PSYCHOMETRIKA—VOL. 34, NO. 4
DECEMBER, 1969

EVALUATION OF FACTOR ANALYTIC RESEARCH PROCEDURES BY MEANS OF SIMULATED CORRELATION MATRICES*

LEDYARD R TUCKER

UNIVERSITY OF ILLINOIS

RAYMOND F. KOOPMAN

SIMON FRASER UNIVERSITY

AND

ROBERT L. LINN

EDUCATIONAL TESTING SERVICE

In order to study the effectiveness of factor analytic methods, a procedure was developed for computing simulated correlation matrices which are more similar to real data correlation matrices than are those matrices computed from the factor analysis structural model. In the present investigation, three methods of factor extraction were studied as applied to 54 simulated correlation matrices which varied in proportion of variance derived from a major factor domain, number of factors in the major domain, and closeness of the simulation procedure to the factor analysis structural model. While the factor extraction methods differed little from one another in quality of results for matrices more dissimilar to the factor analytic model, major differences in quality of results were associated with fewer factors in the major domain, higher proportion of variance from the major domain, and closeness of the simulation procedure to the factor analysis structural model.

Questions as to the power of factor analytic research studies to reflect relations within a domain of psychological phenomena are of considerable scientific importance due to the wide use made of factor analysis in psychology. The basic formulation of factor analysis involves a particular mathematical system which may parallel to a better or poorer degree the relations existing in various domains of psychological phenomena. Further, different experiments within any one domain of phenomena may differ with respect to important characteristics which affect the parallelism between the factor analytic model and the observed data. Undoubtedly, such lack of parallelism has an effect on the capability of a factor analytic research study to obtain results relevant to the structure of relations in the domain of phenomena under study.

* The research was jointly supported by the University of Illinois and the office of Naval Research under contracts Nonr 1834(39) and N00014-67-A-0305-0003.

More precise knowledge is needed as to the nature and extent of such effects as related to the nature and extent of lack of parallelism between the formal factor analytic model and the relations in the domain of phenomena.

The mathematical model basic to factor analysis has been presumed in many theoretical and methodological developments, and various theorems have been developed which may have relevance to the analysis of observed data. Strictly, however, the truth of such theorems has been proven only within the constraints of the mathematical model, and their relevance in the analysis of data may be a function of the degree of parallelism between the mathematical model and the structure of relations in the data. For example, Guttman [1954], Harris [1962], and Kaiser [1961] have developed and advocated use of several measures of the lower bound to the number of common factors involved in a correlation matrix. These theorems are strictly true for correlation matrices constructed from the mathematical factor analytic model. However, questions exist as to their relevance to the analysis of a correlation matrix based upon observed data.

Mathematical statistical developments such as Lawley's [1940] method of maximum likelihood factor analysis and Rao's [1955] canonical factor analysis are attempts to solve some of the problems generated from the analysis of data from a sample of individuals. These developments, however, presume the mathematical model for factor analysis so that the relevance of these methods in the analysis of observed data may be a function of the degree of parallelism between the mathematical model and the structure of the phenomena and properties of the experiment giving rise to the data.

In contrast to the theoretical developments and methodological studies based on the mathematical model as discussed in the preceding paragraphs, other studies may attempt to demonstrate the usefulness of factor analytic research in its application to real data. Thurstone [1938], for example, considered many of his studies as such examples in addition to the contributions to knowledge about the structure of abilities. Others, for example Spearman [1939], Holzinger and Harman [1938], Kaiser [1960], and Zimmerman [1953], have compared results obtained by two or more factorial techniques applied to the same data. This type of study, however, is beset by the problem of validity for interpretation of the results and is forced to subjective judgment as to the nicety of results. No true or target solution is known with which to compare the results obtained.

Three studies, Thurstone's [1940] box problem, Cattell and Dickman's [1962] ball problem, and Cattell and Sullivan's [1962] cups of coffee problem have possessed a solution to this validity problem. In all of these cases, the domain of phenomena was chosen in the physical world and such that previous knowledge indicated the structure of relations existent. Each of these studies involves a single demonstration of the power of the factor analytic approach to represent relations in a domain of observations.

In order to evaluate the effectiveness of proposed methods and the relevance of derived theorems from the factor analytic mathematical model, there is a need for a broad program of methodological studies involving data that possess the following attributes: 1) the structural relations underlying the data are known so that the validity of the factor analytic results may be checked; 2) the degree to which the relations underlying the data are paralleled by the factor analytic mathematical model may be varied so as to facilitate the study of dependence of effectiveness of method to degree of parallelism; and 3) experimental planning and execution may be varied both as to general nature and to quality (from poor to good). The present program of studies attempts to meet these three criteria by the use of a computer procedure to produce correlation matrices which might be obtained from real data. This program does not solve the ultimate validity problem which still exists in the relation between the computed correlation matrices and those that might be obtained from real data. This procedure, however, should yield information on the power of the factor analytic methods when applied to data for which the factorial model is not precisely appropriate.

A point of considerable importance is that the present group of studies envisages data on a population of individuals. Problems arising from analyses based on samples of individuals are not being considered at this time. In case some of the problems associated with sampling of individuals are to be studied in some project, the simulated correlation matrices developed by the present procedure could be considered as the population correlation matrices. Sample correlation matrices could be developed by any of several techniques (see, for example, Odell and Feiveson, [1966]). Linn [1968] has followed this approach. However, the present studies attack the separable problem of relevance of various theorems and procedures developed from the mathematical model for factor analysis for population data uncomplicated by the sampling of individuals problems.

As discussed by Cliff and Hamburger [1967], the present simulation model represents one type of psychometric conceptualization of factor analytic theory. One type of psychometric conceptualization of factor analytic theory advocated and utilized by Guttman [1953, 1954, 1955, 1956], Tryon [1957(a), 1957(b), 1958], and Kaiser and Caffrey [1965] involves the postulations of a population of measures combined with the sampling of measures to form the batteries analysed. Loevinger [1965] has criticized this type of psychometric conceptualization when applied to mental test theory as being unrealistic in its postulation of a population of measures or items. In contrast to the sampling of measures concept the simulation model conceives of each measure involving a sampling of influences on the behavior of the individual. The mixture of this sampling is taken to be partially under the control of the designer of the measuring instruments and of the experimenter during the collection of observations on individuals.

The research reported here concentrates on a procedure to simulate correlation matrices, a partial attack on two data analysis problems, and methodological problems in comparing output factors with the input factors. The two data analysis problems are: 1) the communality problem, and 2) the number of factors problem.

1. *Procedure for Simulation of Correlation Matrices*

The procedure developed for simulation of correlation matrices involves a mathematical, probabilistic model (hereafter called the simulation model) possessing many features similar to the mathematical model for factor analysis (hereafter called the formal model). In fact, the simulation model contains the formal model as a special case. Both models presume the existence of a major domain consisting of those influences on observed scores of individuals for the phenomena which the experimenter wishes to study. Both models presume other factors or influences outside the major factor domain. The two models differ in the conception of these other influences. An important similarity between the models is that both are linear in that they involve addition of contributions to correlations from the various presumed influences on the scores of individuals. A curvilinear model for simulation of correlation matrices could be constructed which would include the formal model and the present simulation model as special cases and which would be more general. However, in our opinion, the procedure described here represents a valuable first step and will provide valuable information concerning methodological problems in analysis of data.

The major departure of the simulation model from the formal model is in the conception of many minor factors in the simulation model in addition to the postulation of unique factors composed of specific influences and errors of measurement for each variable. Subjectively, the postulation of many minor influences, or factors, which affect the values of the observed scores seems to make considerable sense as a description of the real world. Loadings of variables on these minor factors are taken as random variables somewhat out of the control of the designer of the variable and the experimenter collecting the data. Parameters are included in the simulation model to regulate the general nature of these random values.

The three types of factors considered are designated by the subscript s with values of 1 for major factors, 2 for minor factors, and 3 for unique factors. The number of factors of each type is designated by M_s and the factors of each type are designated by the subscript m_s such that:

$$m_s = 1, 2, 3, \cdots, M_s .$$

Variables are designated by the subscript j or j' with the number of variables being J, thus:

$$j \quad \text{or} \quad j' = 1, 2, 3, \cdots, J.$$

https://doi.org/10.1007/BF02290601 Published online by Cambridge University Press

For each type of factor, there is a matrix A_s with entries of "actual input factor loadings." There is a row in A_s for each variable and a column for each factor in section s of factors; thus, A_s is a matrix of order $J \times M_s$. Two sizes of major factor domains were used in the present studies: three factor major domains and seven factor major domains (*i.e.*, $M_1 = 3$ and $M_1 = 7$). All batteries contained twenty variables.

A matrix A_s^* is defined for each matrix A_s by adjusting the rows of A_s to unit length vectors. A matrix P_s is defined from each A_s^* by:

(1) $$P_s = A_s^* A_s^{*\prime}.$$

Matrix P_s is square of order J and is symmetric, positive, semidefinite. Since each row of A_s^* was defined as being of unit length,

(2) $$\text{Diag}\,(P_s) = I,$$

where Diag (P_s) is to be read as the diagonal matrix formed from the diagonal entries in matrix P_s. The simulated correlation matrix R is defined by:

(3) $$R = B_1 P_1 B_1 + B_2 P_2 B_2 + B_3 P_3 B_3$$

where B_1, B_2, and B_3 (in general B_s) are diagonal matrices with entries b_{1j}, b_{2j}, and b_{3j} (in general b_{sj}). These entries b_{sj} are restricted to being real, positive numbers such that:

(4) $$b_{1j}^2 + b_{2j}^2 + b_{3j}^2 = 1.$$

A consequence of the restriction on rows of the matrices A_s^* to being unit vectors and, thus, of the diagonal entries in matrices P_s being unities as stated in (2) is that

(5) $$r_{jj} = 1;$$
$$\text{Diag}\,(R) = I.$$

The matrix A_s of actual input factor loadings may be defined in terms of the matrices B_s and A_s^* by:

(6) $$A_s = B_s A_s^*.$$

From (1), (3), and (6):

(7) $$R = A_1 A_1' + A_2 A_2' + A_3 A_3'$$
$$= (A_1, A_2, A_3)(A_1, A_2, A_3)'$$

where (A_1, A_2, A_3) is a supermatrix containing the matrices A_1, A_2, and A_3 as horizontal sections. Note that the constructed matrix R has the formal properties, required for a correlation matrix, of having unit diagonal entries and being symmetric, positive, and semi-definite.

Coefficients in the B_s matrices are important parameters of the simulation model, functioning to regulate the proportions of the variances of the variables derived from the three types of factors. When B_2 is zero, the simulation model is identical with the formal model and B_1^2 contains the comunalities while B_3^2 contains the uniquenesses of the variables. These coefficients in the B_s matrices may be interpreted as reflecting the design of the measures, both as to individual variables and as to the battery of variables, utilized in a factor analytic study as well as reflecting the care exercised in collection of the observations. Values of the coefficients B_{1i}^2 used are given in Table 1.

The central feature of the simulation model is the development of the matrices A_s of "actual input factor loadings." Corresponding to each matrix A_s is a matrix $\tilde{A}_s$ of "conceptual input factor loadings" which represents the ideas of the designer of the variables. Comments about the matrix $\tilde{A}_1$ for factors in the major factor domain and a procedure used in the present studies are given in following paragraphs. The matrices $\tilde{A}_2$ and $\tilde{A}_3$ for minor factors and unique factors were set to zero, which represents the only sensible idea that the designer of the measures could have for these factors. The actual input factor loadings differ from the conceptual input factor loadings due to imperfections in the design of the variables and in the plan and conduct of the experiment. These imperfections may be conceptualized for the major and minor factor domains as involving some random processes. Possible forms of these processes, one form for each of the two types of factors, are discussed in following paragraphs. The effect of the imperfections on the unique factors is taken to make the ma rix A_3^* an identity matrix.

