DOCUMENT RESUME

ED 051 269 TM 000 583

AUTHOR	Horst, Paul
TITLE	Factor Scores Independent of Item Traits.
INSTITUTION	Hawaii Univ., Honolulu. Education Research and Development Center.
SPONS AGENCY	Office of Economic Opportunity, Washington, D.C.
PUB DATE	Feb 71
NOTE	19p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Factor Analysis, *Motivation, *Personality Tests, Response Mode, *Response Style (Tests), Scores, Scoring Formulas, *Testing Problems
IDENTIFIERS	Gumpgookies

ABSTRACT

During early attempts to interpret factors represented in scores on the Gumpgookies test, an instrument designed to tap motivation to achieve in young children, the factors identified by ordinary factor-analytic techniques were found to be confounded by the subjects' response sets. This paper proposes a method for defining objectively irrelevant item response tendencies. A procedure for deriving independent factor scores from a set of items in which scores are also independent of irrelevant item trait responses is developed. The stages involved in the investigation are discussed and all steps of statistical computation are included. (Author/PR)

EDUCATION RESEARCH AND DEVELOPMENT CENTER

COLLEGE OF EDUCATION
UNIVERSITY OF HAWAII
HONOLULU, HAWAII

Center for Research in Early Childhood Education
Dorothy C. Adkins, Director

Factor Scores Independent of Item Traits
A Section of the Final Report of the University of Hawaii
to the Office of Economic Opportunity

Paul Horst, Consultant to the University of Hawaii
and Professor Emeritus of Psychology, University of Washington

The research reported herein was performed pursuant to a grant (Number CG9929) of the United States Office of Economic Opportunity to the University of Hawaii. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment on the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official position or policy of the Office of Economic Opportunity. The study was also supported in part by Public Health Research Grant 2 R01 MH00743-15 to the University of Washington, Seattle, Washington.

February 1971

FOREWORD

The report that follows was prepared by Dr. Paul Horst, Professor Emeritus of Psychology, University of Washington, as an outgrowth of problems encountered by the University of Hawaii Center for Research in Early Childhood Education in the course of early attempts to interpret factors represented in scores on the Gumpgookies test. This is an instrument, developed in part through support from the Office of Economic Opportunity, which is composed of dichotomous items designed to tap components of motivation to achieve in young children. It had been found that factors identified through ordinary factor-analytic techniques were confounded by response sets of the subjects, notably those associated with position of the keyed answer on the page or section of a page presented to the respondent and with whether the keyed answer was presented first or second. Hence assistance was solicited from Dr. Horst in devising a method of extracting factors with response set or "item characteristic" scores partialled out, so that the resulting factor scores would be independent of response sets. He has worked with the Center as Consultant on this problem, along with others. Some of the later aspects of his work on the general problem that had been posed have been supported by a grant to The University of Washington from the United States Public Health Service.

In addition to the authors of the test, Dr. Dorothy C. Adkins and Dr. Bonnie L. Ballif of Fordham University, who was also serving the Hawaii Center as Consultant, Mr. Renato Espinosa and Mr. Robert Bloedon contributed to the project, particularly in development of appropriate computer techniques.

The success of the approach described by Dr. Horst is attested to in another part of the report of the Hawaii Center, "A New Approach to Response Sets in Analysis of a Test of Motivation To Achieve," by Dorothy C. Adkins and Bonnie L. Ballif. In the study reported therein, it was empirically verified that the method does yield factor scores that are uncorrelated with response set scores, that the factors are characterized by KR-20 reliability coefficients greater than zero, and that the factors can be at least tentatively identified.

ABSTRACT

A method is proposed for defining objectively irrelevant item response tendencies. A procedure is developed for deriving independent factor scores from a set of items in which scores are also independent of irrelevant item trait responses.

