PSYCHOMETRIKA—VOL. 24, NO. 2
JUNE, 1959

DOMAIN SAMPLING FORMULATION OF CLUSTER AND FACTOR ANALYSIS

ROBERT C. TRYON

UNIVERSITY OF CALIFORNIA, BERKELEY

Domain sampling principles permit formulation of a general method of multidimensional analysis. Cluster and factor analysis methods are special cases stemming from decisions made at different stages of the general method, especially in defining an independent dimension. Key cluster analyses define a dimension as a selection of *s* variables drawn from the full *n* set. Centroid, principal axes, and maximum likelihood analyses define it by the *n* variables (raw or residual, weighted or unweighted); bifactor and second-order analysis, by both types of selection; square root analysis, by one variable. Key cluster methods can be designed to test hypotheses.

The definition of a variable as the sum of a sample set of scored responses (e.g., to test items) selected to be representative of a defined domain of behavior is a basic principle of psychometrics. This standard practice may be expressed in a simple algebraic fashion which leads to an integration of the plethora of formulations of the reliability coefficient [39]. When a test is included among *n* variables, domain sampling algebra also provides a definitive solution of its communality [40]. These principles have been shown to implement the broad logic of multidimensional analysis by the psychometric procedures called cluster analysis [42]. The most generally applicable computational variant of cluster analysis, the CC method, has also recently been published [41].

Completing this group of papers on domain sampling formulations, this article has as its purpose, first, to state the general case of multidimensional analysis, and, second, to develop from it important special cases that are variant methods of cluster analysis. Some of these special forms have, however, been otherwise known over the last half century as factor analysis methods, their main originators being Spearman [29, 30], Thomson [31], Burt [2, 3], Kelley [19, 20], Hotelling [14, 15], Thurstone [32, 34], Holzinger [13], and Lawley [21]. The factor methods of Spearman, Kelley, Thurstone, and Holzinger are conceived as issuing from the basic factor theorem. The assumptions are that a test score results from underlying, uncorrelated and additive true (general and multiple), specific, and error factors. These restrictive assumptions of factor theory are difficult to justify on substantive biological and psychological grounds [36]. This paper shows that when the factor methods are recast as variants of cluster analysis such assumptions about the components of a test score are unnecessary restrictions.

TABLE 1

DECISIONS THAT DISTINGUISH VARIOUS METHODS OF MULTIDIMENSIONAL ANALYSIS

REGIONS OF DECISION	KEY CLUSTER				PIVOT VARIABLE	GEN'L AND KEY	GENERAL CLUSTER		
	1	2	3	4	5	6	7	8	9
	Resid Cluster		Precluster		Factor Analysis Variants				
	Tot Com (TC)	Cum Com (CC)	Preclus Cum Com (PCC)^a	Ration Cum Com (RCC)	Square Root (PV)	Bi-factor	Cent-roid (Unwgt)	Prin-cipal Axes (Wgt)	Maxi-mum Like-lihood
PRELIM DECISIONS									
a Reflection	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Factor					*No*	*No*	*Yes*	*No*	*No*
b Precluster	No	No	Empiric	Ration	Yes	No	No	No	No
Factor					*No*	*Yes*	*No*	*No*	*No*
DIAGONAL ENTRIES									
c Initial Communalities	Simul Quad	Approx	Approx^b	Approx^b	Approx	Approx	Approx	Approx	No
Factor					*Unities*	*No*	*High r*	*Unities*	*No*
DIMENSIONALITY									
d Defining Variables of C_x	s_x	s_x	s_x	s_x	3 on 1	n and s_x	n	n(wgt)	n(wgt)
Factor					*1*	*k and s_x*	*n*	*n(wgt)*	*n(wgt)*
e Partial Communalities	Sim Σ	Sim Σ	Sim Σ	Sim Σ	Sim Σ	Sim Σ	Sim Σ	Iter	Iter
Factor					*Sim Σ*^{c}	*Appr B*	*Sim Σ*	*Iter*	*Iter*
f Terminating Criteria^d	T	T	T	T	T	T	T	T	Resid r
Factor					*Salient*	*k + 1*	*Resid r*	*Resid r*	*Resid r*
g Reiterate Factoring	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Prog k
Factor					*No*	*No*	*No*	*No*	*Prog k*
STRUCTURE									
h Oblique Analysis	From d	From d	From b,d	From b,d	From d	From d	Rotate	Rotate	Rotate
Factor					*No*	*No*	*Rotate*	*No*	*No*
SCORING									
i Oblique Dimension Scores	Σ	Σ	Σ	Σ	Σ	Σ	Regress	Regr	Regr
Factor					*Σ*	*Regress*	*Regr*^{e}	*Regr*	*Regr*

[a] For factor analysis methods (e.g.,"multiple group"), see text

[b] In abridged form of the method, no initial communalities are necessary

[c] Specified for the special case of s_x=1 and communalities unity

[d] Sampling criteria may be used in place of or in addition to the T criterion

[e] Regression in theory; cluster score by Σ in practice

The broad plan of this treatment is represented schematically in Table 1. The general case of multidimensional analysis consists of progressive stages, or regions, in each of which decisions by the analyst are required. These regions are summarily described in the lettered rows *a* to *i*, extreme left column of Table 1. The numbered columns 1 to 9 are the special cases, each defined by the particular pattern of decisions listed down its column. The first four cases form the *Key Cluster* methods, columns 1 to 4. The remaining groups are other cluster analysis variants known also as factor analysis methods (or schools): their descriptive names are given below the column numbers. For example, column 7 defines the Thurstone centroid–simple structure method. In this group of methods, note that in each region of decision (or row) there are two entries. The first entry is a decision resting on the conception of the method as a variant of cluster analysis. The second, in italics and labelled at the left as "Factor," is the decision made in orthodox factor formulation. For example, down column 7 Thurstone's decisions are represented by the pattern of second, italicized, entries.

The general case of multidimensional analysis is given in the next section of the paper. Each region of analysis is taken up in order, the principles involved being illustrated by referring to some of the types of decisions listed along its row. In later sections the special cases will be taken up successively, that is, for each column of Table 1 the nature and rationale of its pattern of final decisions will be given, first, conceiving the method as a variant of cluster analysis (first cell entries), and, second, as an orthodox factor formulation (second cell entries).

General Case of Multidimensional Analysis

The basic data of the analysis are the intercorrelations between scores on variables $X_1, X_2, \cdots, X_v, \cdots, X_n$. The over-all objective is *to determine and measure the smallest number, k, of dimensions that reproduce the intercorrelations*, entered as side elements in a correlation matrix. The successive stages of the analysis, rows *a* to *i* of Table 1, achieve this objective. For convenience, these stages are grouped under five subordinate objectives, lettered A to E below.

A. *Preliminary Decisions* (Table 1, rows *a*, *b*)

Reflection of Variables (row *a*). A main desideratum in deciding on whether to reflect variables is the method of computing partial communalities (squared factor loadings), row *e*. In the special cases of columns 1 to 7, where the simple summation (Sim Σ) formula (to be developed later) is used, reflection of variables, denoted by "Yes" in row *a*, is required.

