THE

INTERNATIONAL ENCYCLOPEDIA OF EDUCATION

Research and Studies

Volume 4
F–H

Editors-in-Chief
TORSTEN HUSEN

T. NEVILLE POSTLETHWAITE

THE INTERNATIONAL ENCYCLOPEDIA OF EDUCATION

Research and Studies

Volume 4
F–H

Editors-in-Chief

TORSTEN HUSEN
University of Stockholm, Sweden

T. NEVILLE POSTLETHWAITE
University of Hamburg, FRG

PERGAMON PRESS

OXFORD · NEW YORK · TORONTO · SYDNEY · PARIS · FRANKFURT

U.K.	Pergamon Press Ltd., Headington Hill Hall, Oxford OX3 0BW, England
U.S.A.	Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford, New York 10523, U.S.A.
CANADA	Pergamon Press Canada Ltd., Suite 104, 150 Consumers Rd., Willowdale, Ontario M2J 1P9, Canada
AUSTRALIA	Pergamon Press (Aust.) Pty. Ltd., P.O. Box 544, Potts Point, N.S.W. 2011, Australia
FRANCE	Pergamon Press SARL, 24 rue des Ecoles, 75240 Paris, Cedex 05, France
FEDERAL REPUBLIC OF GERMANY	Pergamon Press GmbH, Hammerweg 6, D-6242 Kronberg-Taunus, Federal Republic of Germany

First edition 1985

Library of Congress Cataloging in Publication Data
Main entry under title:
The International encyclopedia of education

Includes bibliographies.
Index: v. 10.
1. Education—Dictionaries. 2. Education—Research—Dictionaries. I. Husén, Torsten, 1916–
II. Postlethwaite, T. Neville.
LB15.I569 1985 370'.3'21 84–20750

British Library Cataloguing in Publication Data
The International encyclopedia of education

1. Education—Dictionaries

I. Husén, Torsten II. Postlethwaite, T. Neville
370'.3'21 LB15

ISBN 0–08–028119–2

Computer data file designed and computer typeset by Page Bros (Norwich) Ltd.
Printed in Great Britain by A. Wheaton & Co. Ltd., Exeter

CONTENTS

F

Factor Analysis

Factor analysis is a technique for representing the relationships among a set of variables in terms of a smaller number of underlying hypothetical variables. It aims to describe the variation among a set of measures in terms of more basic explanatory constructs, and thus to provide a simpler and more easily grasped framework for understanding the network of relationships among those measures. Correlations might be computed, for example, among the scores of a group of students on measures of addition, subtraction, multiplication, division, vocabulary, and reading comprehension. A factor analysis of these correlations might show that the relationships among the tests could be almost completely explained in terms of two underlying variables, which might well be interpreted as computational ability and verbal ability.

1. Early Development of Factor Analysis

Although the technique of factor analysis is now applied in a wide variety of disciplines, it originated in the field of psychology. Towards the end of the nineteenth century a number of psychologists turned their attention to experimental studies of intelligence and intellectual abilities. Spearman collected data to test his theory that mental activity could be explained in terms of a single central intellective function, "intelligence". Finding high correlations between estimates of intelligence and students' scores on tests of weight, light, and pitch discrimination, he concluded that

> all branches of intellectual activity have in common one fundamental function (or group of functions), whereas the remaining or specific elements of the activity seem in every case to be wholly different from that in all others. (Spearman 1904)

Subsequently, in his two-factor theory, the fundamental function was described as a general factor, "g", and the element specific to a particular activity as its specific factor, "s".

Spearman had noted that his matrices of correlations among intellectual abilities could be arranged hierarchically, showing a progressive decrease in value from left to right and from the upper to the lower rows of the table. He recognized that this would be the expected pattern of correlations if all mental processes reflected the operation of a single central intellective function, which operated at different levels of complexity. To test whether a set of correlations he had obtained among six variables conformed to this pattern, for instance, he computed the tetrad differences among the correlations, for example $(r_{13}r_{26} - r_{23}r_{16})$. Finding that they were approximately zero, he confirmed the hypothesis that the correlations could be explained by one general factor.