FACTOR SCORES INDEPENDENT OF ITEM TRAITS

I. The Problem

Much attention has been given to the influence of those characteristics of stimulus elements (items, for example) on responses of individuals, aside from the actual purported item content. These characteristics may be social desirability, acquiescence, polarity of statement (do, do not), preference value, serial position, attributes of response categories (primacy, recency), etc. The literature in this area is too extensive to make even illustrative references feasible here. Perhaps now well structured procedures should be developed for the management of extraneous item characteristics in the evaluation of personality variables. Such an enterprise should probably proceed as follows:

1. Exhaustive identification of those stimulus element characteristics which are independent of item content and to which persons may respond differentially.

2. The definition of the elements of an item characteristic (IC) vector such that, when a person's response vector is multiplied by the IC vector, one may rationally regard such a product as the person's score for that item characteristic. For example, in the case of dichotomous items, an item preference vector would consist of the preference or "p" values of the items. The general approach might be as follows: A response matrix for a sample of entities is defined with rows as entities. An IC matrix conformable to the response matrix on the right has a column corresponding to each defined item characteristic. The IC matrix is defined so that the product of the response matrix by the IC matrix yields a matrix that may logically be regarded as a matrix of IC scores. Definition of the response

matrix will depend on the structure of response option patterns and on keying procedures.

In any case, it is probably most appropriate to investigate response styles or other extraneous item characteristics only after having operationally and computationally defined the corresponding item characteristic scores.

3. Having defined item characteristic scores, one may then develop models for utilizing these together with item scores in various ways. For example: One may partial out the IC variables from the item variables. One may include them with item or scale scores in factor analyses. One may utilize them with or without item or scale scores in estimating criterion measures. In general, one is free to utilize the IC variables in any way that other variables can be used in multivariate analysis models.

The crucial consideration is that an entity response vector and an item characteristic vector each be defined in such a way that their minor product constitutes an acceptable definition of an item characteristic score.

This general problem of achieving from dichotomous items factor scores that are independent of particular response sets or item characteristic scores was thrust upon the attention of the writer by Dorothy C. Adkins and Bonnie L. Ballif in connection with a project of the Center for Research in Early Childhood Education at the University of Hawaii. In their early factor analyses of items in the Gumpgookies, a test of motivation to achieve in school for young children, attempts to interpret factors were plagued by an apparent contamination of the factors by two types of response sets affecting the dichotomous items: position of the keyed answer to an item--right or left--and the order in which the keyed answer to an item was presented to the subject--first or second. The methods described herein are an outgrowth of their desire for factors based upon an item intercorrelation matrix with response set scores partialled out. They have been applying the method with every indication of success (Adkins and Ballif).

II. The Rationale

Let

n_e be the number of entities

n_a be the number of items

n_t be the number of item traits

n_f be the number of factors

X_{ea} be an $(n_e \times n_a)$ binary item score matrix

X_{at} be an $(n_a \times n_t)$ item trait matrix

M_a be an item mean vector

D_a be an item standard deviation matrix

x_{ea} be a deviation item score matrix

We define an $(n_e \times n_t)$ *trait score matrix* by

$$X_{et} = X_{ea} X_{at} . \tag{1}$$

From (1) we see that a trait score is defined as the scalar product of an entity's item score vector by an item trait vector.

We consider now the general problem of calculating item factor scores which are independent of item trait scores. To do this, we first consider the problem of calculating item scores which are independent of item trait scores. We let x_{et} be an $(n_e \times n_t)$ deviation matrix of item trait scores. It can readily be shown from the definitions and equation 1 that

$$x_{et} = x_{ea} X_{at} . \tag{2}$$

Let

C_{aa} be an n_a'th order matrix of covariances of item scores;

C_{at} be an $(n_a \times n_t)$ matrix of covariances of item scores with item trait scores;

C_{tt} be an $(n_t \times n_t)$ matrix of trait score covariances; and

u_{ea} be an $(n_e \times n_a)$ matrix of item scores independent of item trait scores.