Preclustering Variables (row *b*). Preclustering variables before dimensional analysis permits certain abbreviated or rational variants of multidimensional analysis (columns 3, 4, 5). In the other variants test clusters are located during the dimensional analysis (1, 2, 6) or by rotation of dimensions (7, 8, 9).

B. *Determining Initial Communalities* (Table 1, row *c*)

The diagonal entries of the matrix are "self-correlations." Communalities are selected as diagonal elements because their use yields the smallest number, k, of necessary and sufficient dimensions, or uncorrelated cluster domain scores, that will reproduce the intercorrelations. If one used the reliability coefficient, $r_{tt'}$, of each variable as its diagonal element, more dimensions than k would be required, and if one set each diagonal value at unity, even more dimensions would be necessary.

As a correlation coefficient, the communality h_v^2 of a variable v is defined as the correlation between the observed variable X_v and a hypothetical construct variable $X_{v'}$ measuring a *different* behavior property than X_v but having correlations across the $n - 1$ other variables equal to those of X_v ([40], formula 1) i.e.,

$$h_v^2 = r_{vv'} \,. \tag{1}$$

As a variance, the fundamental definition of h_v^2 is the proportion of total variance of the observed scores of v that is predictable from the construct variable-domain score, C_v, defined as

$$C_v = z_v + z_{v'} + \cdots + z_{v\infty} \,, \tag{2}$$

in which the observed z_v-scores are defined as one sample variable drawn from an infinite set of construct variables, all members of which measure different behavior properties, but whose correlations with the $n - 1$ variables are proportional to those of v. From this construction it follows that ([40], formula 5)

$$h_v^2 = r_{vC_v}^2 \,. \tag{3}$$

Proportionality of the correlations of v with those of another variable i means specifically

$$r_{vj}/r_{ij} = \text{a constant} \qquad (j = 1, \cdots, n; v \neq i \neq j). \tag{4}$$

Taking the following as an index of proportionality ([41], formula 6; [43]),

$$P_{vi}^2 = \frac{(\Sigma r_{vj} r_{ij})^2}{\Sigma r_{vj}^2 \Sigma r_{ij}^2} \qquad (j = 1, \cdots, n; v \neq i \neq j), \tag{5}$$

then under condition (4), P_{vi}^2 is unity. The definition of the variable-domain C_v in (2) is not restricted to a domain of variables with equal correlations but merely to those with proportional r's, as defined in (4).

Communalities are in practice not computed by their defining formulas (1) and (3) because the requisite construct variables are not available. In

Table 1, row *c*, note that in TC analysis (to be described later) a solution is attempted by a quadratic equation. But in remaining methods approximations are taken; after reiteration of the factoring procedure, row *g*, final converged values are achieved.

C. *Determining Dimensionality* (*Factoring*) (Table 1, rows *d*, *e*, *f*, *g*)

The object is to determine the value of k, the number of uncorrelated composite variables or independent dimensions, $C_1 , C_2 , \cdots , C_x , \cdots , C_k$ that could reproduce the correlation matrix, including diagonal communalities.

(a) The communality, h_v^2 , of any variable v may be partitioned into k partial communalities (squared factor loadings) or h_{xv}^2 values as follows:

$$h_v^2 = h_{1v}^2 + \cdots + h_{xv}^2 + \cdots + h_{kv}^2 . \tag{6}$$

(b) The correlation coefficient of v with each other variable i is reproduced, i.e., the observed r_{vi} equals

$$r'_{vi} = h_{1v}h_{1i} + \cdots + h_{xv}h_{xi} + \cdots + h_{kv}h_{ki} , \tag{7}$$

or, said another way, each of its residual correlations after removing the variance from dimensions 1 to k is

$${}_{1\ldots k}r_{vi} = r_{vi} - r'_{vi} = 0. \tag{8}$$

Definition of an Independent Dimension, C_x (*An Orthogonal Factor*) (Table 1, row *d*). The score C_x is a composite, defined as the following independent cluster domain score:

$$C_x = (w_a)_r C_a + (w_b)_r C_b + \cdots + (w_{s_x})_r C_{s_x} . \tag{9}$$

The *defining variables* of the dimension are selected observed variables a, b, $\cdots$, s_x , taken from all n variables. The C's are their variable-domain scores, defined by (2). The prescript r means that scores on preceding dimensions $C_1 , \cdots , C_{x-1}$ are held constant, thus establishing the independence of C_x . The w's are weights. The same variable may appear in different dimensions, though, of course, as different residual scores.

This definition is more general than that written originally by Pearson [25], who initiated multidimensional analysis. It encompasses as special cases the varieties of cluster and factor analysis given in Table 1. In row *d*, note that in key cluster analysis (TC, CC, PCC, RCC), each dimension is defined by a cluster of s_x variables usually less than n. In centroid, principal axes, and maximum likelihood factor analysis, the defining variables are indiscriminately a general cluster of all n variables. These latter methods differ from each other in the values of the weights attached to the different component variables in (9). In bifactor analysis, dimension C_1 is a general cluster of all n variables, but later dimensions are key clusters. In pivot variable ("square root") analysis, each dimension is defined focally by one variable only.

Recall that each C-variable is a variable-domain defined in (2) as a composite of z-scores of variables measuring different behaviors but showing proportional correlations, hence the use of communalities in the diagonals. Were the z-scores in (2) defined as different test samples of the *same* behavior domain, each C-variable would be the construct composite called the "true" or X_{t_∞}-score of the variable [39] and reliability coefficients would be diagonal entries of the matrix. Were each C-variable defined as the single z-score of the defining variable, then the diagonals would be unities. In multidimensional analysis, the definition leading to communalities in the diagonal is chosen for the reason given under objective B.

The simplest weighting of the C-variables is to set all w's to unity in (9), as in TC, CC, PCC, RCC, centroid, diagonal, and bifactor analysis. This simple summation, as Burt calls it [3, 4], intrinsically weights each defining variable of the dimension C_x by its proportional contribution to the variance of C_x, that is, by the sum of its communality and its correlations with the other defining variables. Another choice, characteristic of the principal axes and maximum likelihood methods, is the computationally arduous least squares solution of the sets of weights which yield the maximal sum of their h_x^2 values, that is, of their partial communalities.

Partial Communalities (*Squared Factor Loadings*) (Table 1, row *e*). The portion of the variance of a variable v predictable from a dimension C_x is the square of its correlation with C_x, called its partial communality, $h_{x_v}^2$, defined in (6). In the unweighted case, from the correlation of sums in the limit,

$$(10) \qquad h_{x_v}^2 = r_{vC_x}^2 = \frac{(\Sigma_r r_{vi})^2}{\Sigma_r h_i^2 + 2\Sigma_r r_{ij}} \qquad (i, j = a, \cdots, s_x\, ; i < j).$$

The numerator is simply the square of the sum of the residual correlations of v with the defining variables of C_x. The denominator is simply the sum over the total submatrix of residual correlations of these defining variables, including the diagonal residual communalities. When $i = v$, then ${}_rh_i^2$ is included in the numerator.