Conceptual input factor loadings for factors in the major factor domain represent the experimenter's ideas as to what he hopes is the factorial composition of the variables. His ideas may be more or less precise depending upon the extent and precision of his knowledge of the phenomena constituting the major domain. At the low end of a scale of extent of knowledge about the phenomena, the experimenter may have only a vague idea as to the nature of the phenomena and the influences constituting the major factors. From this vague idea he may be able to construct a number of measures, or variables, that seem to depend upon whatever influences may exist. Further, he may be able to judge that the variables appear to be somewhat related to or different from each other. He may be able to judge that the battery of variables represents the various aspects of the domain. An important aspect of judgment is whether the variables are simple or complex in the sense of depending on few or many influences so that the experimenter may be able to design and select variables judged to be simple. The existence of simple structures for the battery of variables depends on this judgment. As knowledge about the phenomena increases, the precision of design of measures should increase so as to enable experimenters to have more precise ideas about the conceptual input factor loadings of the variables.

The procedure adopted for development of the matrices $\tilde{A}_1$ in the present

Table 1

Values of the Coefficients b_{1j}^2 Used in the Studies

	Three Factor Major Domain				Seven Factor Major Domain		
	Ranges of b_{1j}^2				Ranges of b_{1j}^2		
Variable	High	Wide	Low	Variable	High	Wide	Low
1	.6	.6	.4	1	.8	.6	.4
2	.8	.3	.2	2	.7	.3	.2
3	.6	.2	.3	3	.7	.7	.3
4	.6	.7	.4	4	.6	.7	.4
5	.6	.7	.3	5	.7	.6	.2
6	.7	.4	.3	6	.8	.6	.4
7	.7	.8	.3	7	.6	.8	.3
8	.7	.2	.4	8	.6	.8	.3
9	.6	.7	.3	9	.8	.8	.4
10	.6	.8	.4	10	.8	.7	.2
11	.8	.5	.4	11	.8	.8	.2
12	.8	.8	.4	12	.8	.6	.3
13	.7	.2	.3	13	.6	.4	.2
14	.7	.3	.4	14	.6	.7	.2
15	.7	.8	.4	15	.6	.4	.3
16	.7	.7	.2	16	.7	.6	.2
17	.8	.2	.3	17	.6	.5	.3
18	.6	.5	.2	18	.7	.7	.2
19	.8	.5	.2	19	.8	.5	.3
20	.6	.2	.3	20	.7	.2	.4

studies was considered as representing the case when only vague ideas exist about the major factor domain. Each variable in each battery was developed independently from every other variable. First, relative conceptual input loadings were developed for the variable which constituted a row vector; then, the vector was adjusted to unit length by a multiplying factor. The relative conceptual loadings were integers developed by the following procedure. For a three-factor major domain, the loading on a randomly selected one of these factors was a 0, 1, or 2 chosen at random with equal probability. A limit of 2 was placed on the sum of loadings for the variable, so that if the first loading obtained was 2, the other two loadings were recorded as 0. If the first loading obtained was 1, a loading of 0 or 1 was chosen at random with equal probability for one or the other of the remaining factors, also chosen at random. In case the first loading was 0, the second loading obtained was 0, 1, or 2 chosen at random with equal probability for one or the other of the remaining factors. The loading of the variable on the third factor was chosen so that the sum of loadings was 2. A similar procedure was used for obtaining relative loadings for each variable in a seven factor domain. The sum of loadings for each variable was controlled at 6. The first loading was an integer in the range 0 through 6 (with equal probability) on a randomly chosen factor. The second loading was in the range from 0 through a value of $6 - a$, where a is the value of the first loading, on one of the remaining factors, etc. It is to be noted that this procedure tended to produce a fairly strong simple structure for the conceptual input factors.

An interesting result of the foregoing procedure was that some of the factors are much more represented by loadings of variables in a battery than are other factors. Table 2 presents the sums of squares of conceptual input factor loadings on each of the factors for each of the batteries developed. This result appears to be similar on a subjective impression level to the relative importance of factors in many studies of actual data early in the investigations of domains of phenomena. This trend in representation of factors in the batteries has a strong effect in the results of the analyses to be reported.

A three step procedure was utilized to develop the matrix A_1 of actual input factor loadings for the major factor domain from the matrix $\tilde{A}_1$ of conceptual input factor loadings. In the first of these steps, the conceptual input factor loadings are combined with random normal deviates to represent discrepancies that might occur in the actual construction of measuring instruments. Let $(\tilde{a}_1)_{jm_1}$ be the entry in row j and column m_1 of matrix $\tilde{A}_1$ and x_{jm_1} be a random normal deviate ($\mu = 0$, $\sigma = 1$) drawn independently corresponding to each $(\tilde{a}_1)_{jm_1}$. Let $(y_1)_{jm_1}$ be the output from the first step and defined by:

$$(y_1)_{jm_1} = (\tilde{a}_1)_{jm_1}c_{m_1} + d_{1j}x_{jm_1}(1 - c_{m_1}^2)^{1/2} \tag{8}$$

https://doi.org/10.1007/BF02290601 Published online by Cambridge University Press

Table 2

Sums of Squares of Conceptual Input Factor Loadings
Major Domain, on Factors in the Batteries Studied

Three Factor Major Domain

Factor	Batteries 1	2	3
1	7.03	6.02	10.05
2	7.54	7.03	6.04
3	5.51	6.02	4.02

Seven Factor Major Domain

Factor	Batteries 1	2	3
1	3.48	2.12	3.14
2	5.83	4.49	4.39
3	2.33	3.43	4.44
4	4.22	3.43	5.00
5	4.51	5.72	2.78
6	2.42	7.52	2.44
7	2.93	1.16	5.77

where c_{m_1} is a constant for each factor m_1 and d_{1j} is a constant for each variable j. The constant d_{1j} normalizes each row of x_{jm_1} to a unit length vector and is defined by:

$$d_{1j} = \left(\sum_{m_1} x_{jm_1}^2\right)^{-1/2}. \tag{9}$$

The constants c_{m_1} are conceptualized as representing the general control an experimenter has on the loading of actual variables on the factors. The c_{m_1} are limited to values in the range of 0 to 1, inclusive. Values of c_{m_1} used in the present studies were .7, .8, or .9, chosen at random with equal probability for each factor. These values are given in Table 3.

Table 3

Values of c_{m_1} Used in the Batteries Studied

Three Factor Major Domain

Factor	Batteries 1	2	3
1	.8	.7	.7
2	.8	.8	.9
3	.9	.8	.8

Seven Factor Major Domain

Factor	Batteries 1	2	3
1	.7	.8	.9
2	.8	.8	.7
3	.7	.9	.8
4	.8	.9	.9
5	.9	.9	.8
6	.8	.7	.7
7	.7	.8	.8

https://doi.org/10.1007/BF02290601 Published online by Cambridge University Press

Table 4

Sums of Squares of Actual Input Factor Loadings, Major Domain

Three Factor Major Domain				Seven Factor Major Domain			
Factors	Batteries			Factors	Batteries		
	1	2	3		1	2	3
High Range of b_{1j}^2							
1	5.17	4.72	3.82	1	2.21	1.69	1.64
2	4.16	4.80	5.25	2	3.22	1.83	1.53
3	4.38	4.18	4.63	3	.71	1.71	1.87
				4	2.56	1.80	3.12
				5	2.52	2.48	1.83
				6	1.47	3.45	1.34
				7	1.32	1.04	2.67
Wide Range of b_{1j}^2							
1	2.37	3.10	3.06	1	1.61	1.50	1.61
2	3.57	4.05	3.95	2	2.52	1.67	1.61
3	4.17	2.94	3.09	3	.66	1.49	.95
				4	1.67	1.14	2.88
				5	2.26	3.26	.97
				6	1.46	2.48	1.31
				7	1.82	.47	2.66
Low Range of b_{1j}^2							
1	2.18	2.20	2.93	1	.90	.49	.63
2	2.25	2.05	1.64	2	.93	.98	.94
3	1.97	2.15	1.82	3	.74	.83	.66
				4	1.02	.66	.91
				5	.78	.96	.65
				6	.55	1.60	.83
				7	.77	.16	1.08

The second step in developing the actual input factor loadings involves use of a "skewing function" which was introduced to reduce and limit the negativity of the factor loadings $(\tilde{a}_1)_{im_1}$. This function produces coefficients $(z_1)_{im_1}$ as follows:

$$(z_1)_{im_1} = \frac{(1+k)}{(2+k)} \frac{(y_1)_{im_1}[(y_1)_{im_1} + |(y_1)_{im_1}| + k]}{[|(y_1)_{im_1}| + k]} \tag{10}$$

https://doi.org/10.1007/BF02290601 Published online by Cambridge University Press

where k is a parameter with a value to be chosen within the range of 0 to infinity, inclusive. A value of $k = .2$ was used in the present studies. This function was introduced to limit the actual input factor loadings to approximating a positive manifold. This step was necessitated by the possibility that the operation in (8) might produce fairly negative values of $(y_1)_{jm_1}$. Each vector of $(z_1)_{jm_1}$ was adjusted to a unit vector by:

$$(a_1^*)_{jm_1} = g_{1j}(z_1)_{jm_1} \tag{11}$$

where

$$g_{1j} = [\sum_{m_1} (z_1)_{jm_1}^2]^{-1/2}. \tag{12}$$

The third, and final, step in developing the matrices A_1 of actual input factor loadings for the major domains was to premultiply the matrices A_1^* by the matrices B_1 described previously. This step corresponds to (6).

Table 4 gives the sums of squares of the entries in columns of the matrices A_1. These sums of squares may be interpreted as measures of the strengths of representation of the factors in the several batteries. Similarly, as for the conceptual input factor loadings, there is a wide range of sums of squares of actual input factor loadings for the factors in each of the batteries. To obtain further information on the differential strengths of dimensions of the input factor matrices, each of the matrices was rotated to principal axes factor matrices. Table 5 gives the sums of squares of the loadings on the principal axes. For the seven factor major domain and low range of b_{1j}^2 the sums of squares of loadings on the seventh principal axis are quite small. An effect of this wide range which may be anticipated is that the weaker input factors may not be represented in the output from the data analysis. This point will be considered further in subsequent discussion.

Development of loadings of the variables on factors in the minor factor domain involved random normal deviates ($\mu = 0, \sigma = 1$) drawn independently for each cell in the matrix A_2. These normal deviates are designated by x_{jm_2}. A first operation on these normal deviates was to multiply those for each factor by a constant which formed a decreasing geometric series in the progression of minor factors. That is, $x_{jm_2}^*$ is defined by:

$$x_{jm_2}^* = x_{jm_2}(1 - \epsilon)^{(m_2-1)} \tag{13}$$

where ϵ is a parameter of the simulation model and is restricted to the range of values 0 through 1, inclusive. Consider the possibility of an infinite series of minor factors. Consider, also, the sum of squares of $x_{jm_2}^*$ for variable j over the infinity of minor factors. It is interesting to note that the expected value of this sum of squares is not infinite when ϵ is greater than zero. This expectation turns out to be:

$$E\left(\sum_{m_2=1}^{\infty} x_{jm_2}^*\right) = \frac{1}{1 - (1 - \epsilon)^2}. \tag{14}$$

Table 5

Sums of Squares of Loadings on Principal Axes of

Major Domain Actual Input Factor Matrices

Three Factor Major Domain				Seven Factor Major Domain			
Factors	Batteries			Factors	Batteries		
	1	2	3		1	2	3
High Range of b^2_{1j}							
1	7.55	7.68	8.83	1	4.95	5.58	4.12
2	3.46	3.82	2.67	2	2.95	2.74	2.86
3	2.69	2.20	2.21	3	1.93	1.73	2.14
				4	1.71	1.31	1.84
				5	1.06	1.10	1.54
				6	.93	.96	1.03
				7	.47	.60	.48
Wide Range of b^2_{1j}							
1	5.10	4.95	4.99	1	4.71	5.19	4.23
2	3.41	3.47	3.15	2	2.24	2.16	2.53
3	1.59	1.68	1.96	3	1.98	1.75	1.98
				4	1.21	1.37	1.36
				5	.81	.73	.88
				6	.68	.58	.59
				7	.37	.24	.43
Low Range of b^2_{1j}							
1	4.15	3.14	3.97	1	2.02	2.69	2.13
2	1.27	1.95	1.48	2	1.00	.92	1.05
3	.99	1.30	.95	3	.79	.88	.73
				4	.71	.53	.57
				5	.56	.33	.52
				6	.41	.27	.40
				7	.20	.08	.30