Reversing subscripts to indicate transposition, we have from the definitions

$$C_{aa} = x_{ae} x_{ea} / n_e ; \tag{3}$$

$$C_{at} = x_{ae} x_{et} / n_e ; \tag{4}$$

$$C_{tt} = x_{te} x_{et} / n_e . \tag{5}$$

We now write

$$u_{ea} = x_{ea} - x_{et} B_{ta} , \tag{6}$$

where B_{ta} is a matrix determined so that u_{ea} is independent of the matrix of item trait scores x_{et} , or such that

$$x_{te} u_{ea} = 0 . \tag{7}$$

From (4) through (7)

$$0 = C_{ta} - C_{tt} B_{ta} . \quad (3)$$

And from (8)

$$B_{ta} = C_{tt}^{-} C_{ta} . \quad (9)$$

Equation (9) is the well known expression for the matrix of multiple regression constants calculated from deviation measures.

From (2) and (6),

$$u_{ea} = X_{ea} (I - X_{at} B_{ta}) . \quad (10)$$

Let

$$G_{aa} = u_{ae} u_{ea} / n_e , \quad (11)$$

where G_{aa} is now the covariance matrix of item scores independent of the item trait scores.

From (3), (4), (5), (6), (9), and (11) we have

$$G_{aa} = C_{aa} - C_{at} C_{tt}^{-1} C_{ta} . \quad (12)$$

Also from (2), (3), and (4),

$$C_{at} = C_{aa} X_{at} , \quad (13)$$

and from (9) and (12)

$$G_{aa} = C_{aa} - C_{at} B_{ta} . \quad (14)$$

The matrix G_{aa} in (14) is a covariance matrix. To get factor scores independent of item trait scores, we consider first the correlation matrix derived from G_{aa} . We let

$$D_c = \text{diag} (G_{aa}) . \quad (15)$$

Then the partial correlation matrix of factor scores with item scores partialed out is

$$R_{aa} = D_c^{-\frac{1}{2}} G_{aa} D_c^{-\frac{1}{2}} . \quad (16)$$

We next consider a factor analysis of R_{aa} . First we consider the basic structure or principal axes solution for n_f factors. We let A_f be the first n_f principal axes factors of R_{aa} . The method utilized will be the basic structure successive factor method (Horst, 1965, p. 160).

We let

$$A_f = Q_f \Delta_f , \quad (17)$$

where Δ_f^2 is a diagonal matrix of the n_f largest eigenvalues of R_{aa}, and Q_f is a matrix of the corresponding eigenvectors.

Next we solve for the varimax transformation of the A_f matrix by the simultaneous factor varimax solution (Horst, 1965, p. 423). This solution also incorporates the algorithms used in the principal axes solutions, since the method requires a series of basic structure type solutions. We indicate this solution by

$$B = A_f H , \quad (18)$$

where H is the square orthonormal transformation which satisfies the varimax criterion.

It should be noted that the varimax criterion is independent of the signs of the varimax factor loading matrix. For this reason it may often be necessary to determine both row and column sign changes for a varimax factor loading solution in order to maximize the number of large positive elements. For some factor analytic solutions of item variables, it may be that some items were not keyed to give the optimal positive manifold. Suppose then we let

$$b = i_L B i_R , \quad (19)$$

where i_L is a sign matrix operating on rows of B and i_R operates on columns. As a first approximation we give i_L the signs of the corresponding elements in $A_{.1}$, the first principal axes factor loading vector. This we may indicate by

$$_1i_L = \text{sign} (D_{A_{.1}}) . \quad (20)$$

We let

$$_1b = {_1i_L} B . \quad (21)$$

We then take as i_R the signs of the column sums of $_1b$ in (21) and write

$$_2b = {_1i_L} B i_R . \quad (22)$$

Next from ${}_2b$ we make up a sign matrix ${}_2i_L$ from the signs of the largest absolute values in the rows of ${}_2b$ of (22). We then let

$$i_L = {}_1i_L \, {}_2i_L \, . \tag{23}$$

Hence for the b matrix of (19) the largest element of each row will be positive. In general, it can also be expected that the largest element of each column will be positive.