A residual correlation, from (7) and (8), is

$$(11) \qquad {}_rr_{vi} = {}_{1\ldots(x-1)}r_{vi} = r_{vi} - (h_{1_v}h_{1_i} + \cdots + h_{(x-1)_v}h_{(x-1)_i}),$$

and similarly for the ${}_rr_{ij}$ terms. A residual communality is, from (6),

$$(12) \qquad {}_rh_i^2 = {}_{1\ldots(x-1)}h_i^2 = h_i^2 - (h_{1_i}^2 + \cdots + h_{(x-1)_i}^2).$$

If one has chosen the defining variables of C_x before the dimension analysis as in PCC and RCC analysis, it is *not* necessary to work out the individual residual terms of (11) and (12). Only sums from the raw matrix plus diagonals and sums of partial communalities on prior dimensions are necessary. Thus, from (11) the numerator of (10) is

$$\Sigma_r r_{vi} = \Sigma r_{vi} - (h_{1v} \Sigma h_{1i} + \cdots + h_{(x-1)v} \Sigma h_{(x-1)i}) \qquad (i = a, \cdots, s_x). \tag{13}$$

When $i = v$, h_v^2 is included in Σr_{vi}. The denominator of (10),

$$\Sigma_r h_i^2 + 2\Sigma_r r_{ij}^2 = \Sigma r_{ij} - [(\Sigma h_{1i})^2 + \cdots + (\Sigma h_{(x-1)i})^2] \qquad (i, j = a, \cdots, s_x). \tag{14}$$

Σr_{ij} is the sum over the raw matrix of s_x variables including diagonals.

This simple general formula (10), called "Sim Σ" in Table 1, row *e*, is used by all cluster and factor analysis methods, excepting principal axes and maximum likelihood (to be considered later). Recall that for different methods, it differs only in the value of s_x, e.g., in centroid analysis, $s_x = n$.

Terminating Criteria (*Salient Dimension Analysis*) (Table 1, row *f*). As a simple rational standard for terminating factoring, the writer proposes the *communality exhaustion criterion.* To end factoring by this criterion, one estimates the communalities of all the variables at the beginning of the analysis. Factoring then proceeds up to the dimension C_k at which the communalities of the n variables are exhausted.

Recall that the communality h_v^2 is basically defined in (3), quite independent of the dimension analysis, as the variance of variable v predictable from its variable-domain C_v. This magnitude is estimated before factoring is undertaken by computing formulas given later in the paper, (31) or (32). After factoring is under way, the variance of v predictable from dimensions C_1, C_2, $\cdots$, C_x is represented by h_{vx}^2, this magnitude being the sum of partial communalities of v up to and including h_{xv}^2, as shown in (6).

Writing the ratio of these two variances, i.e.,

$$F_{xv} = \frac{h_{vx}^2}{h_v^2}, \tag{15}$$

factoring may be terminated on that dimension C_k at which the numerator of (15) approaches the denominator, i.e., when

$$F_{kv} = \frac{h_{vk}^2}{h_v^2} = 1.000. \tag{16}$$

To evaluate F for each variable on each dimension as factoring proceeds would, however, be a complex procedure. Furthermore, the magnitude of F for a given variable is subject to substantial error, both in the initial approximation to the denominator term and in the initial dimensional estimate of the numerator. Less subject to these errors is the approximate sum (or average) of the F-values over all n variables, namely,

$$T_x = \frac{\Sigma h_{vx}^2}{\Sigma h_v^2}. \tag{17}$$

Using T_x as a criterion, one stops factoring on the dimension C_k for which T_k by (17) first equals or exceeds unity, i.e., where

$$T_k = \frac{\Sigma h_{vk}^2}{\Sigma h_v^2} \geq 1.000. \tag{18}$$

In practice, the T-criterion in (18) appears to yield the most *salient* dimensions [41]. Consider bias in T: initial estimates of communalities are usually biased downward, suggesting that a salient terminal dimension might be rejected by T. This is unlikely because such a dimension would be the one for which T first *exceeds* 1.000. The effect of an upward bias would be to accept nonsalient dimensions. To minimize such an effect, the analyst may set the criterion a little under unity, say, at .975.

The T-criterion may reject later dimensions that would be accepted on sampling grounds. Many analysts may consider such rejection to be an unimportant loss, for such dimensions contribute minor general variance and are usually difficult to interpret.

Significant Dimension Analysis. On the basis of *sampling criteria,* one may wish, however, to accept all significant (as distinguished from salient) dimensions. The orthodox F-test procedures applied to the communality exhaustion indices represented by (15) and (17) would seem appropriate, but their sampling characteristics are not yet known. There remain the various significance tests applied by factorists to the distribution of residual correlations (11) or to the distribution of the square roots of the partial communalities (i.e., of the factor loadings) of a given dimension (10). The tests developed by Saunders are of special interest because he proposes both types ([6], formulas 44 and 46, p. 300ff).

Reiteration of the Factoring to Converged Communalities (Table 1, row g). After the first factoring is complete, the sum of partial communalities by (6) may not yield the correct value of the communality of each variable, as in artificial or population matrices [40]. In those methods that start with approximations, reiteration of the dimensionality analysis on the k dimensions is required until convergence is secured, as shown in row g, Table 1.

The decision as to which decimal place will define convergence may be made on arbitrary grounds of salience, say, the third place. On sampling grounds, however, one may terminate convergence when for every variable the difference between two successive iterated values of its communality by (6) becomes less than, say, a third of the standard error of the last iterated value of its communality. Treating h_v as a multiple correlation, as in (38), later, the approximate error is

$$\sigma_{h_v} \doteq (1 - h_v^2)/\sqrt{N - (k + 1)} \doteq (1 - h_v^2)/\sqrt{N - 2}. \tag{19}$$

The magnitude of k, at least 1, is usually trivial relative to N, hence leading to the final approximation shown.

D. *Determining the Structure of the Interrelationships* (Table 1, row h)

Having determined the dimensionality of the intercorrelations, one may relax the condition of independence and select or derive the k dimensions that may be oblique to each other and be better defined by the observed variables. Key cluster analysis routinely locates those groups of variables which delineate the k most nearly independent oblique dimensions. Corresponding to the independent dimensions C_1 , C_2 , $\cdots$, C_k there are the matched set of oblique dimensions, respectively, C_I , C_{II} , $\cdots$, C_K . Thus for independent dimension C_x given in (9) there is an oblique dimension C_y defined in simple summation form by the domain score

$$C_y = C_a + C_b + \cdots + C_{s_x} . \tag{20}$$

Scores on the C-variables that form this composite have the same definition as in (9), but they are *not* residual scores as in (9). Geometrically, dimension C_y is an oblique subcentroid in k-space.