The two-factor theory was challenged by Thomson and other psychologists on both theoretical and empirical grounds. Working with larger batteries of tests and larger numbers of cases, Burt identified verbal, numerical, and practical group factors in school subjects in addition to a general factor; a group factor is one which is represented only in certain similar types of tests but not in others. Spearman later admitted the necessity of group factors, and British factorists adopted a factor model which incorporated both a general factor and group factors.

Hierarchical theories of mental structure had little appeal for American psychologists. They preferred a multiple-factor approach in which several factors were extracted directly from a correlation matrix, without any initial assumption about the need for a general factor. In the early 1930s, Kelley and Hotelling sought a unique and exact mathematical solution to the problem of identifying the underlying factors in a correlation matrix, and developed the general method of principal components analysis put forward earlier by Karl Pearson. This method extracts successive uncorrelated components which account for as much of the variation among the scores of students on a set of variables as is possible at each stage.

Thurstone, the major American contributor to the development of factor analysis, noted that the addition of further tests to a battery could affect the factors identified by the principal components approach. He sought a method of analysis which would lead to the discovery of psychologically meaningful factors which were invariant, that is, supporting the same interpretation, over different test batteries.

In 1931, Thurstone accelerated the development of factor analysis by noting that Spearman's tetrad difference of $r_{13}r_{24} - r_{23}r_{14} = 0$ was the equivalent of setting a second order minor or determinant to be equal to zero. In algebraic form,

$$\begin{vmatrix} r_{13} & r_{14} \\ r_{23} & r_{24} \end{vmatrix} = 0$$

He reasoned that "if the second-order minors must vanish in order to establish a single common factor, then must the third-order minors vanish in order to establish two common factors, and so on" (Thurstone

1947). This allowed him to use matrix algebra procedures to express the problem of determining the number of factors needed to account for an observed correlation matrix. He formulated the problem in terms of the fundamental factor theorem $\mathbf{FF}' = \mathbf{R}$, where **R** was the original correlation matrix and **F** was the factor matrix to be identified. **F** would consist of a matrix of coefficients or "loadings" of the original tests or variables on the "factors", and would usually be a rectangular matrix of lower rank than **R**. To avoid the then prohibitive calculations of the principal components solution to this equation, Thurstone developed the centroid method of analysis, which although quite tedious, was widely used until the 1950s, when advances in computer technology made other methods feasible.

Thurstone was also responsible for distinguishing two separate phases in the determination of factors—factor extraction and factor rotation. He recognized that the initial extraction of factors by the centroid method or by variants of the principal components method merely provided an arbitrary orthogonal set of reference axes—a set of axes at right angles to each other in two-dimensional, three-dimensional, or higher dimensional space depending upon the number of factors extracted—to represent the correlations among the tests or the relationships among the test vectors, and that any particular set of axes was only one of a very large number which would represent the correlations equally well. He claimed that the factor loadings determined at the factor extraction stage had no psychological meaning until they were rotated in the common factor space. Starting from the psychological assumption that there are some mental functions not involved in every intellectual task, Thurstone developed the criterion of simple structure to locate new positions for the reference axes. This required that the axes be placed so that each test would have significant loadings on only one or two factors and near-zero loadings on the remaining factors, and so that on each factor, a majority of the tests would have near-zero loadings. Unlike the British factorists, he made no initial assumption about the need for a general factor, but sought to determine "how many factors are indicated by the correlations without restriction as to whether they are general or group factors" (Thurstone 1947). It was left to the configuration of the test vectors to determine whether a general factor was needed in addition to other factors to explain the correlations among the tests.

Factor schools differed on the question of acceptable types of rotation. Most of the British factorists and a few of the American factorists insisted on orthogonal rotations; while a given axis could be rotated through any angle, the angle between that axis and other axes should remain at 90°. The factors therefore represented unrelated constructs. Thurstone, however, claimed that the restriction of unrelatedness or orthogonality of factors should not be imposed on the data. Application of the simple structure criterion would reveal whether the data could be represented by an orthogonal axis system. In most cases, however, the simple structure solution would require an oblique rotation of the initial axes, in which the angles between the rotated axes could be smaller or larger than a right angle. The factors emerging from an oblique rotation therefore tended to be themselves correlated. If the factors were correlated, the correlations among the factors could be further analysed to yield second-order or higher order factors, which to the extent that they were represented in all tests in a battery, could be regarded as analogous to a general factor.