It should be observed that the problem of determining the optimal sign-changing matrices i_L and i_R for varimax factor loading matrices has been largely ignored by most investigators, and that this problem is not restricted to the case where factor scores independent of item trait scores are sought.

To define the trait-free factor score matrix, we return to equation (10). Here u_{ea} is an item score matrix independent of the item trait scores, as indicated by equation (7). But the scores in this matrix are not standardized. Therefore we let

$$v_{ea} = u_{ea} \, D_c^{-\frac{1}{2}} \, . \tag{24}$$

From (11), (15), and (16), we see that

$$v'_{ea} \, v_{ea} \, / \, n = R_{aa} \, , \tag{25}$$

and therefore v_{ea} is a matrix of standard measures.

We now wish to find a matrix of trait-free factor scores y_{ef} which enables us best to approximate v_{ea} in the least squares sense. The appropriate model for this solution may be written

$$v_{ea} \, i_L - y_{ef} \, b' = \varepsilon \, , \tag{26}$$

where of course v_{ea} and b are known and the i_L appears on the left of (26) to conform to i_L in (19).

The least squares solution for y_{ef} in (26) is well known to be

$$y_{ef} = v_{ea}\, i_L\, b\, (b'\, b)^{-1} . \tag{27}$$

From (17), (18), (19), and (27), we get

$$y_{ef} = v_{ea}\, Q\Delta^{-1}\, H\, i_R . \tag{28}$$

But from (17) and (28) we may write

$$y_{ef} = v_{ea}\, A_{af}\, \Delta^{-2}\, H\, i_R . \tag{29}$$

If we let

$$C = \Delta^{-2}\, H\, i_R \tag{30}$$

and

$$\beta = AC , \tag{31}$$

from (29) and (31),

$$y_{ef} = v_{ea}\, \beta . \tag{32}$$

From (24) and (32),

$$y_{ef} = u_{ea}\, D_c^{-\frac{1}{2}}\, \beta . \tag{33}$$

Let

$$B_{af} = D_c^{-\frac{1}{2}}\, \beta . \tag{34}$$

From (33), (34), and (10),

$$y_{ef} = x_{ea}\, (I - X_{at}\, B_{ta})\, B_{af} . \tag{35}$$

Let

$$B_{tf} = B_{ta}\, B_{af} . \tag{36}$$

From (35) and (36),

$$y_{ef} = x_{ea}\, (B_{af} - X_{at}\, B_{tf}) . \tag{37}$$

Finally, let

$$\beta_{af} = B_{af} - X_{at}\, B_{tf} . \tag{38}$$

From (38)

$$y_{ef} = x_{ea}\, \beta_{af} . \tag{39}$$

Now by definition

$$x_{ea} = X_{ea} - 1\, M_a{}' \, . \tag{40}$$

From (39) and (40)

$$y_{ef} = X_{ea}\, \beta_{af} - 1\, M_a{}'\, \beta_{af} \, . \tag{41}$$

It can be proved now that

$$y_{fe}\, y_{ef} \,/\, n_e = I \tag{42}$$

and

$$y_{fe}\, x_{et} = 0 \tag{43}$$

or that the factor scores y are uncorrelated and of unit variances, and also uncorrelated with the item trait scores.