If the analyst wishes to check on the clusterings indicated by the dimensional analysis with an eye, perhaps, to a possible reclustering of the variables including those that had remained unclustered, he may employ a geometric model (for an illustration, see [41], Fig. 1). This model takes the k dimensions as independent axes, and each variable as a point on them. The coordinate of each variable on any axis C_x is its correlation, r_{vC_x} (unaugmented factor loading), which by (10) is

$$r_{vC_x} = h_{x_v} . \tag{21}$$

The resulting interior model of variable-points is perceptually not as descriptive of the interrelationships among them as the surface model, given by plotting each variable-domain C_v by its augmented correlation, this being, from the correlation of sums in the limit,

$$r_{C_vC_x} = h_{x_v}/h_v . \tag{22}$$

The surface model, in which all variable-domains are points at distance 1.00 from the origin, has the perceptual merit of revealing directly as a surface separation of points the relationship between each variable-domain C_v and any one of the other variable-domains C_i . This important "common factor correlation" can, however, be computed directly from the matrix and the communalities, i.e., from the correlation of sums in the limit,

$$r_{C_vC_i} = r_{vi}/h_vh_i . \tag{23}$$

Simple Cluster Structure (*Rotated Primary Factors*). The finally selected cluster domains of type (20) are the most nearly independent k dimensions evident in the data. The degree of their interdependence is given by their intercorrelations. For any two such generally oblique cluster domains, C_{y_i} , C_{y_j} , their correlation is given, from the correlation of sums in the limit, as

$$(24) \qquad r_{C_{y_i}C_{y_j}} = \frac{\Sigma r_{ij}}{\sqrt{\Sigma r_{ii}}\,\sqrt{\Sigma r_{jj}}} ,$$

where Σr_{ii} is the sum of r's in the submatrix of the s_i variables of C_{y_i} including diagonal communalities, Σr_{jj} is the same for the submatrix of s_j variables defining C_{y_j} , and Σr_{ij} is the sum of the $s_i s_j$ coefficients in their cross correlation submatrix. Recall that, as stated under (9), a given variable may appear in more than one oblique dimension, a situation which would, of course, increase obliqueness.

As an aid in interpreting an oblique dimension C_{y_i} , one may compute the correlation of each known observed sample variable v with the dimension (rotated factor loading). By the correlation of sums in the limit it is

$$(25) \qquad r_{vC_{y_i}} = \frac{\Sigma r_{vi}}{\sqrt{\Sigma r_{ii}}} .$$

But of more interest theoretically is the correlation of C_{y_i} with each *kind* of general variation of which v is taken as a test sample, namely, its variable-domain C_v in (2). This correlation (augmented rotated factor loading) is simply, from the correlation of sums in the limit,

$$(26) \qquad r_{C_vC_{y_i}} = r_{vC_{y_i}}/h_v .$$

Difficult problems arise in simple cluster structure analysis in those methods in which the defining variables of the independent C_x dimensions are total clusters of all n variables. As shown in Table 1, row h, columns 7, 8, 9, these dimensions must be rotated [see 9] to meaningful defining orthogonal or oblique clusters. Orthodox factor analysts following Thurstone [34] propose graphical rotation—a cumbersome, subjective procedure, admittedly an art [6]. Recent attempts have been made to achieve rotation by objective analytic methods [5, 18, 24, 26, 28]. Rotation, an unnecessary burden, is not required when the dimensions are defined by key clusters.

E. *Scores on Oblique Dimensions* (Table 1, row i)

Ideally, the best estimate of an individual's sample score C_{ys} on any dimension C_y is the regression of C_y on the n variables, i.e.,

$$(27) \qquad \bar{C}_{ys} = \Sigma \beta_{C_y i} z_i \qquad (i = 1, \cdots, n).$$

Such an estimate is so arduous to compute that in most factor analyses the important job of measuring persons on the reduced dimensions is rarely tackled, and a main benefit of the analysis is lost.

When, however, a dimension is defined by a key cluster, a good estimate may be secured from the cluster score, namely, the simple sum of the z scores of the defining variables of C_v , i.e.,

$$C_{vs} = z_a + \cdots + z_s \, . \qquad (28)$$

In this composite each defining variable takes an intrinsic weight proportional to the sum (or mean) of its correlations with the remaining defining variables. In Table 1, row i, the simple summation, labelled "Σ" may be used in all but the general cluster methods (columns 7, 8, 9).

The *cluster domain validity* of the observed cluster score (28) is its correlation with the full domain score, C_v in (20). By the sums formula in the limit it is

$$r_{C_{vs}C_v} = \sqrt{\frac{\Sigma h_i^2 + \Sigma r_{ij}}{s + \Sigma r_{ij}}} \qquad (i, j = a, \cdots, s; i \neq j). \qquad (29)$$

Note that the numerator in (29) is simply the sum over the submatrix of C_v including diagonal communalities, and the denominator is the same except with unities in the diagonals.

The relationships between cluster scores that fallibly measure the final k oblique dimensions are given by their intercorrelations. Between any two such scores, C_{vis} , C_{vjs} , by the sums formula this correlation is

$$r_{C_{vis}C_{vjs}} = \text{same as (24) but with the unities in the diagonals.} \qquad (30)$$

Special Cases of Multidimensional Analysis

Key Cluster Analysis: Total Communality (TC) and Cummulative Communality (CC) (Table 1, columns 1, 2)

The TC and CC methods directly apply the general formulations outlined above. As shown in Table 1, columns 1 and 2, the correlation matrix is initially made positive (row a) by conventional reflection methods (see [10], Table 16.13) in order to guarantee that variables chosen to define a given dimension will show correlations of positive proportionality in (4).

An electronic computer is required in TC analysis to solve for the communalities by a simultaneous quadratic formula (row c). Difficulty has, however, been experienced in achieving a solution in empirical matrices [16, 17]. The CC method has therefore been developed [41] to meet the possible failure of solutions by the quadratic formula. CC analysis starts with approximations to communalities, and dimensionality analysis is reiterated until convergence is secured. CC analysis is thus a procedure alternative to TC, being a method that provides a solution in all matrices, and one that may be programmed either for electronic or desk calculator computation.

Solution of the communalities in TC analysis is based on the fact that the communality of any variable v is the squared multiple correlation between v and the remaining $n - 1$ variable-domains ([40], formula 11),

$$h_v^2 = R^2_{v.c_1c_2\ldots c_i\ldots c_n} \qquad (i \neq v). \tag{31}$$

Solving for the communalities of all variables requires a simultaneous solution by reiteration of the n quadratics of type (31).

Various initial approximations to communalities required in the CC method are available [40, 45, 46]. On domain sampling grounds, the preferred estimate of h_v^2 is "Approximation B" ([40], formulas 29, 30), the Spearman formula, computed from a cluster of reference variables whose correlations are most nearly proportional to those of v. Here,

$$h_v^2 = \Sigma r_{vi} r_{vj} / \Sigma r_{ij} \qquad (i, j \neq v; i < j), \tag{32}$$

where both i and j are the three variables showing the highest P^2_{vi} and P^2_{vj} values, respectively, by (5).

In both methods the defining variables of each dimension C_x anchor on a selected pivot variable, v_1 , that appears to show relatively high *and* low correlations with other variables. Such a variable would usually center the cluster obliquely to clusters defining other dimensions. To locate the pivot variable a measure of "pivotness" of each of the n variables is first computed, namely, the variance of its squared r's. The pivot variable v_1 then is that variable whose

$$\text{var}\ ({}_r r^2_{v_1 i}) \quad \text{is the maximum} \quad (i = 1, \cdots, n; i \neq v). \tag{33}$$

A quicker and probably less sensitive means of selecting the pivot variable is to choose the one with the highest residual communality as given in (12). For desk calculator work, selecting v_1 from a correlation distribution table is satisfactory, though it entails some subjective elements ([41], Table 2).