Thus, if the $x^*_{jm_2}$ were considered as relative loadings on an infinite series of minor factors, the expected squared length of a vector of these loadings would not be infinite. Such a vector, even in an infinite dimensional space, may be shortened to a unit length with coordinates $(y_2)_{jm_2}$ where:

$$(15) \qquad (y_2)_{jm_2} = d_{2j} x^*_{jm_2}$$

and

$$(16) \qquad d_{2j} = (\sum_{m_2} x^{*2}_{jm_2})^{-1/2}.$$

Let the matrix Y_2 contain the $(y_2)_{jm_2}$ and consider the matrix product

$$(17) \qquad Q = Y_2 Y_2'$$

with entries

$$(18) \qquad q_{jj'} = \sum_{m_2} (y_2)_{jm_2}(y_2)_{j'm_2} .$$

While the expected value of Q over many samplings of random normal deviates x_{jm_2} is an identity matrix, each Q matrix is not restricted to being an identity matrix. Consider the expected value of $q^2_{jj'}$ when $j \neq j'$. An estimate of this expectation may be obtained from the expected value of $(\sum_{m_2} x^*_{jm_2} x^*_{j'm_2})^2$. This expectation turns out to be:

$$(19) \qquad E[(\sum_{m_2} x^*_{jm_2} x^*_{j'm_2})^2] = \frac{1}{1 - (1 - \epsilon)^4}.$$

From (15), (16), and (18),

$$(20) \qquad q^2_{jj'} = \frac{(\sum_{m_2} x^*_{jm_2} x^*_{j'm_2})^2}{(\sum_{m_2} x^{*2}_{jm_2})(\sum_{m_2} x^{*2}_{j'm_2})}.$$

Taking as an estimate of the expectation of $q^2_{jj'}$ the ratio of the expectations of the numerator and denominator of (20) yields:

$$(21) \qquad \text{Estimated} \quad Eq^2_{jj'} = \frac{1 - (1 - \epsilon)^2}{1 + (1 - \epsilon)^2}.$$

It is clear that as ϵ becomes very small, approaching zero, this estimate approaches zero, also. Thus, as ϵ approaches zero, each Q matrix should approach an identity matrix; but, for a finite ϵ, each Q matrix is not expected to be an identity matrix. The consequence of this observation in the present context is that for any particular developed battery, the minor factors may be expected to affect the off-diagonal correlations when ϵ is chosen as a finite value. A value of .02 was used for the present studies.

A second step was taken in development of the matrices A_2^* using the

https://doi.org/10.1007/BF02290601 Published online by Cambridge University Press

same "skewing function" on the $(y_2)_{im_2}$ as was employed for the major factor domain and is given in (8). The resulting $(z_2)_{im_2}$ were reduced to unit length vectors for each variable by steps parallel to (11) and (12).

In the computations, loadings on minor factors were computed on 180 factors, stopping quite short of the infinite number considered in the preceding discussion, since the loadings became of very trivial size beyond this point. Also, the expected contributions to the correlations became very small.

The matrices P_3 were taken to be identity matrices, a result that could be obtained by either of two definitions of A_3^*. In the first such definition, the matrix A_3^* could be defined as an identity matrix. This corresponds to the common conception of one unique factor for each variable in a battery. A second definition could be similar to the procedure followed for loadings of the variables in the minor factor domain but using an infinitesimal value of ϵ. This might be considered as more realistic for errors of measurement in that it implies a very large number of random influences on the measures. There is no argument to be made here for one or the other conception since they both result in P_3 being an identity matrix.

2. *Study Plans*

The study plans may be conceptualized in terms of three factors in the sense of the use of this word "factor" in experimental design. These factors, along with levels employed, are listed below.

1) Range of values of coefficients b_{1j}^2 :
   high range: $b_{1j}^2 = .6, .7,$ or $.8$;
   wide range: $b_{1j}^2 = .2, .3, .4, .5, .6, .7,$ or $.8$;
   low range: $b_{1j}^2 = .2, .3,$ or $.4$.
2) Size of major domain:
   three factors in major domain;
   seven factors in major domain.
3) Relation between coefficients b_{2j}^2 and b_{3j}^2 :
   formal model: $b_{2j}^2 = 0, b_{3j}^2 = (1 - b_{1j}^2)$;
   middle model: $b_{2j}^2 = b_{3j}^2 = (1 - b_{1j}^2)/2$;
   simulation model: $b_{2j}^2 = (1 - b_{1j}^2), b_{3j}^2 = 0$.

For each range of b_{1j}^2 a matrix B_1 was drawn for each size of major domain as given in Table 1. For each size of major domain, three matrices $\tilde{A}_1$ were developed. For each $\tilde{A}_1$, three matrices A_1 were computed, one for each range of b_{1j}^2, thus yielding a total of 18 matrices A_1. For each of these matrices, a matrix A_2^* was developed. The matrices A_3^* were taken as identity matrices.

The factor of relation between coefficients b_{2j}^2 and b_{3j}^2 is represented by three levels of ratio. The first level of ratio may be termed the formal model since, as discussed previously, correlation matrices constructed with b_{2j}^2 set at zero satisfy, exactly, the formal model for factor analysis. The third level of ratio may be termed the simulation model, while the second level of ratio may be termed the middle model. For each of the eighteen matrices A_1 and associated matrices A_2^* and A_3^*, three correlation matrices were computed, one for each class of relation between coefficients b_{2i}^2 and b_{3i}^2. †

While there are three correlation matrices for each cell of this factorial design, these replications are not independent from cell to cell. The correlation matrices for all cells for a given size of major domain were developed from the same three matrices $\tilde{A}_1$. The same matrix B_1 was used for all correlation matrices in each cell for the three classes of relation between coefficients b_{2i}^2 and b_{3i}^2 and a given size of major domain. This design was chosen to facilitate major comparisons between cells by reduction of between cell variation due to random effects. A more completely balanced design would have resulted in many more correlation matrices and would have increased the computational work inordinately.

Analytic methodological problems considered in the present group of studies include: 1) the communality problem, and 2) the number of factors to be extracted problem. Principal axes factoring was used in all analyses performed. For the communality area of problems, each simulated correlation matrix was factored three times: 1) with unity in every diagonal cell (designated as RwU), 2) with the matrix of correlations rescaled to a variance-covariance matrix so that the variance of errors of predicting each variable from all other variables is unity (designated as C), and 3) with the squared multiple correlation of each variable with all other variables in the diagonal cell for each variable (designated as RwSMC).

A number of workers, starting with Hotelling [1933] and Kelley [1935], have advocated the principal components factoring represented by the first procedure. The second procedure was proposed by Harris [1962] and is related both to Guttman's lower bound for the communality and to the maximum likelihood factor method of Lawley [1940] and to Rao's [1935] canonical factoring procedure. The matrix C used in this procedure is defined by (22):

$$C = S^{-1}RS^{-1} \tag{22}$$

where S is a diagonal matrix containing the standard errors of estimating each variable from the remaining variables. The output factor matrix, F_r, for r factors is obtained by (23):

$$F_r = SV_r(\beta_r - I)^{1/2} \tag{23}$$

where β_r is a diagonal matrix containing the r largest characteristic roots of C and V_r contains, as column vectors, the corresponding characteristic vectors scaled to unit length. The third procedure is the result of Guttman's theorem that the squared multiple correlation of each variable with all remaining

† The 18 matrices A_1 of actual output factor loadings and the 54 correlation matrices are on file at the American Documentation Institute as document number 10060.

https://doi.org/10.1007/BF02290601 Published online by Cambridge University Press

variables is a lower limit to the communality of that variable. This procedure is widely used in computations with digital computers.

Results from the analyses were inspected for indications of relevance of Guttman's [1954] lower bounds for the number of factors to be extracted. Further, the series of characteristic roots obtained from each factoring was inspected for indications to a solution for the number of factors to be extracted problem.

Output factor spaces were compared with the input factor spaces for the major factor domain by the procedures described in the next section.

3. *Methods for Comparison of Output Factors with Input Factors of the Major Factor Domain*

A major procedural problem involves the comparison of output factors with the actual input factors of the major factor domain. A rephrasing of this problem is: how well do dimensions in the output factor space represent dimensions in the actual input factor space? Two alternatives are considered: joint rotations of the actual input factors and the output factors to optimal matching, and rotations of the output factors only to optimal matching with each of the actual input factors. Note that the matrices A_1 of actual input factor loadings were used in these comparisons.

The first of these alternatives yields indications whether or not corresponding subspaces exist in the input factor space and the output factor space which are highly related. If r factors are extracted for the output factor space, a subspace of dimensionality Q', equal to or less than r, may be highly related to a Q' dimensional subspace of the input factor space; but, the $r - Q'$ dimensional complementary subspace in the output factor space may not be highly related to a corresponding subspace of the input factor space. Thus, the answer to the question would be that the output space included a subspace that represented a subspace of the input space, but that the output space included a complementary subspace which did not represent a subspace of the input space.

The second of the alternatives involves the representation of each input factor in the output factor space. If the output space projects onto a subspace of the input factor space, it would be desirable for this input subspace to be defined by a subset of the input factors. Since the method for construction of the input factor matrix involved a procedure which tended to produce a simple structure, the projection of the output space onto a subspace defined by a subset of the input factors would permit rotation of the output factor matrix to simple structure which, in turn, would permit identification of the individual input factors which define the input subspace. This is an important possibility for factor analytic procedures.

In order to investigate the problem of number of factors to extract for the output factor matrix, comparisons between the input and output factor spaces

https://doi.org/10.1007/BF02290601 Published online by Cambridge University Press

were computed for a series of number of factors extracted, in most cases $r = 1, 2, \cdots, M_1$, the number of input factors in the major domain. For a few of the correlation matrices, the series of number of factors extracted was extended beyond M_1.

Let F_r represent the output factor matrix for r factors extracted. Note that the methods of extraction of the factors result in matrices F_r of full column rank. In the following comparison techniques, it will be presumed that the input factor matrix A_1 is of full rank, this being true in fact for the matrices used in this study as demonstrated in Table 8 by none of the sums of squares of loadings on principal axes being zero. Rotation of axes for the output and input factor spaces are accomplished by (24) and (25).

$$\ddot{A}_{1r} = A_1 T_{Ar} \tag{24}$$

$$\ddot{F}_r = F_r T_{Fr} \tag{25}$$

where $\ddot{A}_{1r}$ and $\ddot{F}_r$ are the rotated matrices for r factors extracted, and where T_{Ar} and T_{Fr} are the corresponding transformation matrices to accomplish the rotations. Since coefficients of congruence, as described shortly, are used to compare columns of $\ddot{A}_{1r}$ with columns of $\ddot{F}_r$, the scaling of the columns of T_{Ar} and T_{Fr} is immaterial. Arbitrarily, the columns of the transformation matrices are defined as unit length vectors.

As mentioned in the preceding paragraph, the coefficient of congruence is used as a measure of similarity between a column of loadings in matrix $\ddot{A}_{1r}$ and a column of loadings in matrix $\ddot{F}_r$. This coefficient, which was first suggested by Burt [1949], named and used by Tucker [1951], and used by Wrigley and Neuhaus [1955], is the cosine of the angle between two column vectors of factor loadings in a space having an orthogonal axis for each variable. The form of this coefficient is given by (25).

$$\text{Coefficient of Congruence} = \frac{\sum_j a_{jm} f_{jp}}{\sqrt{\sum_j a_{jm}^2}\,\sqrt{\sum_j f_{jp}^2}} \tag{25}$$

and is analogous to a coefficient of correlation, but not identical since the loadings are not converted to deviations from their means.

In (25), no indication was given as to rotation of axes for either the input factor loadings a_{jm} or the output factor loadings f_{jp}. Rotated output factor loadings were considered for all cases used in the present study. In the first alternative method of comparison of the input and output factors, both the input and output factor matrices were rotated so as to maximize the squared coefficient of congruence which is designated for this case as ϕ^2. This solution is analogous to a canonical correlation solution and yields as many rotated dimensions as the number of factors in the input factor matrix or in the output factor matrix, whichever of these number of factors is less. Let Q be the lesser of M_1 and r. The matrices $\ddot{A}_{1r}$, $\ddot{F}_r$, T_{Ar}, and T_{Fr} will have Q columns which may be denoted by the subscript $q = 1, 2, \cdots, Q$. There will be Q values of ϕ^2, one corresponding to each value of q, which may be designated ϕ_q^2 and be entered as diagonal entries in a matrix Φ_r^2.