Suppose we wish to transform the factor scores so that they have means of a_c and standard deviations of s_c. We may write

$$Y_{ef} = y_{ef}\, s_c + 1\, 1'\, a_c \, . \tag{44}$$

Let

$$B_{af} \equiv B_{af}\, s_c \tag{45}$$

and

$$V_a{}' = 1'\, a_c - M_a{}'\, B_{af} \, . \tag{46}$$

From (43) through (46),

$$Y_{ef} = X_{ea}\, B_{af} + 1\, V_a{}' \, . \tag{47}$$

Suppose now we wish to construct an integer scoring matrix E_{af} from B_{af} such that the largest absolute value in each column of E_{af} is c_1. We let D be a diagonal matrix of the largest absolute values in the columns of B_{af},

calculate

$$B_{af} \equiv (B_{af}\, D^{-1})\, (c_1 + .999) \, , \tag{48}$$

and take the integer function of B_{af}; thus

$$E_{af} = \text{int}\,(B_{af}) \, . \tag{49}$$

The maximum absolute value in each column of E_{af} will now be c_1. To further simplify the scoring, we may for any integer c_1 greater than 1 let E_{ij} = sign E_{ij} for $E_{ij} \neq 0$. With this procedure, the number of items discarded because of all 0 elements in a row of E_{af} decreases as c_1 increases.

For the integer scoring method for dichotomous items, one may reverse the keying of all items with negative scoring weights so that all weights become positive.

The interpretation of the negative scoring weights is that the items with negative weights for a factor suppress unwanted variance of other factors which contaminate the items with positive weights for the factor.

III. Computational Procedure

The computational procedure for the foregoing rationale will now be described. The symbol Ξ is read "the expression on the left is replaced by the expression on the right." Unless otherwise indicated, subscripts refer to order of matrices and reversal of subscripts means transposition of the matrix. Exponents in parentheses mean elemental exponentiation.

Calculate as a minor product of type II vectors (Horst, 1963, p. 143)

$$C_{aa} = X_{ae} X_{ea} , \tag{50}$$

$$M_a = X_{ae} 1 , \tag{51}$$

$$S_D = X_{ae}^{(2)} 1 . \tag{52}$$

Calculate the item means (M_a) and item standard deviation vectors:

$$M_a \Xi M_a / n_e , \tag{53}$$

$$S_D \Xi S_D / n_e , \tag{54}$$

$$S_D \Xi (S_D - M_a^{(2)})^{(\frac{1}{2})} . \tag{55}$$

Calculate the item covariance matrix:

$$C_{aa} \Xi C_{aa} / n_e - M_a M_a ' . \tag{56}$$

Calculate the item-by-trait covariance matrix as the major product of type II vectors (Horst, 1963, p. 144):

$$C_{at} = C_{aa} X_{at} . \tag{57}$$

Calculate the trait-by-trait covariance matrix as a major product of type II vectors (Horst, 1963, p. 144):

$$C_{tt} = X_{ta} C_{at} . \tag{58}$$

Calculate the inverse of a symmetric matrix (Horst, 1963, p. 461):

$$C_{tt} \equiv C_{tt}^{-1} . \qquad (59)$$

Calculate the regression matrix for estimating the item score matrix from the item trait scores:

$$B_{at} = C_{at} C_{tt} . \qquad (60)$$

Calculate the covariance matrix of item scores with trait scores partialed out:

$$C_{aa} \equiv C_{aa} - C_{at} B_{ta} . \qquad (61)$$

Let

$$D_c = \text{diag} (C_{aa}) . \qquad (62)$$

Calculate

$$D_c \equiv D_c^{-\frac{1}{2}} . \qquad (63)$$

Calculate the item correlation matrix with trait scores partialed out:

$$C_{aa} \equiv D_c C_{aa} D_c . \qquad (64)$$

Calculate the first n_f principal axes factor loading vectors of C_{aa} (Horst, 1965, p. 160):

$$A = Q_f \Delta_f . \qquad (65)$$

Calculate the varimax transformation matrix H (Horst, 1965, p. 428) iteratively as follows:

$$B = A H , \qquad (66)$$

$$D = \text{diag} (B' B) , \qquad (67)$$

$$D \equiv D / n_a , \qquad (68)$$

$$B \equiv B^{(3)} - B D , \qquad (69)$$

$$C = A' B . \qquad (70)$$

The basic structure of C (Horst, 1965, p. 437) is

$$p d q' = C \qquad (71)$$

and

$$H = p q' . \qquad (72)$$

Continue (66) through (72) until H stabilizes, beginning with $H = I$.