Around the pivot variable one collects the remaining defining variables of C_x . These are the variables with highest indices of proportionality with v_1 . Three such variables at a minimum are selected. If these variables are called, in order of magnitude of $P^2_{v_1 i}$, i_1 , i_2 , i_3 then any additional variable i may also be included in the cluster if its $P^2_{v_1 i}$ value is equal to or above .81 and also is within twice $P^2_{v_1 i_1} - P^2_{v_1 i_3}$, that is, if its

$$P^2_{v_1 i} \geqq (2P^2_{v_1 i_3} - P^2_{v_1 i_1}) \geqq .81. \tag{34}$$

Partial communalities, row e, are computed by the general formula (10). In TC analysis the factoring process is terminated by the T-criterion at the end of one factoring procedure only. But in CC analysis, since approximations initiate the analysis, the first factoring process is terminated by the T-criterion, then new values of the communalities are computed from (6), and the factoring process is reiterated until the communalities converge (rows f, g). In both TC and CC analysis, the dimensional analysis locates the

k oblique dimensions, row h, and scores of individuals on the dimensions, row i, follow the formulations as given earlier under the general method.

Key Cluster Analysis: Preclustered Cummulative Communality (PCC) and Rational Cummulative Communality (RCC) (Table 1, columns 3, 4)

Recall that in TC and CC analyses the cluster of variables selected to define a dimension C_x is chosen *during* the dimension analysis. One may, however, choose them *prior* to factoring, empirically in PCC, rationally in RCC analysis. Preselection of clusters makes multidimensional analysis a quick desk calculator operation because the complete residual correlation matrices essential to CC analysis are not required. PCC and RCC analyses are procedurally identical after the analyst has clustered the variables (see Table 1, columns 3 and 4).

In PCC analysis (for a recent illustration see [38]), one *empirically* groups the n variables into k' clusters, $C_I , \cdots , C_y , \cdots , C_{K'}$. Each cluster is made to be as "tight" as possible, i.e., is composed of variables whose correlations are maximally proportional by (5). Some variables may remain unclustered, but their number is kept as small as possible. As an aid in selecting the groupings one may use a correlation distribution table.

But in RCC analysis the rational groupings stem from the analyst's *theory* from which he generated the n variables under study. An investigator commonly conceives the n variables to sample different behavior domains or properties of the individuals, such as the facets of Guttman [11]. The s variables that fall in each such theoretical subgroup are a rational cluster. The n variables will usually be organized in k' such clusters, though a few may remain as isolates.

As in CC analysis, one starts both PCC and RCC analyses by computing approximations to the communalities, preferably by formula (32). Thereafter the work is procedurally identical to the CC analysis excepting that only mean residuals are necessary. The mean residual correlation of a variable v with any cluster C_y is, from (13),

$$(35) \qquad {}_r\bar{r}_{vi} = (1/s_y)\Sigma\, {}_r r_{vi} \qquad (i = a, \cdots , s_y).$$

The mean residual communality of the variables that compose C_y is, from (12),

$$(36) \qquad \overline{{}_r h_i^2} = (1/s_y)\Sigma\, {}_r h_i^2 \ .$$

The quickest means of choosing the *pivotal defining cluster* of any dimension C_x is to select the one with largest mean residual communalities as given in (36). The partial communalities are then computed by the simple summation formula (10). As shown in Table 1, the remainder of PCC and RCC analyses is the same as in CC. The final k dimensions are thus defined by a selection from the k' original clusters.

A complication may arise if one should run out of clusters before dimensionality has been completely determined. If one wishes precision on dimensionality he would compute the $n \times n$ residual correlation matrix and would perform the CC procedures on it and on any later residual matrices that are necessary. PCC and RCC analyses can be made as precise as CC analysis. To do so the analyst forms new estimates of the communalities by (6) and then, as in CC analysis, reiterates the factoring procedures until communalities converge.

Abridged PCC and RCC (*Multiple Group Factor Analysis or "Poor Man's Cluster Analysis"*). The analyst need only spend a few hours of work on a correlation matrix if he is satisfied with an approximate multidimensional analysis. After reflecting the variables, preclustering them, and approximating their communalities as in PCC analysis proper he can compute the correlations between the resulting k' cluster domains by (24). He may, if he wishes, even *skip* the step of approximating communalities, leave the diagonal vacant, and use mean r's instead of Σr's in (24).

The result is a $k' \times k'$ matrix with diagonal elements of unity. The correlation between each variable v and a given domain C_v is then estimated by (25); if communalities are not used, he would use mean r's instead of Σr values in (25). These calculations result in an $n \times k'$ matrix. From a study of these data he may make a reasonable estimate of the dimensionality, interpret the oblique dimensions, and compute cluster scores. If he wishes a more accurate estimate of the dimensionality and of the most oblique k clusters, he can factor the $k' \times k'$ matrix by the diagonal method of factoring (see "Pivot variable analysis" below). By this means he quickly locates the k most nearly independent cluster dimensions.

Orthodox PCC Analysis (*Group, Grouping, and Multiple Group Methods of Factoring*). In 1939, the writer published dimensional analysis by the PCC method in approximately the form presented here ([37], Sec. 7); the quick abridged form was also given (Sec. 5, and Analyses 14, 15a for centroids). Five years later Holzinger [12] and then Thurstone [33, 35] presented the abridged form. Thurstone labels it the multiple group method of factoring. At the end of abridged analysis he rotates the k oblique dimensions to orthogonal positions in order to compute residual correlations, and to see if further dimensions might be necessary. In his book, Thurstone [34] added the group and the grouping methods (see also [6], ch. 11). They differ from the 1939 and current PCC methods in the procedure of grouping variables in clusters on grounds of absolute *magnitudes* of correlations rather than of proportionality of correlations, our P^2 criterion.

Pivot Variable (*PV*) *Cluster Analysis* (*Diagonal or Square Root Factor Analysis*) (Table 1, column 5)

In PV analysis (see Table 1, column 5), each variable i among the n

variables may be selected as the central sample variable of an oblique cluster, C_{y_i}. Defined generally in (20), C_{y_i} here consists of three variables only. The other two variables are those which, after the matrix is reflexed, yield the highest P^2 values with i. On domain sampling principles one may thus conceptualize $k' = n$ preclustered domains, $C_{y_I}, \cdots, C_{y_i}, \cdots, C_{y_n}$. After determining approximations to the communality of each variable from its reference variables by (32), one computes the correlations between the n domains by (24). Then a dimensional analysis of the $n \times n$ matrix of $r_{C_{y_i}C_{y_j}}$ coefficients, with diagonal elements of unity, is performed. The cluster with highest column sum defines the first dimension; locating the pivot cluster of later dimensions requires only computing residual communalities and selecting the one with highest value. With an electronic computer, however, one can instead compute residual matrices and more sensitively select the pivot variable by (33). If the defining pivot cluster of any dimension C_x is denoted C_{y_p}, then the augmented partial communality of an oblique cluster is a special case of (10), i.e.,

$$(37) \qquad h^2_{xC_{y_i}} = {}_r r^2_{C_{y_i}C_{y_p}}/(1 - {}_r h^2_{C_{y_p}}).$$

The numerator of (37) calls only for the residual interdomain correlations of the selected pivot cluster with each of the remaining $n - 1$ clusters; only a simple $n \times 1$ matrix of such residual correlations is necessary. Factoring is terminated when the sum of the augmented partial communalities of all variables over all dimensions, that is, the numerator of the T-criterion, first becomes equal to or greater than the denominator term, n.