The most significant developments during this early period of factor analysis were Spearman's conceptualization of the two-factor theory, its subsequent extension by British psychologists to a general plus group factor model, and a number of crucial contributions from L. L. Thurstone—his generalization of the two-factor notion to a multiple factor analysis model, his recognition of the need to rotate initially-extracted factors to arrive at scientifically interpretable results, and his development of the concept of oblique factors and of criteria for identifying factors. While the basic techniques of factor analysis were well-established by the 1950s, many problems remained. The initial extraction of factors still involved approximate methods, as did the estimation of test communalities, that is, that part of the variance of a test which it has in common with other tests in a battery. Criteria for determining the number of factors needed to explain the correlations were still approximate, and there was a substantial element of subjectivity in the graphical rotational procedures employed by factor analysts. Over the ensuing years, many of these problems have been resolved or considerably refined, with theoretical advances being greatly facilitated by advances in computer technology.

2. *The Basic Factor Model*

The basic factor model assumes that a score on a variable can be expressed as a linear combination or as a weighted sum of scores on factors underlying performance in that variable. If three hypothetical factors F_1, F_2, F_3 were assumed to underlie performance in test j, scores (expressed in standardized form, that is with a mean of zero and a standard deviation of 1) on test j could be represented by the equation

$$z_j = a_{j1}F_1 + a_{j2}F_2 + a_{j3}F_3 + U_j \qquad (1)$$

where the a coefficients represent the loadings of test j on the respective common factors; F_1, F_2, and F_3 represent standard scores on these factors; and U_j represents scores on a factor unique to test j, including error of measurement. The standard scores of two persons on test j, for instance, might be expressed as follows:

$$\text{Person 1: } z_{j1} = a_{j1}F_{11} + a_{j2}F_{21} + a_{j3}F_{31} + U_{j1} \quad (2)$$

$$\text{Person 2: } z_{j2} = a_{j1}F_{12} + a_{j2}F_{22} + a_{j3}F_{32} + U_{j2} \quad (3)$$

Thus the loadings of test j on any one factor are the same for all persons, but the scores on a factor, whether common or unique, differ among persons.

Continuing with the above example, the standard score of Person 1 on test k would be given by

$$z_{k1} = a_{k1}F_{11} + a_{k2}F_{21} + a_{k3}F_{31} + U_{k1} \quad (4)$$

The product of the z scores of Person 1 on tests j and k, that is, $z_{j1}\, z_{k1}$ can be found by multiplying the expressions on the right-hand side of Eqns. (2) and (4). Summing the product of the standard scores on tests j and k over all N persons in the sample, and dividing the result by N gives

$$\frac{1}{N}\left(\sum_{i=1}^{N} z_{ji}z_{ki}\right) = a_{j1}a_{k1} + a_{j2}a_{k2} + a_{j3}a_{k3} \quad (5)$$

since the scores on the three factors are standard scores and the sum of the squares of standard scores is equal to N, and since product terms involving scores on different factors, whether common or unique, are equal to zero, as the factors are by definition uncorrelated.

The expression on the left-hand side of Eqn. (5) defines the correlation between tests j and k, so that

$$r_{jk} = a_{j1}a_{k1} + a_{j2}a_{k2} + a_{j3}a_{k3} \quad (6)$$

That is, the correlation between any pair of variables can be expressed as the sum of the product of the loadings of those variables on each of the common factors. Using the vector terminology of matrix algebra, Eqn. (6) can be written as

$$r_{jk} = [a_{j1}\ a_{j2}\ a_{j3}] \begin{bmatrix} a_{k1} \\ a_{k2} \\ a_{k3} \end{bmatrix} \quad (7)$$