Calculate sign matrices for S_a and S_f, as indicated in equations (19) through (23) ($S_a \equiv i_L$, $S_f \equiv i_R$).

Calculate the optimal sign varimax factor loading matrix

$$B \equiv S_a \, B \, S_f \; . \qquad (73)$$

Calculate the varimax regression matrix as follows:

$$H \equiv H \, S_f \; , \qquad (74)$$

$$H \equiv \Delta^{-2} \, H \; , \qquad (75)$$

$$B = A \, H \; , \qquad (76)$$

$$B \equiv D_c \, B \; , \qquad (77)$$

$$B_{tf} = B_{ta} \, B \; , \qquad (78)$$

$$B \equiv b - X_{at} \, B_{tf} \; , \qquad (79)$$

$$B \equiv B \, S_c \; , \qquad (80)$$

$$V_a' = 1' \, a_c - M_a' \, B \; . \qquad (81)$$

Calculate E as an integer matrix from B as indicated in equations (48) and (49).

Calculate as minor products of type II vectors (Horst, 1963, p. 143) the exact factor scores by

$$y_B = X_{ea} \, B - 1 \, V_a' \; , \qquad (82)$$

the integer weight factor scores by

$$y_E = X_{ae} \, E \; , \qquad (83)$$

and the item trait scores by

$$y_t = X_{ae} \, X_{at} \; . \qquad (84)$$

Letting

$$y = (y_B, \, y_E, \, y_t) \; , \qquad (85)$$

calculate the correlations among the three sets of y scores in (85) as follows:

Calculate as a minor product Type II vector:

$$C_{aa} = y' y , \tag{86}$$

$$M_a = y' 1 , \tag{87}$$

$$S_D = y^{(2)'} 1 . \tag{88}$$

Calculate the means and standard deviations for the y supermatrix by

$$M_a \equiv M_a / n_e , \tag{89}$$

$$S_D \equiv S_D / n_e , \tag{90}$$

$$S_D \equiv (S_D - M_a^{(2)})^{(\frac{1}{2})} . \tag{91}$$

Calculate the correlation supermatrix for the y scores by

$$C_{aa} \equiv (C_{aa} - M_a M_a') / (S_D S_D') , \tag{92}$$

where $/$ means elemental division.

We may now define

$$C_{aa} = \begin{bmatrix} R_{BB} & R_{BE} & R_{Bt} \\ R_{EB} & R_{EE} & R_{Et} \\ R_{tB} & R_{tE} & R_{tt} \end{bmatrix} \tag{93}$$

Then we should have

$$R_{BB} = I , \tag{94}$$

$$R_{Bt} = 0 , \tag{95}$$

and R_{EE} and R_{BE} should approximate identity matrices while R_{Et} should approximate a null matrix.

IV. Kuder-Richardson Factor Score Reliability

We may estimate the Kuder-Richardson reliability for the y_B scores for the j'th factor by

$$r_j = \frac{n_a}{(n_a - 1)\, S_c^2} (S_c^2 - B_j' D_s^2 B_j) \, , \qquad (96)$$

where B_j is the j'th column vector of the exact varimax weighting matrix which yields factor scores with standard deviations of S_c, and D_s is a diagonal matrix of item standard deviations. This formula is based on the assumption that the average item retest covariance is equal to the average inter-item covariance. Presumably, therefore, it may markedly underestimate the retest reliability of a factor score.

REFERENCES

Horst, Paul. *Matrix algebra for social scientists*. New York: Holt, Rinehart and Winston, 1963.

Horst, Paul. *Factor analysis of data matrices*. New York: Holt, Rinehart and Winston, 1965.

Adkins, Dorothy C., and Ballif, Bonnie L. "A new approach to response sets in analysis of a test of motivation to achieve," 1970 (Submitted for publication to *Educational and Psychological Measurement*).