Having now located k pivot clusters by the dimensional analysis, the analyst assigns each of the remaining $n - k$ variables to a final set of k oblique clusters. Each may be assigned to that pivot cluster which defines the dimension on which the variable in question has its highest partial communality. Tighter clusters may, however, be grouped by criterion (34). An illustration of PV diagonal factoring procedures applied to an interdomain matrix is given elsewhere ([38], Appendix D). The potentialities of PV analysis should be explored. It is a rapid means of estimating k, the dimensionality (rank) of the matrix, and hence it could precede maximum likelihood analysis where foreknowledge of the approximate value of k is desirable.

PV analysis would give precise results if one started with correct values of the communalities. But it may be employed to find such values, as follows. After the first factoring to determine the k most nearly independent pivotal clusters, these clusters may be used as a constant reference set of predictors to compute the communality of each variable, v. Elsewhere it has been shown ([40], formula 44) that h^2_v is the squared multiple correlation between v and a set of k oblique cluster domains, i.e.,

$$(38) \qquad h^2_v = R^2_{v.C_I C_{II} \ldots C_K}.$$

From knowledge of the k oblique clusters by initial PV analysis, and with initial trial values of the constants required in (38) computed from (32), (24), and (25), simultaneous solutions of all n communalities, reiterated to convergence, can be programmed electronically. Such a program may be integrated with a periodic replication of PV analysis in order to discover whether the increased precision of the communalities produces changes in dimensionality.

Orthodox PV Analysis. One of the oldest factoring methods, diagonal factoring, used in PV analysis has recently been relabelled square root factor analysis [44]. In orthodox practice the analyst pivots a dimension on a *single* variable by arbitrarily inserting unities in the diagonal and factoring the unreflected r matrix. This practice prevents one from determining the dimensionality of the matrix, rests the partial communalities rather unstably on the coefficients of the pivot variable alone, and leaves unclear the oblique cluster structure of the variables. The method may, however, be useful in studies with large n as a preliminary analysis to locate the most promising predictors of a criterion.

General and Key Cluster Analysis (Bifactor and Second-order Factor Analysis) (Table 1, column 6)

Historically, the urge to discover one general dimension in cognitive behaviors provided the impetus to the ultimate development of multidimensional analysis. It produced the two-factor theory of Spearman [29, 30], and its subsequent generalization by Holzinger and Harman [13] to bifactor analysis.

Applying domain sampling principles to this case, one defines the first dimension as a general domain score on a full battery of all n variables; thereafter each dimension is a residual score on an empirically discerned key cluster. In Table 1, column 6, notice that in bifactor analysis the decisions follow the same pattern as in CC analysis, except for the one particular of the definition of the dimensions, row d. Here, the deviation is *only* with respect to the first dimension C_1 which is defined in general formula (9) as the sum of scores on all n variable domains, i.e., $s_1 = n$. In the first residual correlation matrix and thereafter the regular CC procedure is carried through, each subsequent dimension being defined by s_x variables.

Orthodox Bifactor Analysis. As illustrated by Holzinger and Harman [13], the orthodox procedure is applied to predominately positive matrices that may not require reflection. The variables are preclustered into k' clusters, each consisting of variables which (in our terms) show high P^2 values with each other. One of these clusters is a general cluster, consisting of one variable drawn from each of the remaining $k' - 1$ clusters. The first dimension, C_1, is defined by this general cluster and is called a general factor, g, the first partial communalities of all n variables on it being called their squared

g-saturations. In the first residual matrix and thereafter, the dimensions are successively defined in turn as each of the residual $k' - 1$ clusters (group factors). The partial communalities are computed by approximation (32) but for the defining variables of each key cluster only. Zero partial communalities are by *fiat* assigned to variables in other clusters. This procedure is unwieldy computationally. Results from it will correspond closely to those more efficiently achieved by the CC method, modified as in Table 1, column 6, by defining the first dimension as a general cluster domain, the remaining dimensions as key clusters; preclustering is unnecessary.

Orthodox Second-order Factor Analysis. In Thurstone centroid analysis, when a generally positive matrix of correlations between primary factors results from rotation, the correlations between these first-order factors may then be subjected to a new centroid analysis on one dimension only. This dimension is termed a second-order factor. On domain sampling principles the k correlated primaries represent hypothetical oblique clusters which, unlike the oblique clusters discovered in key cluster analysis, are normally poorly defined by actual variables. The second-order factor is simply a composite general cluster domain—a battery score on the k oblique clusters. This general composite is difficult to interpret because of its vague omnibus character and because of the complex redundancy of some variables that are common to two or more primaries.

A cleaner general composite, if desired, would be secured by a CC analysis designed as described above, namely, by defining the first dimension C_1 as a composite domain of all n variables without redundancy, later dimensions as key clusters. To illustrate that such a first dimension corresponds closely to the Thurstone second-order dimension (sometimes called g), the writer compared the correlations between the 11 WAIS variables and the Thurstone second-order factor ([7], Table 5, 18–19 yr. olds) with their correlations with the C_1 dimension defined as an nonredundant general cluster domain. The P^2-value of the paired columns of correlations was unity in the second decimal place.

Second-order analysis can be extended, of course, to more than a single dimension and to matrices with positive and negative correlations between the first-order factors. This problem is more fruitfully approached, however, under higher-order composites of oblique clusters discussed in the last section of this paper (see "Designed reanalyses").

General Cluster Analysis (*Centroid–Simple Structure, Principal Axes, Maximum Likelihood Factor Analysis*) (Table 1, columns 7, 8, 9)

In the remaining group of methods each dimension C_x is defined successively as general battery residual scores on all n variable domains, unweighted or weighted.

Unweighted General Clusters (*Centroid–Simple Structure Factor Analysis*)

(Table 1, column 7). If the analyst wishes to define each dimension C_x as the total unweighted residual domain score on the n variables, he sets $s_x = n$ in (9) and all weights equal. The procedures of CC analysis are now called for down to structure analysis, as shown in Table 1, column 7. As pointed out earlier, the indiscriminate use of all n variables requires graphical or analytic rotational methods to describe the simple cluster structure. Following Thurstone, the simple-structure factorist who uses graphical methods usually does not place the oblique rotated dimensions through oblique clusters of variables but rather the rotated dimensions bound the variables. As a result, the calculation of individual's scores on these oblique dimensions or factors requires the use of the complex regression equation given in (27).

This complexity usually leads the simple-structure factorist, after having laboriously located the "underlying" primaries through rotational devices, to abandon efforts to estimate the scores of individuals on them. Thurstone recommends ([34], p. 515) that the complex regression score be replaced by computing a simple cluster score on the nearest oblique cluster—precisely the type of cluster domain that is directly located and measured by key cluster analysis.

Orthodox unweighted general cluster analysis has been fully developed by Thurstone [32, 34] and his followers under the name "multiple factor analysis." In Table 1, column 7, if one compares the orthodox procedure (2nd cell entries) originally formulated on the basic factor theorem with the procedures based on domain sampling (1st cell entries), they are seen to be identical except for certain unrefined features of the orthodox method: the use of highest r's as initial estimates of communalities, the terminating of factoring by a statistical test of residual r's, and the lack of reiteration of factoring to converged communalities.