Generalizing Eqn. (7) to represent the intercorrelations among n variables in terms of the three factors gives

$$\begin{array}{c} \text{Test} \\ 1 \\ 2 \\ \cdot \\ j \\ k \\ \cdot \\ n \end{array} \overset{\begin{array}{ccccccc} \text{Test } 1 & 2 & . & j & k & . & n \end{array}}{\begin{bmatrix} r_{11}^* & r_{12} & . & r_{1j} & r_{1k} & . & r_{1n} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ r_{j1} & r_{j2} & . & r_{jj}^* & r_{jk} & . & r_{jn} \\ r_{k1} & r_{k2} & . & r_{kj} & r_{kk}^* & . & r_{kn} \\ . & . & . & . & . & . & . \\ r_{n1} & r_{n2} & . & r_{nj} & r_{nk} & . & r_{nn}^* \end{bmatrix}} =$$

$$\begin{array}{c} \text{Test} \\ 1 \\ 2 \\ \cdot \\ j \\ k \\ \cdot \\ n \end{array} \overset{\begin{array}{ccc} F_1 & F_2 & F_3 \end{array}}{\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ . & . & . \\ . & . & . \\ a_{j1} & a_{j2} & a_{j3} \\ a_{k1} & a_{k2} & a_{k3} \\ . & . & . \\ a_{n1} & a_{n2} & a_{n3} \end{bmatrix}} \begin{bmatrix} a_{11} & . & . & a_{j1} & a_{k1} & . & a_{n1} \\ a_{12} & . & . & a_{j2} & a_{k2} & . & a_{n2} \\ a_{13} & . & . & a_{j3} & a_{k3} & . & a_{n3} \end{bmatrix} \quad (8)$$

which is conveniently represented by the matrix equation

$$\mathbf{R}_c = \mathbf{F}\,\mathbf{F}' \quad (9)$$

where $\mathbf{R}_c$ is the matrix of correlations among the tests (which differs from the data-generated matrix $\mathbf{R}$ in that the asterisked diagonal entries consist of the correlation shared by the respective test with other tests in the battery and is less than unity), $\mathbf{F}$ is the matrix of test loadings on the factors, and $\mathbf{F}'$ is the transpose of the latter matrix. In Eqn. (8), r_{jk} of Eqn. (7) appears as the product of the jth row of the $\mathbf{F}$ matrix and the kth column of the $\mathbf{F}'$ matrix. Equation (9) indicates that a given $\mathbf{F}$ matrix would yield a unique $\mathbf{R}_c$ matrix, but that a given $\mathbf{R}_c$ matrix could be analysed to yield many different factor matrices.

Equation (9) represents the common factor model. The complete factor model also incorporates the variance (ψ_j) unique to each test, thus:

$$\mathbf{R} = \mathbf{R}_c + \psi = \mathbf{FF}' + \psi \quad (10)$$

where $\mathbf{R}$ is the correlation matrix with unities in the diagonal cells, and ψ is the diagonal matrix

$$\begin{bmatrix} \psi_1 & 0 & . & 0 & . & 0 \\ 0 & \psi_2 & . & 0 & . & 0 \\ 0 & 0 & . & \psi_j & . & 0 \\ 0 & 0 & . & 0 & . & \psi_n \end{bmatrix}$$

Each of the unique test variances (ψ_j) is regarded as consisting of a reliable component (specific variance, s_j^2) and an unreliable component (error variance, e_j^2). The common factor variance or communality for each test is represented by the symbol h_j^2. Thus in the factor model the variance of a test is expressed as the sum of several components:

$$\begin{aligned} \sigma_j^2 = 1 &= (a_{j1}^2 + a_{j2}^2 + \ldots + a_{jm}^2) + (s_j^2 + e_j^2) \quad (11) \\ &= \qquad\qquad h_j^2 \qquad\qquad + \quad \psi_j \end{aligned}$$

The reliability coefficient (r_{jj}) of a test is the sum of the reliable components of variance, ($h_j^2 + s_j^2$) or ($1 - e_j^2$).

It was assumed in the derivation of Eqn. (5) that the factors in the $\mathbf{F}$ matrix were uncorrelated, and this assumption is also implicit in Eqn. (10). Regarding this assumption as unnecessarily restrictive, Thurstone advocated the acceptance of oblique or correlated factors if warranted by the configuration of the test vectors. When Eqn. (10) is expanded to accommodate correlated factors, the basic factor equation becomes

$$\mathbf{R} = \mathbf{R}_c + \psi = \mathbf{F}\phi\mathbf{F}' + \psi \tag{12}$$

where ϕ represents the matrix of correlations among the factors.