Let the columns of $\ddot{A}_{1r}$ and $\ddot{F}_r$ be so ordered that the ϕ_q^2 form a decreasing series. Note that each pair of corresponding columns of $\ddot{A}_{1r}$ and $\ddot{F}_r$ are unrelated to all other columns in these two matrices. Thus the series of corresponding pairs of columns of $\ddot{A}_{1r}$ and $\ddot{F}_r$ defines rotations in the input factor and output factor spaces which are maximally related. Further, consideration of any subset of pairs of columns for pairs $q = 1, 2, \cdots, Q'$ with $Q' < Q$ defines subspaces of dimensionality Q' in the two factor spaces which are maximally related. When one or more ϕ_q^2 are zero, say Q'' of them, the corresponding columns of $\ddot{A}_{1r}$ and $\ddot{F}_r$ define Q'' dimensional subspaces which are unrelated to each other as well as to the complementary subspaces in $\ddot{A}_{1r}$ and $\ddot{F}_r$. Consequently, this subspace of dimensionality Q'' in the output factor matrix F_r is unrelated to the entire major factor input factor matrix A_1. Low values of ϕ_q^2 would indicate dimensions in the output factor space with low relations to the major factor input factor matrix.

In the second alternative for comparison of the output factors with the input factors, only the output factors were rotated so as to obtain a maximum coefficient of congruence with each of the input factors. This solution is analogous with a multiple regression solution using the output factors as predictor variables to predict each of the input factors, in turn, as a criterion variable. The squared coefficient of congruence, denoted $R_{rm_1}^2$, with input factor m_1 is analogous to the squared multiple correlation.

A high value of $R_{rm_1}^2$ indicates that a rotation of axes is possible in the output space to a factor with loadings highly related to the loadings on a given input factor. A low value of $R_{rm_1}^2$ indicates that only a low relation may be obtained between a rotated output dimension and the given input factor. In case the number of factors, r, extracted in the output factor matrix is less than M_1, the number of input major factors, and, in case there are r input factors for which the $R_{rm_1}^2$ are high, then a rotation is possible of this output space to the selected input factors. In case the preceding were possible for several different selections as to number of output factors, the number of factors to extract problem would not be highly critical. The possibility exists that the very weak input factors would not be easily matched by output dimensions; low values of $R_{rm_1}^2$ for these input factors would result.

4. *Results*

The squared multiple correlation (commonly designated SMC) of each variable with all other variables in each battery was obtained for all twenty variables in all 54 batteries. These SMC's are summarized in Table 6 which presents the mean and standard deviation of the SMC's for each value of b_{1j}^2

Table 6

Summary of SMC's

Three Factors in Major Domain

b^2_{1j}	Number	Mean SMC Formal Model	Mean SMC Middle Model	Mean SMC Simulation Model	Standard Deviation Formal Model	Standard Deviation Middle Model	Standard Deviation Simulation Model
		High Range of b^2_{1j}					
.8	15	.749	.777	.834	.010	.021	.037
.7	21	.651	.679	.749	.021	.026	.040
.6	24	.556	.605	.709	.014	.025	.048
		Wide Range of b^2_{1j}					
.8	12	.716	.746	.820	.027	.026	.043
.7	12	.624	.670	.774	.020	.020	.026
.6	3	.510	.577	.697	.053	.061	.117
.5	9	.452	.509	.659	.009	.032	.073
.4	3	.364	.438	.626	.006	.013	.015
.3	6	.268	.338	.505	.003	.033	.091
.2	15	.176	.252	.467	.010	.033	.073
		Low Range of b^2_{1j}					
.4	24	.294	.363	.530	.023	.045	.095
.3	24	.226	.243	.491	.014	.042	.138
.2	12	.152	.232	.437	.010	.040	.092

Seven Factors in Major Domain

b^2_{1j}	Number	Mean SMC Formal Model	Mean SMC Middle Model	Mean SMC Simulation Model	Standard Deviation Formal Model	Standard Deviation Middle Model	Standard Deviation Simulation Model
		High Range of b^2_{1j}					
.8	12	.658	.684	.747	.056	.063	.075
.7	18	.577	.615	.701	.053	.056	.070
.6	21	.504	.548	.635	.035	.035	.054
		Wide Range of b^2_{1j}					
.8	12	.628	.656	.720	.093	.096	.123
.7	15	.558	.593	.685	.058	.063	.053
.6	15	.472	.519	.620	.038	.038	.052
.5	6	.360	.400	.507	.055	.051	.069
.4	6	.327	.380	.513	.034	.060	.089
.3	3	.236	.315	.502	.026	.037	.081
.2	3	.157	.229	.419	.023	.029	.031
		Low Range of b^2_{1j}					
.4	15	.189	.254	.415	.038	.054	.094
.3	21	.153	.220	.396	.032	.050	.089
.2	24	.109	.185	.377	.105	.027	.066

https://doi.org/10.1007/BF02290601 Published online by Cambridge University Press

for each model in each range of b_{1j}^2 used for each size of major domain.

As indicated by Guttman's [1954] proposition, every SMC for the formal model is less than the input b_{1j}^2 which are the theoretical communalities for this case. The mean differences between the b_{1j}^2 and the corresponding SMC's are greater for the seven factor major domain than for the three factor major domain. Further, the standard deviations of these differences, which equal the standard deviations of the SMC's given in Table 6, are larger for the seven factor major domain than for the three factor major domain. It appears, then, that the SMC's may be less adequate estimates of the communalities for the formal model for analyses involving larger ratios of number of common factors to number of variables than for analyses involving smaller such ratios.

A major feature of the SMC's is the progression in mean magnitude for each size of major domain, range of b_{1j}^2 , and value of b_{1j}^2 from least for the formal model to largest for the simulation model. This progression exists for every row in Table 6 without exception. While all SMC's for the formal model are less than the corresponding b_{1j}^2 , some of the SMC's for the middle model and many of them for the simulation model are greater than the input b_{1j}^2 . It is not possible, however, to equate the b_{1j}^2's in the latter two cases with "true" communalities and to make statements as to whether the SMC's are over-estimates or under-estimates as can be done in the case for the formal model. An appropriate question is whether the SMC's are effective coefficients to be inserted into the diagonals of correlation matrices as a step in factor analyses so that the results of the factor analyses reflect the input factors in the major domain. Some answers to this question are given by the results of the factor analyses of the correlation matrices for the middle and simulation models to be discussed in subsequent paragraphs.

A decreasing relation of the SMC's to the b_{1j}^2 from the formal model to the middle model to the simulation model for a given size of major factor domain and range of b_{1j}^2 is shown by two features of the results given in Table 6. First, while the mean SMC appears linearly related to the b_{1j}^2 for each size of major domain, range of b_{1j}^2 , and model, the slope of the regression of SMC on b_{1j}^2 appears to reduce from the formal model to the middle model to the simulation model. Second, the standard deviations of the SMC's for each given value of b_{1j}^2 , range of b_{1j}^2 , and size of major domain tend to increase from the formal model to the middle model to the simulation model. This progression is true for the majority of the rows in Table 6.

As already indicated, the mean SMC's for the formal model tend to be larger for the three factor major domain than for the seven factor major domain. This relation is true also for the middle model and for the simulation model. This observation, while noted here, will receive no further attention in this report. It deserves further intensive study as a separate subject.

A summary of the series of characteristic roots obtained in the principal axes factoring of the correlation matrices is presented in Table 7. Inspection of the series of roots obtained for the three correlation matrices in each cell of the experimental design indicated that mean values of corresponding roots over the three replications could be used without loss of general features of the series.

Guttman's [1954] lower bounds for the number of factors have been indicated by asterisks after the appropriate roots for all series of roots in Table 7. For the series of roots for factoring the correlation matrices with unity in the diagonal cells (designated RwU) the lower bound is the last root equal to or greater than unity, while for factoring the correlation matrices with squared multiple correlations in the diagonal cells (designated RwSMC) the lower bound is the last non-negative root. The lower bounds for factoring the C matrices and for factoring the RwSMC matrices must be identical and are Guttman's stronger lower bounds for the number of factors.

Each series of roots was inspected for breaks in the rate of decrease of the roots and such breaks have been indicated by lines drawn between two roots. In several of the series for factoring the C and RwSMC matrices, two breaks in the rate of decrease of the roots were observed and have been indicated. These breaks in the rate of decrease of the roots have been used as indications of the number of factors in the correlation matrices.

For the high range of b_{1j}^2 and three factors in the major domain, Guttman's weaker lower bound (asterisked root in column RwU) is three factors for all three models while Guttman's stronger lower bound (asterisked roots in columns C and RwSMC) is 3, 8, and 12 factors for the formal model, middle model, and simulation model, respectively. The breaks in the series of roots appear after three factors for all these series of roots. It appears, then, that both Guttman's lower bounds and the breaks in the series of roots indicate correctly there are three common factors for the formal model. In contrast, for the middle model and simulation model, Guttman's weaker lower bound and the breaks in the series of roots are accurate indicators of the number of factors in the major domain while Guttman's stronger lower bound indicates a number of more factors. This result, undoubtedly, is due to the inclusion of the influences of the many minor factors on the correlations which result in there being many more common factors. However, if the purpose of the analyses is to explore the factor structure in the major factor domain, the output of more factors than the number in the major domain may be undesirable. Consequently, Guttman's stronger lower bound would result in "over-factoring".

For the high range of b_{1j}^2 and the seven factor major domain, Guttman's weaker lower bound indicates six factors for all models; Guttman's stronger lower bound indicates 7, 8, and 12 factors for the three models. Breaks in the series of roots indicate seven factors for all models. Thus: Guttman's weaker lower bounds understate, Guttman's stronger lower bounds overstate, and the

Table 7

Characteristic Roots for Principal Axes Factoring
Mean Values over Three Replications

High Range of b_{1j}^2

Three Factors in Major Domain

Factor Number	Formal Model RwU	Formal Model C	Formal Model RwSMC	Middle Model RwU	Middle Model C	Middle Model RwSMC	Simulation Model RwU	Simulation Model C	Simulation Model RwSMC
1	8.32	25.67	7.98	8.56	29.39	8.25	8.79	41.73	8.56
2	3.61	10.86	3.23	3.62	12.13	3.31	3.65	16.43	3.41
3	2.71*	6.89*	2.31*	2.69*	7.58	2.32	2.66*	10.32	2.39
4	.40	.94	-.03	.53	1.47	.16	.69	2.78	.43
5	.40	.93	-.03	.48	1.32	.11	.58	2.29	.31
6	.40	.91	-.03	.43	1.21	.07	.50	1.98	.24
7	.40	.90	-.03	.38	1.10	.03	.40	1.66	.15
8	.40	.90	-.03	.38	1.05 *	.01*	.38	1.48	.12
9	.38	.90	-.04	.34	.99	-.00	.34	1.36	.08
10	.36	.89	-.04	.32	.95	-.02	.31	1.24	.05
11	.32	.88	-.04	.30	.91	-.03	.28	1.13	.03
12	.30	.87	-.04	.28	.87	-.04	.25	1.01*	.00*
13	.30	.86	-.05	.27	.81	-.06	.23	.85	-.04
14	.30	.85	-.05	.26	.79	-.07	.20	.83	-.04
15	.30	.84	-.05	.24	.78	-.07	.18	.78	-.05
16	.26	.83	-.05	.23	.72	-.08	.15	.64	-.07
17	.23	.82	-.06	.20	.71	-.09	.13	.61	-.09
18	.21	.81	-.06	.17	.66	-.10	.10	.50	-.11
19	.20	.80	-.07	.16	.64	-.12	.09	.45	-.13
20	.20	.76	-.07	.15+	.62	-.13	.08	.37	-.15