Weighted General Clusters (*Principal Axes or Components*) (Table 1, column 8). One may wish to attach differential weights to scores on the n variable domains in (9) in order to maximize the sum of partial communalities on each successive dimension. The definition of the dimensions in this case is otherwise identical to that in the preceding unweighted case. Procedurally (see Table 1, column 8), no initial reflection of variables is required, but an electronic computer is essential to solve reiteratively by least squares for the values of the partial communalities, row e.

In the orthodox method of weighted general clusters as described by Pearson [25], Hotelling [14, 15], and Kelley [20], considerable confusion has been introduced by their use of unities (and sometimes reliabilities) in the diagonals. A rationale for unity diagonals would be equally applicable to any of the other methods described above and equally inappropriate, for communalities are called for, as stated earlier in this paper (see also [13], ch. 7). The only basic feature that distinguishes the weighting of general clusters from the preceding centroid analysis is the decision to apply weights in the definition of each dimension. A main feature that distinguishes both approaches from the key cluster method is the decision to set s_x equal to n. As a consequence both general cluster methods require rotational procedures in order to describe cluster structure.

Maximum Likelihood General Clusters (Table 1, column 9). As with the preceding two methods, Lawley's maximum likelihood procedures [21, 22, 23, 27] determine k general cluster dimensions, $s_x = n$ in (9). But Lawley's dimensions are significant in the sense that the final partial and total communalities reiteratively determined from trial values are model population values that would produce a correlation matrix from which there is maximum likelihood that the observed statistical matrix represents sampling fluctuations. A chi square test of residual correlations is taken as evidence that the model cannot be rejected. The procedures are onerous even for an electronic calculator, since models with k progressively increased must be tested one at a time. Efficiency is greatly improved if good trial values of k and of the partial communalities can be found to initiate the reiterative procedures. This maximum likelihood approach has special appeal because of its capabilities of significance testing, but its gargantuan computation requirements and the need of rotation to a final cluster structure are serious limitations.

To sum up the three general cluster approaches and offer a prospectus, all three general methods define dimensions as indiscriminate composite domain scores on all n variables. This definition is a common limitation because it leads to the uncertain procedures of rotation to simple structure. Paradoxically, the Thurstonian unweighted case, the most generally used method of factor analysis [6, 8, 10, 34], is least justifiable. A few books, notably British ones, do put the method in the right perspective [1, 3, 31]. Centroid factoring is not as exact as the principal axes method in determining dimensionality. It lacks a test of significance, unlike the maximum likelihood method. For a general cluster solution the analyst would now use a modern computer programmed for the principal axes or maximum likelihood methods. As for the computational simplicity of the centroid method, the key cluster methods are simpler and have the additional merit of routinely describing oblique structure without the need of rotational devices.

Designing principal axes and maximum likelihood solutions so as to describe oblique cluster structure would remove their present inadequacies. Coupling key cluster analysis to them would, it would seem, turn the trick. A preliminary pivot variable (PV) analysis would quickly demarcate the desired set of k oblique clusters. If a method of least squares fit of k axes through these oblique subcentroids can be devised, the result would be a weighted key cluster or principal cluster solution. To increase the efficiency of a maximum likelihood method a preliminary PV analysis should, it would seem, provide good initial trial values, not only of dimensionality but also of partial communalities (appropriately unaugmented). If this method can be further modified to pass hypothetical population dimensions through the k key clusters, rotation would not be required, but it would retain the significance testing features of Lawley. The result would be a maximum likelihood key cluster solution.

Rationally Designed Dimensional Analysis

The order in which clusters are selected to define successive dimensions is determined in the key cluster methods of Table 1, columns 1 to 4, completely objectively; the object is to maximize the independence of the variables that define different dimensions and to select them in decreasing order of salience. But key cluster analysis need not be so blindly empirical; it may be designed to test hypotheses based on theories about the structure and order of variance determination among the n variables.

Designed TC and CC Analysis. An analyst may generate the hypothesis, for example, that clusters among n_1 variables of type A (say, a group of sociological variables) may better predict the communality variances of all clusters among n_2 variables of type B (say, a group of social attitude variables), than would be the case in the converse direction, B to A. To test the hypothesis, he would perform a TC or CC analysis of the $n_1 + n_2$ matrix of correlations, but restrict the defining variables of the factored dimensions solely to the A block of variables. After the terminating criterion T has been met by the A-variables, the residual communalities of the B-variables would be their communality variances unpredictable by the A set. A fresh TC or CC analysis in reverse order, B to A, would reveal residual communality variances of the A-variables unpredictable by the B set. Under the hypothesis, percentage determination of the variances in the A to B design should be higher than in B to A.

Designed PCC and RCC Analysis. Having empirically or rationally preclustered n variables, for example, of the sociological and attitude types, an analyst may on the basis of theory generate the hypothesis that only certain ones of the k' clusters are salient predictors of the remaining clusters, and in a hypothetical order of salience. His test would consist of a dimensional analysis in which the successively factored dimensions would be defined by the different selected clusters arranged in the order of their hypothetical decreasing salience. After the T-criterion is met, then under the hypothesis the analyst would discover that, for the variables not in the selected clusters, their residual communalities should in general have progressively decreased as factoring proceeded and should have become negligible on dimensions following those hypothesized as being salient.

Designed Reanalyses. Any of the above designs may follow upon a preliminary purely empirical key cluster analysis. Such a preliminary study, including the oblique structure analysis, may lead to new ideas that the analyst may wish to test by additional types of dimensional analyses.

For example, will a small number of *higher-order*, or composite, clusters maximally predict the communality variances of the *n* variables? To illustrate, a study of the oblique structure of sociological and social attitude clusters may suggest that one composite of sociological clusters and one composite of attitude clusters may leave but a minor amount of the communalities of the variables unpredictable. The test in this case would be a new dimensional analysis in which the first two dimensions would be defined respectively by the two composite clusters. The residual communalities of the variables would constitute the parts unpredictable from these two dimensions. Thurstone's single second-order factor analysis is a special design, discussed earlier, in which *one* such general composite cluster may be so tested.

Another type of reanalysis consists of bringing together in one common new dimensional analysis clusters found to be the most nearly independent sets in different prior analyses. To illustrate, using urban neighborhoods as objects, the writer performed three separate CC analyses: on sociological variables in 1940, on the same characteristics in 1950, and on voting variables in 1954. Each of the separate analyses yielded three salient oblique dimensions. The nine oblique clusters were then projected into a single common dimensional analysis. This master analysis still yielded three salient oblique dimensions, demonstrating thereby the common tridimensionality of these characteristics over more than a decade. In such reanalyses the analyst may also wish to include brand new variables that he considers theoretically to belong to the common structure.