In the present example,

$$\phi = \begin{bmatrix} 1 & r_{F_1F_2} & r_{F_1F_3} \\ r_{F_2F_1} & 1 & r_{F_2F_3} \\ r_{F_3F_1} & r_{F_3F_2} & 1 \end{bmatrix} \tag{13}$$

If the data can be satisfactorily explained by a set of uncorrelated factors, then ϕ reduces to an identity matrix,

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and Eqn. (12) reduces to Eqn. (10).

3. *Exploratory vs. Confirmatory Factor Analysis*

Usually, the first objective in carrying out a factor analysis of a correlation or covariance matrix is to arrive at an **F** matrix of the following form:

$$\begin{array}{l} \\ \text{Variables} \\ \text{Test 1} \\ \text{Test 2} \\ \vdots \\ \text{Test } j \\ \vdots \\ \text{Test } n \end{array} \begin{array}{c} \text{Factors} \\ \begin{array}{cccccccc} I & II & III & : & p & : & m \end{array} \\ \begin{bmatrix} a_{1I} & a_{1II} & a_{1III} & : & a_{1p} & : & a_{1m} \\ a_{2I} & a_{2II} & a_{2III} & : & a_{2p} & : & a_{2m} \\ \vdots & \vdots & \vdots & : & \vdots & : & \vdots \\ a_{jI} & a_{jII} & {}_{jIII} & : & a_{jp} & : & a_{jm} \\ \vdots & \vdots & \vdots & : & \vdots & : & \vdots \\ a_{nI} & a_{nII} & a_{nIII} & : & a_{np} & : & a_{nm} \end{bmatrix} \end{array} \tag{14}$$

This is the matrix of the loadings (a_{jp}) of a set of tests or other variables on a set of m underlying common factors, $m < n$. It is also referred to as a factor structure matrix, representing the correlations of each of the tests with each of the factors. As pointed out earlier, it is only one of a large number of matrices which would satisfy the relationship expressed in Eqn. (10), and some rotation of the axes represented by the factors would be required to arrive at a meaningful representation of the original data.

Both in its early development and in the large majority of its present-day applications, factor analysis has been used in an exploratory manner, to explore the underlying dimensions of a set of data. While there has been some indulgence in blind exploration among the uninitiated, in the sense of seeing what factors emerge from any ill-assorted set of variables, the use of factor analysis to explore the dimensions of an educational or psychological or sociological domain of interest has mostly been in the context of well-designed studies in which hypotheses have been carefully formulated and variables have been carefully selected. Exploratory factor analysis, however, does not place specific restrictions on the number of factors which should appear in the **F** matrix or the subsequent rotated matrix, or on whether particular entries in the factor matrices or factor correlation matrices should be zero or nonzero; it is an unrestricted factor model.

The idea of testing the hypothesis that the relationships among a set of variables might be accounted for in terms of a restricted factor model emerged in the mid-1950s, and following the work of such authors as Howe, Anderson and Rubin, Lawley, Jöreskog and Gruvaeus, had led to the development of procedures for confirmatory factor analysis. In contrast with exploratory factor analysis, confirmatory factor analysis sets out to test whether the original correlation or covariance matrix can be represented by an underlying factor matrix with a specific number of factors and/or specified zero or nonzero entries in factor matrices and/or factor correlation matrices. Instead of extracting an initial arbitrary **F** matrix and subsequently rotating that matrix, confirmatory factor analysis tests the specific hypothesis that the correlation or covariance matrix can be explained by an **F** matrix of a specified form, for example by a matrix involving exactly three factors with a specified pattern of loadings as in (A) and of factor correlations as in (B) below:

	(A)				(B)		
	I	II	III		I	II	III
Test 1	x	x	0	I	1	x	x
Test 2	x	x	0	II	x	1	0
Test 3	x	x	0	III	x	0	1
Test 4	x	x	0				
Test 5	x	0	x				
Test 6	x	0	x				
Test 7	0	0	x				
Test 8	0	0	x				

Maximum likelihood methods are used to estimate the nonzero (x) elements in these matrices, given the original correlations or covariances among the variables. If a goodness of fit test then shows that the observed matrices do not deviate significantly ($p < 0.05$) from the hypothesized factor solutions, the specific theoretical hypothesis is confirmed.