Seven Factors in Major Domain

Factor Number	Formal Model RwU	Formal Model C	Formal Model RwSMC	Middle Model RwU	Middle Model C	Middle Model RwSMC	Simulation Model RwU	Simulation Model C	Simulation Model RwSMC
1	5.18	13.63	4.77	5.43	15.76	5.06	5.69	22.14	5.40
2	3.13	8.46	2.75	3.11	9.16	2.76	3.09	11.80	2.81
3	2.23	5.54	1.81	2.24	6.07	1.85	2.25	7.95	1.95
4	1.91	4.42	1.46	1.89	4.32	1.48	1.87	6.24	1.55
5	1.54	3.41	1.08	1.52	3.74	1.11	1.49	4.70	1.17
6	1.28*	2.86	.82	1.29*	3.12	.87	1.31*	3.94	.97
7	.81	1.80*	.36*	.80	1.96	.39	.80	2.56	.47
8	.40	.85	-.06	.46	1.13*	.05-*	.56	1.76	.19
9	.39	.84	-.07	.41	1.00	-.00	.46	1.44	.13
10	.37	.82	-.08	.37	.94	-.02	.40	1.34	.10
11	.37	.78	-.08	.35	.89	-.04	.37	1.17	.06
12	.35	.77	-.09	.33	.82	-.06	.32	1.01*	.00*
13	.32	.75	-.10	.30	.78	-.08	.29	.90	-.03
14	.30	.73	-.10	.28	.72	-.10	.25	.79	-.07
15	.28	.71	-.12	.25	.69	-.12	.22	.71	-.09
16	.27	.70	-.12	.24	.62	-.14	.17	.58	-.12
17	.24	.67	-.14	.21	.59	-.15	.15	.55	-.14
18	.23	.65	-.15	.20	.58	-.16	.13	.47	-.15
19	.20	.60	-.16	.18	.53	-.18	.11	.41	-.16
20	.20	.56	-.19	.15	.49	-.19	.08	.35	-.19

https://doi.org/10.1007/BF02290601 Published online by Cambridge University Press

Table 7 (Continued)

Characteristic Roots for Principal Axes Factoring
Mean Values over Three Replications

Wide Range of b_{ij}^2

Three Factors in Major Domain

Factor Number	Formal Model RwU	Formal Model C	Formal Model RwSMC	Middle Model RwU	Middle Model C	Middle Model RwSMC	Simulation Model RwU	Simulation Model C	Simulation Model RwSMC
1	5.40	14.19	4.95	5.75	16.60	5.33	6.12	26.04	5.82
2	3.73	9.11	3.28	3.71	10.22	3.30	3.69	14.89	3.40
3	2.21*	4.55*	1.67*	2.21*	5.31	1.74	2.25	8.13	1.93
4	.80	.98	-.02	.98	1.50	.30	1.18	2.93	.72
5	.80	.97	-.02	.89	1.41	.23	1.06*	2.62	.61
6	.79	.97	-.02	.82	1.30	.17	.90	2.24	.48
7	.76	.96	-.03	.74	1.19	.10	.78	1.94	.34
8	.73	.96	-.03	.66	1.12	.06	.69	1.77	.26
9	.68	.95	-.03	.60	1.05*	.02*	.55	1.57	.17
10	.61	.95	-.04	.56	.99	-.00	.51	1.38	.13
11	.55	.92	-.04	.52	.92	-.04	.42	1.19	.06
12	.50	.91	-.05	.45	.88	-.06	.38	1.05*	.02*
13	.48	.90	-.05	.40	.85	-.07	.30	.98	-.01
14	.40	.86	-.06	.35	.79	-.09	.28	.85	-.04
15	.33	.83	-.06	.30	.75	-.11	.22	.73	-.08
16	.30	.81	-.07	.27	.72	-.12	.19	.66	-.10
17	.28	.78	-.07	.23	.67	-.13	.15	.55	-.14
18	.23	.75	-.08	.21	.64	-.15	.13	.48	-.15
19	.22	.73	-.09	.19	.62	-.17	.11	.43	-.18
20	.20	.70	-.12	.16	.57	-.19	.07	.33	-.22

Seven Factors in Major Domain

Factor Number	Formal Model RwU	Formal Model C	Formal Model RwSMC	Middle Model RwU	Middle Model C	Middle Model RwSMC	Simulation Model RwU	Simulation Model C	Simulation Model RwSMC
1	5.05	11.64	4.58	5.31	13.32	4.88	5.59	17.99	5.26
2	2.65	6.11	2.18	2.67	6.68	2.24	2.72	8.70	2.38
3	2.28	4.63	1.77	2.26	5.02	1.79	2.26	6.46	1.88
4	1.70	3.13	1.14	1.71	3.46	1.19	1.73	4.48	1.32
5	1.23	2.22	.66	1.23	2.44	.71	1.28	3.27	.87
6	1.05*	1.87	.47	1.07*	2.12	.54	1.08*	2.80	.67
7	.83	1.31*	.18*	.84	1.38	.21	.87	1.92	.40
8	.68	.93	-.05	.68	1.20	.10	.77	1.81	.32
9	.57	.89	-.07	.61	1.11*	.06*	.64	1.55	.22
10	.54	.87	-.08	.51	.98	-.01	.51	1.31	.12
11	.48	.84	-.09	.46	.86	-.05	.43	1.10*	.03*
12	.44	.79	-.10	.40	.83	-.07	.36	.99	-.00
13	.41	.77	-.11	.37	.78	-.10	.34	.91	-.03
14	.37	.74	-.11	.35	.74	-.12	.30	.82	-.07
15	.35	.73	-.12	.33	.71	-.14	.26	.72	-.11
16	.32	.71	-.13	.30	.68	-.15	.25	.68	-.12
17	.30	.69	-.14	.27	.65	-.17	.21	.58	-.16
18	.28	.66	-.16	.24	.59	-.19	.16	.48	-.19
19	.25	.63	-.17	.21	.55	-.20	.13	.42	-.20
20	.21	.60	-.19	.18	.50	-.22	.10	.34	-.24

https://doi.org/10.1007/BF02290601 Published online by Cambridge University Press

Table 7 (Continued)

Characteristic Roots for Principal Axes Factoring
Mean Values over Three Replications

Low Range of b_{1j}^2

Three Factors in Major Domain

Factor Number	Formal Model RwU	C	RwSMC	Middle Model RwU	C	RwSMC	Simulation Model RwU	C	RwSMC
1	4.42	5.96	3.67	4.89	7.37	4.22	5.40	11.84	4.92
2	2.23	2.95	1.47	2.29	3.36	1.60	2.40	5.08	1.90
3	1.74*	2.30*	.98*	1.68	2.45	.99	1.69	3.59	1.21
4	.80	.95	-.04	1.07*	1.58	.39	1.41	3.05	.93
5	.79	.94	-.05	.97	1.41	.27	1.19	2.40	.67
6	.78	.94	-.05	.90	1.24	.17	1.05*	2.04	.52
7	.77	.93	-.06	.84	1.19	.13	.96	1.86	.44
8	.70	.92	-.06	.77	1.12	.08	.84	1.63	.32
9	.70	.91	-.07	.74	1.05 *	.03*	.77	1.45	.24
10	.70	.91	-.07	.68	.97	-.02	.67	1.25	.13
11	.70	.90	-.07	.63	.90	-.07	.59	1.06*	.04*
12	.70	.89	-.08	.60	.87	-.09	.51	.99	-.01
13	.68	.89	-.08	.58	.84	-.11	.48	.91	-.05
14	.66	.88	-.08	.55	.81	-.13	.42	.84	-.08
15	.63	.88	-.09	.55	.78	-.15	.39	.75	-.13
16	.60	.87	-.09	.51	.75	-.17	.34	.70	-.15
17	.60	.86	-.10	.48	.72	-.19	.30	.61	-.20
18	.60	.86	-.10	.45	.69	-.21	.24	.52	-.22
19	.60	.84	-.12	.43	.66	-.23	.19	.45	-.26
20	.60	.82	-.13	.41	.63	-.25	.16	.37	-.30

Seven Factors in Major Domain

Factor Number	Formal Model RwU	C	RwSMC	Middle Model RwU	C	RwSMC	Simulation Model RwU	C	RwSMC
1	2.98	3.57	2.14	3.42	4.49	2.65+	3.92	6.96	3.34
2	1.69	1.99	.84	1.77	2.32	1.00+	2.07	3.64	1.49
3	1.48	1.75	.63	1.53	1.96	.75-	1.71	2.81	1.10
4	1.30	1.50	.43	1.38	1.74	.59	1.50	2.51	.89
5	1.17	1.36	.31	1.18	1.48	.38	1.29	2.10	.67
6	1.10*	1.24	.21	1.07	1.33	.26	1.15	1.92	.55
7	.93	1.07*	.06*	1.00 *	1.26	.20	1.10*	1.76	.47
8	.80	.91	-.08	.96	1.20	.16	.99	1.60	.37
9	.79	.90	-.09	.88	1.10	.08	.91	1.48	.29
10	.79	.90	-.09	.84	1.05 *	.04*	.83	1.33	.20
11	.78	.88	-.10	.77	.98	-.01	.76	1.23	.14
12	.76	.87	-.11	.71	.90	-.08	.64	1.05*	.03*
13	.75	.86	-.12	.67	.87	-.10	.59	.96	-.03
14	.72	.85	-.13	.64	.81	-.15	.51	.83	-.11
15	.70	.84	-.14	.60	.78	-.17	.47	.79	-.13
16	.70	.82	-.15	.57	.72	-.22	.41	.66	-.21
17	.67	.80	-.16	.53	.69	-.24	.35	.58	-.25
18	.65	.79	-.18	.52	.67	-.26	.31	.52	-.28
19	.64	.77	-.19	.50	.64	-.28	.27	.47	-.32
20	.62	.76	-.20	.45	.61	-.30	.21	.39	-.35

breaks in the series of roots give an accurate indication of the number of factors in the major domain.

For the wide range of b_{1j}^2 and three factors in the major domain, Guttman's weaker lower bound indicates three factors for the formal and middle models but indicates five factors for the simulation model, thus, indicating an over-estimate of the number of factors in the major domain. Again, Guttman's stronger lower bound yields over-estimations of the number of factors in the major domain for the middle and simulation models. The breaks in the series of roots are all accurate at three factors for all models. For the wide range of b_{1j}^2 and seven factors in the major domain, Guttman's weaker lower bound indicates six factors for all models, thus being an under-estimate, while Guttman's stronger lower bound yields over-estimates for the middle and simulation models of the number of factors in the major domain. Breaks in the series of roots has become a more ambiguous judgment than for the previous cases considered with two breaks seeming to occur for a number of the series. The breaks appear to be more obvious for the series for factoring the matrices C and RwSMC than for factoring the matrix RwU. It appears that the breaks in the series for the formal model for factoring the matrices C and RwSMC give accurate indications of the number of factors while the breaks for all series for the middle model and simulation model yield under-estimates of the number of factors in the major domain.

Results for the low range of b_{1j}^2 for the formal model, for both the three factor major domain and the seven factor major domain, appear to be quite similar to the results for the high range of b_{1j}^2 in yielding accurate indications of the number of factors. For the middle model and the simulation model, results for the low range of b_{1j}^2 are even more ambiguous than for the wide range of b_{1j}^2. Guttman's lower bounds all yield over-estimates of the number of factors in the major domain while the breaks in the series of roots are ambiguous except for the middle model and three factors in the major domain. These breaks yield under-estimates of the number of factors in the major domain with the previous exception of the middle model and three factors in the major domain.

While the indications as to number of factors have been discussed in the preceding paragraphs in terms of the number of factors in the major domains, further evaluations of these indications may be made in terms of the relations of the output factor matrices to the input major domain factor matrices. It is distinctly possible that as many factors are needed as indicated by Guttman's stronger lower bound in order to obtain a good representation of the input major domain factors in the output factor matrices. In contrast, stopping factoring at fewer factors such as indicated by the breaks in the series of roots may yield an output factor space that projects satisfactorily onto a subspace of the input major domain factor space. These possibilities will be considered in connection with the results of matching the output factors with the input major domain factors.

