangular Portion)

Scale	Scale		
	1	2	3
1		-.46	.09
2	-.49		.66
3	.35	.53	

they did in the 9-variable solution. Variables 7, 8, and 9 grouped together along with the reflected Variable 10.

With the exception of Scalar Factor 3, it appears that the scalar factors defined nonindependent scales. The Guttman reproducibility for the variable grouping (1, 2, 3, 6, 10), Scalar Factor 1, was only .83. For the variable grouping (7, 8, 9, 10), Scalar Factor 3, the Guttman reproducibility was .88. Clearly, for the grouping (4, 5, 6), Scalar Factor 2, the Guttman reproducibility was still 1.0. (Note that all reproducibilities were computed according to the conservative Goodenough, 1944, technique.)

The comparability of the scalar factor intercorrelations to the observed scalar intercorrelations (Table 8) was severely distorted for the modified Tucker data. Indeed, only the signs of the relationships were comparable. Without doubt, Variable 10 distorted the scalar factor solution. If it is assumed that each variable should be associated with a single scalar factor, then Variable 6 would clearly be associated with Scalar Factor 2, the variable group (4, 5, 6). Because it had its largest loading on Scalar Factor 1, Variable 10 would be included with the variable grouping (1, 2, 3, 10), thereby leaving Scalar Factor 3 defining three variables (7, 8, 9). Such an assumption would result in two of the three scalar factors defining perfect scales.

In Table 9 the variable reproducibilities are reported as they were determined by various variable groupings. The total scale reproducibility reported in Table 9 is just the average variable reproducibility.

Table 9  
 Summary of Reproducibilities of Various Variable Groupings for Tucker Data

Scale	Variable reproducibility										Scale reproducibility	
	1	2	3	4	5	6	7	8	9	10		
Original Tucker Data												
Scale 1	1.00	1.00	1.00	--*	--	--	--	--	--	--		1.00
Scale 2	--	--	--	1.00	1.00	1.00	--	--	--	--		1.00
Scale 3	--	--	--	--	--	--	1.00	1.00	1.00	--		1.00
Modified Tucker Data												
Scale 1	.80	.70	.97	--	--	--	.73	--	--	.70		.83
Scale 2	--	--	--	1.00	1.00	1.00	--	--	--	--		1.00
Scale 3	--	--	--	--	--	--	.93	.90	.93	.77		.83
Scale 1 (modified)	.93	.90	.93	--	--	--	--	--	--	.77		.88

\*Symbol (--) implies variable is not in the row scale.

bility of the variables defining the scale. Regardless of grouping, Variable 10 always had a low reproducibility. Furthermore, its inclusion with a variable grouping always reduced the reproducibility of each variable in the grouping. A variable with such low reproducibilities is frequently a factorially complex variable, and as such it should not be included with any group of variables assumed to be defining an independent scale.

### Conclusions

There are several conclusions that may be made directly from the illustrative data set and from data analyses not reported.

1. To the extent that there are perfect scales in the data, the MHA algorithm will identify them.
2. The MHA algorithm, when utilizing the orthotran solution, will identify scalar factors that are either orthogonal or oblique.
3. The MHA algorithm will identify pseudoscales in a data set.
4. The MHA algorithm will identify both independent and nonindependent scales.

The Guttman scales identified by the MHA algorithm suffer from the same problems that beset the traditional Guttman scales, e.g., extreme marginals causing chance scales and illogical variables leading to illogical scales. Implicit in the discussion herein is the fact that the same precautions and concerns associated with traditional Guttman scales should also be associated with the Guttman scales identified by the MHA algorithm. Multiple Hierarchical Analysis is only an algorithm for exploratory scalogram analysis.

### References

- Carroll, J. The nature of data or how to choose a correlation coefficient. *Psychometrika*, 1961, 26, 347-372.
- Cattell, R. B., & Jaspers, J. A general plasmode (No. 30-10-5-2) for factor analytic exercises and research. *Multivariate Behavioral Research Monographs*, 1967, 2, 3.
- Christofferson, A. Factor analysis of dichotomized variables. *Psychometrika*, 1975, 40, 5-32.
- Cureton, E. Note on  $\phi/\phi_{max}$ . *Psychometrika*, 1959, 24, 89-91.
- Goodenough, W. H. A technique for scale analysis. *Educational and Psychological Measurement*, 1944, 4, 179-190.
- Guilford, J. P. *Fundamental statistics in psychology and education*. New York: McGraw-Hill, 1965.
- Guttman, L. A basis for scaling qualitative data. *American Sociological Review*, 1944, 9, 139-150.
- Guttman, L. The Cornell technique for scale and intensity analysis. *Educational and Psychological Measurement*, 1947, 7, 247-279.
- Guttman, L. Relation of scalogram analysis to other techniques. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, & J. A. Clausen (Eds.), *Measurement and prediction*. Princeton, NJ: Princeton University Press, 1950.
- Harris, C. W., & Kaiser, H. Oblique factor analytic solutions by orthogonal transformations. *Psychometrika*, 1964, 29, 347-362.
- Hofmann, R. J. Indices descriptive of factor complexity. *Journal of General Psychology*, 1977, 96, 103-110.
- Hofmann, R. J. The orthotran solution. *Multivariate Behavioral Research*, 1978, 13, 99-124.
- Johnson, J. M. Maximal selectivity, correctivity and correlation obtainable in  $2 \times 2$  contingency tables. *American Journal of Psychology*, 1945, 58, 65-68.
- Lingoes, J. Multiple scalogram analysis: A set-theoretic model for analyzing dichotomous items. *Educational and Psychological Measurement*, 1963, 23, 501-524.
- Loevinger, J. The technique of homogeneous tests compared with some aspects of "scale analysis" and factor analysis. *Psychological Bulletin*, 1948, 45, 507-529.
- Purzek, R. Factor analysis. In J. Belzer et al. (Eds.), *Encyclopedia of computer science and technology* (Vol. 8). New York: Marcel-Deker Press, 1977.

Tucker, L. *Report of working group on factor analysis problems*. Paper presented at the joint meeting of the Psychometric and Psychonomic Society, Niagra Falls, 1964.

### **Acknowledgments**

*A majority of this article was written while the author was a Faculty Research Fellow; thus, the support necessary for both development and preparation of this article was provided by the Dean of the Graduate School and Research, Miami University, and the*

*Faculty Development Fund of the Miami University Alumni Association. I am indebted to Bill Gray, who provided interaction during the germination stages of this article, and to Bob Pruzek for providing much encouragement and interaction during its writing.*

### **Author's Address**

Send requests for reprints or further information to Richard J. Hofmann, Child Development Research Center, Department of Educational Psychology, 201 McGuffey Hall, Miami University, Oxford, OH 45056.