4. *Initial Extraction of Factors*

Many approaches to the determination of the initial **F** matrix have been developed since Thurstone proposed his centroid method of analysis, and many earlier methods have been superseded as a result of the development of computers. One set of reference axes and its associated **F** matrix have the important property that they enable a set of correlated variables to be described in terms of a set of orthogonal (uncorrelated) axes which account for the maximum amount of variance remaining among the variables as each axis in the new set is determined. The axes in this set are called the principal components of the original correlation matrix.

Equation (15) presents a correlation matrix for three tests: Vocabulary (V), Comprehension (C), Arithmetic problems (A).

$$\mathbf{R} = \begin{array}{c} \\ (V) \\ (C) \\ (A) \end{array} \begin{array}{c} \begin{array}{ccc} (V) & (C) & (A) \end{array} \\ \begin{bmatrix} 1.0 & 0.6 & 0.2 \\ 0.6 & 1.0 & 0.4 \\ 0.2 & 0.4 & 1.0 \end{bmatrix} \end{array} \tag{15}$$

This matrix can be represented by an ellipsoid of points in three-dimensional space, defined by three orthogonal axes X, Y, and Z. The ellipsoid would take the shape of an elongated football oriented from one corner of a room at floor level (the origin of the three-dimensional space) upwards towards the ceiling and outwards to the opposite walls. The first principal axis of the correlation matrix would be the major axis of the football; the second principal axis would pass through the centroid of the set of points and would be perpendicular to the first principal axis; the third principal axis would be perpendicular to both the first and second principal axes, representing the length of the line across the football if it had been flattened in one of its shorter dimensions. These three axes are called the principal components of the correlation matrix. The variances of the principal components are the latent roots or eigenvalues of **R** which are determined by solving the characteristic equation

$$|\mathbf{R} - \lambda \mathbf{I}| = 0 \tag{16}$$

These eigenvalues show the variance of the points along the first, second, and third principal axes of the football to be 1.823, 0.817, and 0.360 respectively.

The orientation of the principal axes with respect to the original axes is given by a set of eigenvectors corresponding to each eigenvalue; these are the direction cosines of each principal axis. By multiplying the elements of the eigenvectors by the square root of the corresponding eigenvalues, the loadings of the tests on the new axes would be found to be

$$\begin{array}{c} \\ V \\ C \\ A \end{array} \begin{array}{c} \begin{array}{ccc} \text{1st principal component} & \text{2nd principal component} & \text{3rd principal component} \end{array} \\ \begin{bmatrix} 0.800 & -0.475 & 0.366 \\ 0.888 & -0.110 & -0.446 \\ 0.627 & 0.762 & 0.164 \end{bmatrix} \end{array} \tag{17}$$

Principal component analysis describes the relationships among the original n variables in terms of n new uncorrelated factors, rather than in terms of a reduced number of factors. Principal axes can be found, however, for the matrix $\mathbf{R}_c$ [see Eqn. (9)], in which the correlations in the diagonal cells represent the variance which each variable has in common with other variables in the set, not including the unique variance. This application of the principal axes method is referred to as principal factor analysis.

The principal factor method will be illustrated with the aid of the fictitious matrix in Eqn. (18), which is based on the correlations among the scores of 200 15-year-old secondary-school students on examinations in English, French, Italian, physics, and chemistry, but in which the diagonal values of unity have been replaced by the communality, h_j^2, [see Eqn. (11)] of each variable. The communality is that part of the variance of each variable which it holds in common with one or more other variables in the set, or that part of the variable's self-correlation attributable to common factor variance in the set of variables. The squared multiple correlation of each variable with all of the other variables in the set is now usually accepted as the communality estimate, and has replaced the original values of unity in the diagonal cells of the matrix in Eqn. (18), which is therefore designated as $\mathbf{R}_c$.