The output factor matrices were matched with the input factor matrices by the method of maximum congruence described in the procedures section. Consider Table 8: it presents mean coefficients over three replications for the factoring of matrices RwSMC for seven factors in the major domain and the low range of b_{1i}^2 separately for the three models. When only the first output factor was considered for the formal model and a transformation of the input factors was performed to an input dimension having maximum congruence with this single output factor, ϕ^2's of .99975, .99976, and .99949 for the three replication matrices were obtained. The mean of these values is .99967 which is listed at the left of the first row for the formal model in Table 8. When the first two output factors were considered and transformations were performed on both the input and output factors to two pairs of maximum congruent dimensions, the mean ϕ^2's given in the second row of Table 8 were obtained. Thus, each row of Table 8 corresponds to utilization of the specified number of output factors and lists, in descending order, the mean ϕ^2's over the three replications. There are as many ϕ^2's in each row as the number of output factors considered or as the number of factors in the major factor domain, whichever of these numbers is less. These tables are designated as the complete tables of ϕ^2.

An interesting property of the complete tables of ϕ^2 is that every column must be non-decreasing; that is, each entry is equal to or larger than the entry above it. A second interesting and important property is that the entries in the diagonal from upper left to lower right of each table must be non-increasing. As a consequence of these properties and the listing of the values in decreasing order in each row is that the least value in the table is that last diagonal entry. Thus, for the formal model in Table 8 the least entry is .94191 as an average over the three replications.

The ϕ^2's in Table 8 for the formal model all appear to be relatively high so that it may be concluded that the seven factor output space has a close agreement with the input major domain factor space. However, for the middle model the ϕ^2's in the diagonal decrease from a first entry of .98736 to a seventh entry of .32860, indicating a reducing agreement between the output factor space and the most related subspace of the input factor space. A judgment might be made that the first five factors extracted define a space with a satisfactory projection on a five dimensional subspace of the input factor space since the fifth diagonal is moderately high at .79695. It is interesting to note that for the series of roots in Table 7 for low range of b_{1i}^2, middle model, factoring of RwSMC there appears to be a possible break in the rate of decrease in the roots after four or five factors. If factor extraction had been stopped at four or five factors, the ϕ^2's in Table 8 indicate that an output factor space would have been obtained that projected satisfactorily on a

Table 8

Complete Tables of ϕ^2

Low Range of b^2_{1j}, Seven Factor Major Domain, Factoring of RwSMC

Mean Values over Three Replications.

Formal Model

Number of Output Factors	Congruent Dimensions 1	2	3	4	5	6	7
1	.99967						
2	.99968	.99780					
3	.99968	.99809	.99528				
4	.99969	.99832	.99619	.99391			
5	.99973	.99852	.99715	.99440	.98727		
6	.99973	.99868	.99720	.99543	.99049	.98490	
7	.99974	.99885	.99724	.99576	.99148	.98638	.94191

Middle Model

Number of Output Factors	Congruent Dimensions 1	2	3	4	5	6	7
1	.98736						
2	.98921	.92439					
3	.99142	.95198	.91000				
4	.99314	.95961	.93144	.80423			
5	.99362	.96079	.93994	.87644	.79695		
6	.99458	.97383	.94799	.89447	.82199	.63173	
7	.99601	.98299	.96296	.92184	.88086	.64929	.32860
8	.99639	.98839	.97146	.93646	.90223	.66629	.57828
9	.99751	.99185	.97670	.94546	.90759	.82508	.59609
10	.99868	.99438	.98508	.96597	.92150	.87775	.65108

Simulation Model

Number of Output Factors	Congruent Dimensions 1	2	3	4	5	6	7
1	.96433						
2	.96866	.77418					
3	.97758	.84726	.63537				
4	.98144	.85504	.77521	.60046			
5	.98214	.87716	.79824	.62499	.38629		
6	.98581	.89480	.83424	.63605	.47240	.17910	
7	.98781	.93726	.87311	.76535	.59087	.32348	.00371
8	.98882	.96598	.90309	.81912	.72421	.39597	.23882
9	.99423	.97369	.90960	.86394	.78525	.45048	.30524
10	.99687	.98487	.95308	.89656	.80774	.63465-	.41101
11	.99880	.98904	.97017	.93586	.85479	.71068	.48631
12	.99953	.99103	.97608	.95027	.92629	.74975+	.56318

subspace of the major factor domain input factors. Factor extraction was continued for the middle model, low range of b^2_{1j} to ten factors, the stronger lower bound as to number of factors. The least ϕ^2 for ten factors is .65108 indicating only moderate relation between the last dimension of a subspace of seven dimensions with the input factor space. With extraction of ten

https://doi.org/10.1007/BF02290601 Published online by Cambridge University Press

factors, there appears to be a six-dimensional subspace adequately related to a corresponding subspace of the input factors.

The ϕ^2's in Table 8 for the simulation model are much lower than for the other two models with the least ϕ^2 getting down to .00371. Even with extraction of 12 factors the least ϕ^2 increases to only .56318. Inspection of the diagonal indicates a possible solution after extraction of four factors such that the least ϕ^2 is .60046. An inspection of the corresponding series of roots in Table 7 for the low range of b_{1i}^2, seven factors in the major domain, simulation model, factoring of RwSMC indicates a moderate break in the rate of decrease of the roots after four factors. In case factor extraction had been stopped after extraction of four factors, a factor space would have been obtained which projected moderately onto a four-dimensional subspace of the input factor space.

Table 9 presents the diagonal entries of the complete tables of ϕ^2 for all cells in the experimental design. These entries are the means over the three replications for each cell.

For three factors in the major domain the dominant feature of the ϕ^2's is that almost all are high (greater than .94). Lower values of ϕ^2 occur only for the low range of b_{1i}^2 for the simulation model and when three factors are considered in the output. The preceding is true for factoring of all three matrices: RwU, C, and RwSMC. If factor extraction had been terminated after two factors, which is the first break in the series of roots for this case, quite high ϕ^2's in the range of .87–.89 would have been obtained for the two-dimensional output space as related to the corresponding two-dimensional subspace of the major domain input factor space. In case factoring had been continued to four dimensions, the second break in the root series, the least ϕ^2 for the three-dimensional subspace of the output factor space would be .78989, .71527, and .75001 for factoring of the RwU, C, and RwSMC matrices, respectively. In case factoring had been continued to the Guttman weaker lower bound of six factors, the corresponding three values of ϕ^2 would have risen to .90244, .89214, and .90050. For this possibility there is a three-dimensional subspace in each output factor space which has zero relation with the input major domain factor space. A question not considered in the present analysis is whether rotational procedures applied to the output factor matrix which did not utilize the input factor information would be successful in separating the three-dimensional subspace which is related to the input factors from the three-dimensional space that has zero relation with the input factors. Answers to this question could be the subject of further research.

One surprise in the ϕ^2 results is that the procedures designed to correct for communalities failed to materially increase the ϕ^2's over those obtained for factoring the correlation matrices with unity in the diagonal cells. The possibility exists that other procedures involving communality estimation

Table 9

Diagonals of ϕ^2 Tables
Means over Three Replications

High Range of b_{1j}^2

Three Factors in Major Domain

Congruent	Formal Model			Middle Model			Simulation Model		
Dimensions	RwU	C	RwSMC	RwU	C	RwSMC	RwU	C	RwSMC
1	.99991	.99999	.99999	.99953	.99973	.99971	.99869	.99890	.99881
2	.99949	.99997	.99998	.99861	.99889	.99893	.99596	.99498	.99573
3	.99923	.99996	.99998	.99807	.99860	.99856	.99418	.99398	.99419

Seven Factors in Major Domain

Congruent	Formal Model			Middle Model			Simulation Model		
Dimensions	RwU	C	RwSMC	RwU	C	RwSMC	RwU	C	RwSMC
1	.99977	.99995	.99995	.99935	.99957	.99960	.99828	.99847	.99852
2	.99925	.99986	.99989	.99872	.99927	.99922	.99677	.99715	.99698
3	.99878	.99962	.99960	.99657	.99801	.99785	.99119	.99344	.99254
4	.99741	.99889	.99900	.99390	.99702	.99590	.98576	.98831	.98555
5	.99621	.99818	.99814	.99275	.99338	.99354	.98172	.97175	.97825
6	.99501	.99750	.99770	.99133	.98975	.98919	.97466	.95991	.96395
7	.98324	.99812	.99130	.96457	.96686	.96729	.90142	.81991	.86875

Wide Range of b_{1j}^2

Three Factors in Major Domain

Congruent	Formal Model			Middle Model			Simulation Model		
Dimensions	RwU	C	RwSMC	RwU	C	RwSMC	RwU	C	RwSMC
1	.99875	.99995	.99997	.99444	.99804	.99708	.98615	.99268	.98872
2	.99631	.99985	.99993	.99270	.99759	.99676	.98337	.99097	.98729
3	.98503	.99891	.99943	.97985	.99029	.98934	.94283	.95610	.95058

Seven Factors in Major Domain

Congruent	Formal Model			Middle Model			Simulation Model		
Dimensions	RwU	C	RwSMC	RwU	C	RwSMC	RwU	C	RwSMC
1	.99934	.99992	.99993	.99749	.99881	.99832	.99380	.99588	.99485
2	.99761	.99960	.99972	.99374	.99749	.99706	.98327	.98952	.97097
3	.99496	.99927	.99954	.98957	.99532	.99502	.97436	.98333	.98096
4	.98304	.99672	.99714	.97350	.98827	.98813	.93508	.95188	.94713
5	.96845	.99372	.99446	.94186	.97899	.97403	.84867	.88043	.88194
6	.89071	.98741	.99165	.83374	.94237	.93863	.71541	.81729	.78698
7	.57157	.92378	.93343	.44213	.64028	.58632	.30130	.06588	.21989

Low Range of b_{1j}^2

Three Factors in Major Domain

Congruent	Formal Model			Middle Model			Simulation Model		
Dimensions	RwU	C	RwSMC	RwU	C	RwSMC	RwU	C	RwSMC
1	.99967	.99995	.99997	.99357	.99313	.99340	.97911	.97071	.97605
2	.99805	.99958	.99964	.97066	.97501	.97410	.88784	.87633	.88667
3	.99492	.99915	.99927	.95496	.95064	.95319	.74501	.66242	.69711

Seven Factors in Major Domain

Congruent	Formal Model			Middle Model			Simulation Model		
Dimensions	RwU	C	RwSMC	RwU	C	RwSMC	RwU	C	RwSMC
1	.99909	.99961	.99967	.98764	.98723	.98736	.96671	.96199	.96433
2	.99544	.99756	.99780	.91972	.92580	.92439	.76982	.78170	.77418
3	.99011	.99488	.99528	.87012	.91585	.91000	.64486	.63258	.63537
4	.98674	.99317	.99391	.79914	.80102	.80423	.60574	.58847	.60046
5	.97768	.98570	.98727	.78200	.79440	.79695	.40616	.37465	.38629
6	.96347	.98344	.98490	.60285	.63596	.63173	.23784	.16157	.17910
7	.87802	.93202	.94191	.34950	.32819	.32860	.01255	.00267	.00371

would be more effective. This possibility should be investigated in subsequent research.

Two important trends appear in the ϕ^2's in Table 9 for seven factors in the major domain. First, for each range of b_{1j}^2 and matrix factored, there is a downward trend of the ϕ^2's from the formal model to the middle model to the simulation model. Second, for the middle model and for the simulation model, there is a downward trend in the ϕ^2's from the high to wide to low range of b_{1j}^2. Corrections for communalities by factoring the matrices C and RwSMC as compared with factoring the matrix RwU leads to incresaes of corresponding low ϕ^2's only for the wide range of b_{1j}^2. The effect is greatest for the formal model, still strong for the middle model, and evident only for the fifth and sixth ϕ^2 for the simulation model. The possibility of stopping factor extraction of the matrices for the low range of b_{1j}^2 has been discussed. A similar possibility appears for the results for the wide range of b_{1j}^2, especially for the simulation model for which the sixth ϕ^2 appears to indicate a moderate correspondence with a six-dimensional subspace of the input factors, particularly for factoring of the C matrix or the RwSMC matrix. This possibility corresponds to a possible break in the rate of decrease of the corresponding series of roots in Table 7. Stopping factor extraction after four factors, the first break noted in the root series, would yield quite high ϕ^2's. A possible similar effect occurs for the middle model for the wide range of b_{1j}^2.