Summary

The general method of multidimensional analysis, designed on domain sampling principles, covers as special cases all the main types of cluster and factor analysis. The different methods vary primarily in special decisions about the nature of an independent dimension. Such a dimension is defined in general as a composite score on a cluster of variable-domains. In cluster analysis terms the special methods of key cluster analysis, denoted as TC, CC, PCC, and RCC, define each dimension as a selected set or subgroup from all the *n* variables. CC analysis is the most generally applicable. Square root or diagonal factor analysis, called PV analysis, pivots each dimension on one central variable. Bifactor and second-order factor analysis defines the first dimension as a general cluster of all *n* variables, the later dimensions as key clusters. Among the general cluster methods that define each dimension as a composite of all *n* variables, centroid factor analysis defines it as an unweighted total cluster domain, principal axes as a weighted one, and maximum likelihood factor analysis also a weighted one but having the additional feature of supplying a technique for estimating the number of statistically significant dimensions required. The key cluster methods de-

termine simple cluster structure as a routine aspect of the factoring process, whereas the general cluster methods require laborious rotations to determine structure. The key cluster methods can be applied blind, but they can also be designed to test hypotheses.

REFERENCES

[1] Adcock, C. J. *Factorial analysis for non-mathematicians.* Carlton: Melbourne Univ. Press, 1954.

[2] Burt, C. *The distribution and relations of èducational abilities.* London: King, 1917.

[3] Burt, C. *The factors of the mind.* New York: Macmillan, 1941.

[4] Burt, C. Alternative methods of factor analysis and their relations to Pearson's method of principal axes. *Brit. J. Psychol., Statist. Sec.,* 1949, **2**, 98–121.

[5] Carroll, J. B. An analytic solution for approximating simple structure in factor analysis. *Psychometrika,* 1953, **18**, 23–38.

[6] Cattell, R. B. *Factor analysis.* New York: Harper, 1952.

[7] Cohen, J. The factorial structure of WAIS between early adulthood and old age. *J. consult. Psychol.,* 1957, **21**, 283–290.

[8] Fruchter, B. *Introduction to factor analysis.* New York: Van Nostrand, 1954.

[9] Garnett, J. C. M. On certain independent factors of mental measurements. *Proc. Roy. Soc.,* 1919, A, **96**, 91–111.

[10] Guilford, J. T. *Psychometric methods.* (2nd ed.) New York: McGraw-Hill, 1954.

[11] Guttman, L. A new approach to factor analysis: the radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences.* New York: Columbia Univ. Press, 1956.

[12] Holzinger, K. J. A simple method of factor analysis. *Psychometrika,* 1944, **9**, 257–262.

[13] Holzinger, K. J. and Harman, H. *Factor analysis.* Chicago: Univ. Chicago Press, 1941.

[14] Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. educ. Psychol.,* 1933, **24**, 417–441, 498–520.

[15] Hotelling, H. Simplified calculation of principal components. *Psychometrika,* 1935, **1**, 27–35.

[16] Kaiser, H. F. Solution for the communalities: a preliminary report. Rep. No. 5, AF 41(657)-76, Univ. Calif., Berkeley, Sept., 1956.

[17] Kaiser, H. F. Further numerical investigation of the Tryon-Kaiser solution for the communalities. Rep. No. 14, AF 41(657)-76, Univ. Calif., Berkeley, May, 1957.

[18] Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika,* 1958, **23**, 187–200.

[19] Kelley, T. L. *Crossroads in the mind of man.* Stanford: Stanford Univ. Press, 1928.

[20] Kelley, T. L. *Essential traits of mental life.* Cambridge: Harvard Univ. Press, 1935.

[21] Lawley, D. N. The estimation of factor loadings by the method of maximum likelihood. *Proc. Roy. Soc., Edinburgh,* 1940, **60**, 64–82.

[22] Lawley, D. N. The maximum likelihood method of estimating factor loadings. Ch. 21 in G. Thomson, *The factorial analysis of human ability.* (5th ed.) London: Univ. London Press, 1951.

[23] Lord, F. M. A study of speed factors in tests and academic grades. *Psychometrika,* 1956, **21**, 31–50.

[24] Neuhaus, J. O. and Wrigley, C. The quartimax method: an analytic approach to orthogonal simple structure. *Brit. J. statist. Psychol.,* 1954, **7**, 81–91.

[25] Pearson, K. On lines and planes of closest fit to systems of points in space, *Phil. Mag.,* 1901, 6th Ser., 559ff.

[26] Pinzka, C. and Saunders, D. R. Analytic rotation to simple structure. II. Extension to an oblique solution. Princeton, N. J.: Educ. Test. Serv. Res. Bull., Aug., 1954.

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

[28] Saunders, D. R. An analytic method for rotation to orthogonal simple structure. Princeton, N. J.: Educ. Test. Serv. Res. Bull., Aug., 1953.

[29] Spearman, C. General intelligence objectively determined and measured. *Amer. J. Psychol.*, 1904, **15**, 201–293.

[30] Spearman, C. *The abilities of man.* London: Macmillan, 1927.

[31] Thomson, G. *The factorial analysis of human ability.* (5th ed.) London: Univ. London Press, 1951.

[32] Thurstone, L. L. *The vectors of mind.* Chicago: Univ. Chicago Press, 1935.

[33] Thurstone, L. L. A multiple group of factoring the correlation matrix. *Psychometrika,* 1945, **10**, 73–78.

[34] Thurstone, L. L. *Multiple-factor analysis.* Chicago: Univ. Chicago Press, 1947.

[35] Thurstone, L. L. Note about the multiple group method. *Psychometrika,* 1949, **14**, 43–45.

[36] Tryon, R. C. A theory of psychological components—an alternative to "mathematical factors." *Psychol. Rev.*, 1935, **42**, 425–454.

[37] Tryon, R. C. *Cluster analysis.* Ann Arbor, Mich.: Edwards, 1939.

[38] Tryon, R. C. Identification of social areas from cluster analysis. *Univ. Calif. Publ. Psychol.*, 1955, **8**, No. 1, 1–100. Berkeley: Univ. Calif. Press.

[39] Tryon, R. C. Reliability and behavior domain-validity: reformulation and historical critique. *Psychol. Bull.*, 1957, **54**, 229–249.

[40] Tryon, R. C. Communality of a variable: formulation from cluster analysis. *Psychometrika,* 1957, **22**, 241–259.

[41] Tryon, R. C. Cumulative communality cluster analysis. *Educ. psychol. Measmt,* 1958, **18**, 3–35.

[42] Tryon, R. C. General dimensions of individual differences: cluster analysis vs. multiple factor analysis. *Educ. psychol. Measmt,* 1958, **18**, 477–495.

[43] Wrigley, C. F. and Neuhaus, J. O. The matching of two sets of factors. *Amer. Psychologist,* 1955, **10**, 418–419. (Abstract)

[44] Wrigley, C. F., Cherry, C. N., Lee, M. C., and McQuitty, L. L. Use of the square root method to identify factors in the job performance of aircraft mechanics. *Psychol. Monogr.*, 1956, **71**, No. 1 (Whole No. 430).

[45] Wrigley, C. The effect upon the communalities of changing the estimate of the number of factors. Rep. No. 13, AF 41(657)-76, Univ. Calif., Berkeley, March, 1957.

[46] Wrigley, C. F. An empirical comparison of various methods for estimating communalities. *Educ. psychol. Measmt,* in press.

Manuscript received 5/7/58

Revised manuscript received 9/10/58