The second method for comparing the output factor spaces with the input major domain factors was to consider each input factor separately and determine a transformed output factor which had maximum congruence with the single input factor being considered. These coefficients of congruence are termed R^2 due to their similarity to squared multiple correlations of each input factor on the output factors as predictors. The only difference between the present R^2 and the squared multiple correlation is that no corrections are made for mean loadings for the present R^2. In summarizing the R^2's as in Table 10, the entries for each number of factors extracted were ordered into a decreasing sequence separately for each factoring of the 54 correlation matrices. Mean R^2's were obtained, then, over the three replications for each cell in the experimental design. Thus, the first entry in the first row for the formal model in Table 10 is the mean largest R^2 in the first row of the corresponding three replications. The second entry in this row is the mean second largest R^2 in the first row of the corresponding replications. It is to be noted that the ordering of the R^2's could involve a different ordering of the input factors from one number of factors extracted to another number of factors extracted for the same replication and from one replication to another. Table 10 presents the complete tables of ordered R^2's for the low range of b_{1j}^2, seven factor major domain, and factoring of the RwSMC matrices. The three sections are for the three models.

Note that the entries in every column of each section of Table 10 form an increasing sequence. It can be shown that these sequences must be at least non-decreasing, allowing for ties.

Diagonal entries in each section of Table 10 have been underlined since they have special meaning in the present context. In case factor extraction is

Table 10

Complete Tables of Ordered R^2

Low Range of b_{1j}^2, Seven Factor Major Domain, Factoring of RwSMC

Mean Values over Three Replications*

Formal Model

Number of Output Factors	Input Factors Ordered by R^2: 1	2	3	4	5	6	7
1	.66461	.47125	.40511	.34844	.29269	.23650	.15874
2	.77584	.69317	.58637	.56540	.48231	.36336	.27251
3	.91421	.84462	.75559	.69993	.60402	.49233	.33773
4	.96572	.93759	.91356	.82662	.70697	.62380	.41999
5	.98395	.96906	.95599	.92589	.88392	.75305	.61093
6	.99474	.98960	.98058	.96913	.95158	.94054	.78786
7	.99766	.99685	.99657	.99540	.99437	.99286	.98252

Middle Model

Number of Output Factors	Input Factors Ordered by R^2: 1	2	3	4	5	6	7
1	.63526	.42985	.41501	.35388	.27722	.23010	.18999
2	.81827	.62044	.57636	.46776	.41794	.31227	.27626
3	.91894	.78772	.69943	.63525	.54299	.49126	.38613
4	.93723	.88072	.78276	.75336	.66536	.62258	.43605
5	.94171	.91672	.88327	.86576	.77297	.74079	.57768
6	.94931	.93718	.91302	.89221	.87143	.82233	.68564
7	.96703	.95853	.92854	.91162	.89799	.89156	.76479
8	.97366	.96320	.96171	.92911	.92012	.90149	.80735
9	.98112	.97492	.96880	.95033	.93757	.92925	.84383
10	.98930	.97969	.97483	.96678	.95705	.94204	.85576

Simulation Model

Number of Output Factors	Input Factors Ordered by R^2: 1	2	3	4	5	6	7
1	.59819	.42074	.38331	.34174	.26542	.23262	.20082
2	.76910	.61481	.47345	.40572	.35764	.29656	.28521
3	.84202	.69081	.60700	.55202	.45855	.43124	.34484
4	.84798	.77712	.70329	.68324	.60030	.55069	.41067
5	.86032	.80156	.73481	.72682	.66810	.60903	.43865
6	.89422	.80503	.74051	.73331	.69710	.65030	.53322
7	.92202	.89833	.79341	.75798	.74109	.71272	.59316
8	.93016	.91828	.88111	.83576	.80068	.76480	.64803
9	.94240	.92697	.89751	.86121	.83266	.79009	.67988
10	.96647	.94193	.93486	.91343	.88003	.82007	.73703
11	.97430	.95553	.94513	.93583	.91390	.88620	.77093
12	.97453	.96207	.95548	.95023	.93703	.91467	.77180

* For each battery analysed (replication), the input factors were arranged in decreasing order of R^2, separately for each number of output factors. The mean values of R^2 were computed following this ordering.

https://doi.org/10.1007/BF02290601 Published online by Cambridge University Press

terminated after some particular number of factors less than the number of input factors, then the R^2's in that row of the table from the left to, and including the diagonal R^2 are indicative of degree of correspondence between rotated output factors and the same number of unrotated input factors. For example, if factor extraction had been terminated for the formal model for the case considered in Table 10 after five factors, the first five entries in row 5 of this section of Table 10 give the mean R^2's with five input factors ordered by decreasing value of R^2. Since all five of these entries are high, there exists a rotation in the five factor output space which represents quite well the five input factors. A consequence of this fact for the fifth row for the formal model is that valid psychological conclusions probably could be made concerning five of the input factors even though not all dimensions of the input space were represented. The diagonal entries give the lowest value for a group of R^2's equal in number to the number of factors extracted when this group is selected as those having the highest R^2's possible from among the R^2's for all input factors. Thus, the diagonal entries are usable indices for minimum relation between rotated output factors and selected input factors to which the rotated factors are maximally related. Table 11 presents the diagonal entries for the ordered R^2 tables for all cells in the experimental design. The entries are the mean values over the three replications for each cell.

While the diagonal entries in the ordered R^2 tables do not necessarily form an increasing or decreasing sequence, there is a very strong tendency for them to form an increasing sequence with increasing number of factors. Major exceptions occur for the simulation model for seven factors in the major domain for both the wide range and low range of b_{1i}^2 . In these cases the series of diagonal entries start increasing, seem to reach a maximum, and then decrease. This effect appears to a lesser degree for the middle model, seven factors in the major domain, and wide and low ranges of b_{1i}^2 . A most interesting observation is that these maxima in the series of ordered R^2 diagonals occur in the neighborhood of the number of factors for the last noted break in the series of roots in Table 7. A consequent possibility is that stopping factor extraction at the last observed break in the series of roots could lead to a rotation of the factors obtained so that the rotated factors correspond moderately to very well with an equal sized subset of the input factors in the major domain. It appears that extraction of one more or one fewer factors could still lead to rotated factors in the major domain. This statement must be limited to extraction of no more factors than exist in the major factor domain. A further note is that the subset of matching input factors may differ from the analysis of one battery to the analysis of another battery.

Resulting ordered R^2's in Table 10 for the middle model and simulation model, seven factors in the major domain, low range of b_{1i}^2 , factoring of the matrix RwSMC, have been listed for extraction of more factors than existed in the major factor domain. The number of extracted factors was carried to

Table 11

Diagonals of Ordered R^2 Tables

Mean Values over Three Replications

High Range of b^2_{1j}

Three Factors in Major Domain

	Formal Model			Middle Model			Simulation Model		
Factors	RwU	C	RwSMC	RwU	C	RwSMC	RwU	C	RwSMC
1	.72341	.70751	.72304	.72114	.70767	.72117	.70873	.71017	.71912
2	.83188	.82958	.83017	.81534	.82810	.81861	.81450	.83088	.81645
3	.99958	.99997	.99998	.99906	.99923	.99921	.99707	.99668	.99696

Seven Factors in Major Domain

	Formal Model			Middle Model			Simulation Model		
Factors	RwU	C	RwSMC	RwU	C	RwSMC	RwU	C	RwSMC
1	.52037	.56594	.52010	.52879	.55935	.52579	.53466	.55703	.53475
2	.72723	.73904	.72577	.73319	.72753	.73195	.73905	.68224	.73629
3	.74905	.74268	.75052	.76781	.75762	.75767	.78354	.78862	.78247
4	.80734	.79305	.81038	.80720	.79067	.81107	.80846	.82532	.80950
5	.86187	.85695	.86320	.86561	.84573	.85664	.86628	.84059	.84742
6	.92527	.93394	.93052	.93229	.93854	.93572	.93447	.93623	.93482
7	.99338	.99702	.99682	.98615	.98602	.95263	.95263	.93361	.94604

Wide Range of b^2_{1j}

Three Factors in Major Domain

	Formal Model			Middle Model			Simulation Model		
Factors	RwU	C	RwSMC	RwU	C	RwSMC	RwU	C	RwSMC
1	.55920	.67804	.55213	.55626	.65976	.55501	.54912	.69613	.51835
2	.76957	.78725	.77310	.76041	.79906	.77187	.74956	.82933	.76260
3	.99298	.99947	.99969	.98527	.99236	.99156	.95843	.96698	.96364

Seven Factors in Major Domain

1	.64310	.66871	.63847	.64839	.65372	.64589	.65006	.63492	.64977
2	.83534	.77400	.82534	.83739	.79142	.83086	.83371	.80259	.83080
3	.81674	.78526	.81140	.82899	.81118	.83265	.82162	.83183	.83475
4	.92907	.84328	.84017	.82444	.84753	.83303	.80926	.82174	.82313
5	.86854	.88586	.89271	.88685	.88667	.89612	.87742	.88095	.88642
6	.93894	.93965	.94817	.91431	.94523	.94829	.89150	.91708	.91126
7	.82840	.96279	.96461	.75516	.84513	.81723	.69404	.62913	.67419

Low Range of b^2_{1j}

Three Factors in Major Domain

	Formal Model			Middle Model			Simulation Model		
Factors	RwU	C	RwSMC	RwU	C	RwSMC	RwU	C	RwSMC
1	.71084	.72021	.71476	.70225	.67728	.70371	.68969	.65464	.68880
2	.86055	.85635	.85918	.86286	.85723	.86268	.83239	.80964	.82714
3	.99789	.99963	.99968	.97798	.97699	.97819	.88145	.83890	.85827

Seven Factors in Major Domain

	Formal Model			Middle Model			Simulation Model		
Factors	RwU	C	RwSMC	RwU	C	RwSMC	RwU	C	RwSMC
1	.66408	.66437	.66461	.63183	.63654	.63526	.59533	.59456	.59819
2	.69555	.68949	.69317	.63891	.61738	.62044	.60814	.61957	.61489
3	.73640	.76539	.75559	.71843	.69496	.69943	.54844	.60611	.60700
4	.82089	.82746	.82662	.75603	.75510	.75336	.70255	.67438	.68324
5	.87971	.88546	.88392	.74485	.78570	.77297	.66627	.66198	.66810
6	.92812	.93996	.94054	.82183	.81805	.82233	.66644	.63253	.65030
7	.95935	.97912	.98252	.78029	.76438	.76479	.61630	.58782	.59316

https://doi.org/10.1007/BF02290601 Published online by Cambridge University Press

Guttman's stronger lower bound. With the addition of more factors extracted, the R^2's increased to moderately high or very high values. This indicates that reasonably good matches could be obtained with the individual input major domain factors when a sufficient number of factors are extracted. The rotations made here, however, utilized knowledge of the input major domain factors and there remains a considerable question whether rotational procedures not using this information could determine the seven-dimensional subspaces which are congruent with the major domain input factors. This problem is generated by the existence in the output factor space of a subspace not related to the major domain input factor space.

A further interesting comparison is that of the diagonals of the ϕ^2 tables in Table 9 with the diagonals of the ordered R^2 tables in Table 11. As previously noted, the series of diagonals of ϕ^2 must be non-increasing and are usually decreasing. However, as previously noted, the series of diagonals of the ordered R^2 tables tend to be increasing with the exceptions previously noted when there appears to be a maximum followed by a decreasing effect. The ϕ^2 start out greater than the R^2. However, these two coefficients approach equality at the number of factors for the maximum R^2. This number of factors, as previously noted, is in the neighborhood of the last observed break in the corresponding series of roots. For the more critical cells in the experiment, seven factors in the major domain, middle and simulation model, wide and low ranges of b_{1i}^2, the ϕ^2 appear to drop off more rapidly after the number of factors for maximum diagonal R^2 than do the R^2. An implication of a much lower ϕ^2 is that the rotated factors for congruence with the individual input factors would be quite oblique. For an example, the seventh ϕ^2 is .00000 for the second battery, simulation model, seven factors in the major factor domain, low range of b_{1i}^2, and factoring of the RwSMC matrix. The seventh ordered diagonal for this battery is .67611. However, from a special analysis for this battery, it turns out that the transformation to dimensions matching the individual major domain input factors is so oblique as to be singular. In other cases when the ϕ^2 is not exactly zero but is very small, the rotated factors matching the individual input major domain factors are very correlated, approaching unity. When fewer factors are extracted and the rotation made to a subset of input factors for which the R^2 are moderately high and the ϕ^2 is larger than in the preceding case, the correlations between the factors are lower. (More extensive studies of the correlations between factors have been deferred to inclusion with studies on rotation of axes.)

5. *Discussion*

In evaluating the results obtained from the analyses performed, the major criterion will be that of degree of reproduction of the major domain factors from the obtained factor matrices. Procedures for development of correlation matrices were devised to simulate various aspects and properties in the plan-

https://doi.org/10.1007/BF02290601 Published online by Cambridge University Press

ning and execution of factor analytic experiments. Included in the planning and execution of factor analytic experiments was the design of the battery of variables and the care in construction of the measuring instruments and in the collection of data. Possible influences on the data were divided into three categories: 1) a major factor domain representing the area of phenomena of interest to the experimenter, 2) a minor factor domain representing other influences outside the area of interest, and 3) specific and measurement error influences on the results for each variable. Only for the formal model which includes only the first and third of the preceding influences, does the major domain factor matrix coincide with a common factor matrix. For the middle model and simulation model, minor factor domain influences are added which are common factors. Therefore, the major question of inquiry in the present investigation is not the ability to reproduce the common factors by various techniques but is the extent to which results of the analyses may be informative as to properties within the major factor domain.

The present investigation has not included the sampling of individuals problems. All correlation matrices have presumed population parameters for the battery of variables. The problems attacked are of a psychometric nature: that of design of measures, selection of measures, and collection of data.

Observations were made in the preceding section on a number of technical points related to procedures of analysis. These will be reviewed briefly.

The SMC's (squared multiple correlations of the variables with all other variables in the batteries) were less than and strongly related to the b_{1j}^2's only for the formal model. The b_{1j}^2's are parameters in the simulation procedure and are the proportion of the variances on the variables due to the major domain factors. The SMC's tended to increase from the formal model to the middle model to the simulation model accompanied by a reduction in the strength relation of the SMC's to the b_{1j}^2 .

Guttman's [1954] lower bounds for the number of common factors became less effective indicators of the number of factors in the major domain from the formal model to the middle model to the simulation model. The weaker lower bound of number of roots equal to or greater than unity for principal axes factors of the correlation matrices with unity in the diagonal cells proved to be superior to the stronger lower bound which consistently would lead to an over-estimation of the number of factors in the major domain for the middle and simulation models. Perceived breaks in the series of roots from principal axes factoring appeared to be superior in most cases in yielding estimates of the number of factors in the major domain.

Two methods for matching the output factors with the input major domain factors were employed: 1) maximizing congruence when both input and output factors were transformed, 2) maximizing congruence of transformed output factors with each untransformed input factor. Excellent matches were obtained for the formal model when as many factors were extracted as were contained in the major factor domain except for the cases for the wide range of b_{1j}^2 when the correlation matrix with unity in the diagonals was factored. In every case, and especially for the cases for the wide range of b_{1j}^2 , substitution of the squared multiple correlations into the diagonal cells of the correlation matrices improved the matching of the output factors with the input factors for the formal model. However, the substitution of the squared multiple correlations into the diagonal cells of the correlation matrices for the middle model and simulation model did not lead invariably to improved matches between output factors and input factors. When as many factors were extracted as the number of major domain input factors for the middle model and the simulation model some very poor matches were obtained between the output factors and the input factors. This was especially true for the low range of b_{1j}^2 and for seven factors in the major domain.

When factor extraction was terminated at the last perceived break in the series of roots, excellent to moderately acceptable matches were found between the output factors and the input factors. This was true even when fewer factors were extracted than the number of factors in the major domain in which case the match was between the output factor space and a subspace of the input factor space. A further observation was made that in this latter case when fewer factors were extracted, moderately adequate matches were possible with each factor in a subset of the input factors equal in number to the number of extracted factors. In these cases when factor extraction was terminated at the breaks in the series of roots, there was some improvement in results when the squared multiple correlations were substituted for the unities in the diagonal cells of the correlation matrices.

Factor extraction was continued to the Guttman stronger lower bound for the middle model and the simulation model for seven factors in the major domain, low range of b_{1j}^2 , and factoring of the matrices RwSMC. These were the cases for which the matches were poorest of the output factors with the input major domain factors when the number of factors extracted equalled the number of factors in the major domain. Improved matches were found for a subspace of the output factors when the larger number of factors were extracted. However, there may be considerable problem when the rotation of axes is not guided by knowledge of the input factors. Information imbedded in the output factor space may be inadequate for any rotational procedure using only this imbedded information to separate the subspace related to the major domain input factors from the subspace not so related. This is a problem for further investigation.

A point not brought out in the observations of the results is the possible relation between the strengths of the major domain factors and the number of usable extracted factors. As noted in the section on Procedure for Simulation of Correlation Matrices in connection with Tables 2, 4, and 5, the major factor input dimensions varied extensively in strength of representation as measured by the sums of squares of loadings on the factors. This variation was most pronounced for the batteries for seven factors in the major domain and the low range of b_{1j}^2 which corresponds with the batteries for which output results indicated fewer factors to be extracted according to the criterion of last perceived break in the series of roots and to the number of factors for which moderate matching was obtained of the output factors to dimensions of the major factor domain.

Major points in these discussions relate to the quality of factor analysis output to the major factors in the experimental design of 1) range of b_{1j}^2 , 2) size of major factor domain, and 3) class of relation between b_{2j}^2 and b_{3j}^2 . The first fact is related to the ability of the experimenter to construct and use variables with high dependence on the major domain; the second of these experimental design factors is related to the ratio of the number of variables in the battery employed to the dimensionality of the major domain; the third of these experimental design factors is related to the control by the experimenter in reducing the relative effects of minor factors as compared to the measurements error effects. This last experimental design factor has been expressed in terms of the three models employed: formal, middle, and simulation models. These three factors resulted in a $3 \times 2 \times 3$ experimental design. Ratings of the quality of factor analysis for each of the 18 cells of this design are given in Table 12. These ratings are subjective evaluations on a scale analogous to a grade scale from A, high to E, failure.

From Table 12 it appears that all three experimental design factors

Table 12

Ratings* of Results from Factoring of RwSMC
Classified by Size of Major Domain, Range of b_{1j}^2, Model

	Three Factors			Seven Factors		
	Model			Model		
Range of b_{1j}^2	Formal	Middle	Simulation	Formal	Middle	Simulation
High	A+	A	A	A+	A-	B+
Wide	A+	A	A-	A	B	C-
Low	A+	A-	B	A	C	D

* Ratings are subjectively given on grading scale
A - Excellent, B - Good, C - Marginal, D - Poor, E - Failing

are influencing the quality of the results. Poorer results are associated with seven factors in the major domain (higher ratio of number of variables to number of dimensions of the major domain), with the low range of b_{1i}^2 , and with the simulation model. While none of the results were rated as failing (E), the lowest rating given of D was for the cell for seven factors in the major domain, low range of b_{1i}^2 , and simulation model. Ratings of C, marginal, were given in two cells, both for seven factors in the major domain, one for the low range of b_{1i}^2 and middle model and the other for the wide range of b_{1i}^2 and the simulation model.

A major conclusion of the investigation is that the quality of factor analytic results depends greatly upon the quality of the design and conduct of the factor analytic experiment. While, for the formal model especially, communality corrections may be effective in compensating for larger errors of measurement, utilization of data with large influences of minor factors will lower the quality of the output of the factor analysis. Further, it appears that the ratio of number of variables to dimensionality of the major domain should be high. A general conclusion would be of the form that the major domain should be highly represented in the data. This is very important. The results very definitely demonstrate that poorly designed and conducted research cannot be salvaged by efforts expended in data analysis.

REFERENCES

Burt, C. Factor analysis and canonical correlations. *British Journal of Statistical Psychology*, 1949, **1**, 95–106.

Cattell, R. B. and Dickman, K. A dynamic model of physical influences demonstrating the necessity of oblique simple structure. *Psychological Bulletin*, 1962, **59**, 389–400.

Cattell, R. B. and Sullivan, W. The scientific nature of factors: a demonstration by cups of coffee. *Behavioral Sciences*, 1962, **7**, 184–193.

Cliff, N. and Hamburger, C. D. The study of sampling errors in factor analysis by means of artificial experiments. *Psychological Bulletin*, 1967, **68**, 430–445.

Guttman, L. Image theory for the structure of quantitative variates. *Psychometrika*, 1953, **18**, 277–296.

Guttman, L. Some necessary conditions for common-factor analysis. *Psychometrika*, 1954, **19**, 149–161.

Guttman, L. The determinacy of factor score matrices with implications for five other basic problems of common-factor theory. *British Journal of Statistical Psychology*, 1955, **8**, 65–81.

Guttman, L. To what extent can communalities reduce rank? *Psychometrika*, 1958, **23**, 297–308.

Harris, C. W. Some Rao-Guttman relationships. *Psychometrika*, 1962, **27**, 247–263.

Holzinger, K. J. and Harman, H. H. Comparison of two factorial analyses. *Psychometrika*, 1938, **3**, 45–60.

Hotelling, H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 1933, **24**, 417–441, 498–520.

Kaiser, H. F. Varimax solution for primary mental abilities. *Psychometrika*, 1960, **25**, 153–158.

https://doi.org/10.1007/BF02290601 Published online by Cambridge University Press

Kaiser, H. F. A note on Guttman's lower bound for the number of common factors. *British Journal of Statistical Psychology*, 1961, **14**, 1–2.

Kaiser, H. F. and Caffrey, J. Alpha factor analysis. *Psychometrika*, 1965, **30**, 1–14.

Kelley, T. L. Essential traits of mental life. *Harvard Studies in Education*, **26**, Cambridge, Mass.: Harvard Univ. Press, 1935.

Lawley, D. N. The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinborough*, 1940, **A40**, 64–82.

Linn, R. L. A Monte Carlo approach to the number of factors problem. *Psychometrika*, 1968, **33**, 37–71.

Loevinger, Jane. Person and population as psychometric concepts. *Psychological Review*, 1965, **72**, 143–155.

Odell, P. L. and Feiveson, A. H. A numerical procedure to generate a sample covariance matrix. *Journal of the American Statistics Association*, 1966, **61**, 199–203.

Rao, C. R. Estimation and tests of significance in factor analysis. *Psychometrika*, 1955, **20**, 93–111.

Spearman, C. Thurstone's work "reworked". *Journal of Educational Psychology*, 1939, **30**, 1–16.

Thurstone, L. L. Primary mental abilities. *Psychometric Monograph* No. **1**. Chicago: Univ. of Chicago Press, 1938.

Thurstone, L. L. Current issues in factor analysis. *Psychological Bulletin*, 1940, **37**, 189–236.

Tryon, R. C. Communality of a variable: formulation by cluster analysis. *Psychometrika*, 1957, **22**, 241–260. (b)

Tryon, R. C. Reliability and behavior domain validity: reformulation and historical critique. *Psychological Bulletin*, 1957, **54**, 229–249. (a)

Tryon, R. C. Cumulative communality cluster analysis. *Educ. Psychol. Measurement*, 1958, **18**, 3–35.

Tucker, L. R. A method for synthesis of factor analysis studies. *Personnel Research Section Report, No. 984*. Washington, D. C.: Department of the Army, 1951. (Mimeographed)

Wrigley, C. S. and Neuhaus, J. O. The matching of two sets of factors. *American Psychologist*, 1955, **10**, 418–419.

Zimmerman, W. S. A revised orthogonal rotational solution for Thurstone's original Primary Mental Abilities test battery. *Psychometrika*, 1953, **18**, 77–93.

Manuscript received 9/12/67
Revised manuscript received 10/10/68