$$\mathbf{R}_c = \begin{array}{c} \\ (1) \\ (2) \\ (3) \\ (4) \\ (5) \end{array} \begin{array}{c} \begin{array}{ccccc} \text{(1) English} & \text{(2) French} & \text{(3) Italian} & \text{(4) Physics} & \text{(5) Chemistry} \end{array} \\ \begin{bmatrix} (0.59) & 0.63 & 0.65 & 0.31 & 0.20 \\ 0.63 & (0.41) & 0.45 & 0.27 & 0.18 \\ 0.65 & 0.45 & (0.44) & 0.10 & 0.05 \\ 0.31 & 0.27 & 0.10 & (0.36) & 0.55 \\ 0.20 & 0.18 & 0.05 & 0.55 & (0.31) \end{bmatrix} \end{array} \tag{18}$$

The principal factors for $\mathbf{R}_c$ can be determined by finding the eigenvalues and eigenvectors of the above matrix. As the communality estimates are approximations, however, it is common practice to recompute them from the loadings determined for the principal factors, and to iterate this process until the communality estimates are stabilized. The iterated principal axis factor solution for the $\mathbf{R}_c$ matrix in Eqn. (18) is

Variables (j)	Factors (p) I	II	III	IV	V	h^2
$\mathbf{F}$ = English	0.880	−0.239	−0.009	−0.069	−0.017	0.837
French	0.684	−0.129	−0.204	0.068	0.005	0.531
Italian	0.642	−0.372	0.190	0.030	0.013	0.587
Physics	0.506	0.588	−0.021	−0.064	0.016	0.606
Chemistry	0.398	0.604	0.091	0.068	−0.013	0.536
Eigenvalues	2.069	0.923	0.087	0.019	0.001	

(19)

In the $\mathbf{F}$ matrix in Eqn. (19), an $n \times m$ matrix, the values of the communalities are obtained by $\sum_{p=1}^{m} a_{jp}^2$ for each test, and the eigenvalues by $\sum_{j=1}^{n} a_{jp}^2$ for each factor. It will be seen that the eigenvalues decrease in size from the first to later factors. The question arises as to how many of these factors are worth retaining for subsequent processing. If the original set of correlations or covariances can in fact be expressed in terms of a smaller number of underlying factors, the determination of the rank of the correlation matrix with appropriately chosen communality values would indicate the minimum number of factors needed to describe the original set of relationships among variables. The rank of a matrix is defined as the order of the highest nonvanishing determinant, or geometrically, as the minimum number of linearly independent dimensions or vectors needed to explain the data. If a matrix is of rank 2, the relationships among a set of variables can be expressed in two-dimensional space; if it is of rank 3, a three-dimensional space is required, and so on. With correlation or covariance matrices based on observed data in the social sciences, however, a clear-cut determination of the rank of a matrix is seldom possible. Apart from the problem of estimating communalities, observed data are subject to fluctuations due to the sampling of individuals and errors of measurement in the variables being analysed.

The number of factors of the original $\mathbf{R}$ matrix with eigenvalues greater than or equal to 1 is often taken as an indication of the number of initially extracted factors to be retained for further processing; such factors account for at least the equivalent of the total variance of any of the variables being analysed. In the $\mathbf{R}$ matrix on which Eqn. (18) is based, two eigenvalues are greater than 1. While this criterion is a useful starting point, it may underestimate the number of factors required to account for the correlational data, and may well be supplemented by other criteria. In Cattell's Scree test (1966), the eigenvalues are graphed from highest to lowest, and factors are accepted only for those eigenvalues above the point on the graph where the eigenvalues level off. Subjective criteria, such as discarding factors which account for less than 5 percent, say, of the total variance, on the grounds of their lack of practical importance, may also be considered. A further useful guide is the number of factors built into a well-designed factor analytic study.

The principal factor method is the most commonly used of the least squares approaches to the estimation of the initial $\mathbf{F}$ matrix; it is described as a least squares approach, since extracting the maximum variance at each stage is equivalent to minimizing the unexplained variance or residual correlations between the variables. An $\mathbf{F}$ matrix can also be generated directly from an iterative least squares solution involving the minimization of the residual correlations for an hypothesized number of factors; the Minres method (Harman 1976) is a variant of this approach.

Increasing use is being made of the method of maximum likelihood to determine the initial factor matrix, $\mathbf{F}$. The theoretical basis of the method had been given by Lawley in 1940, but its application did not become feasible until the development of new methods of maximum likelihood factor analysis (Jöreskog 1966, 1969, Jöreskog and Lawley 1968). The method is more efficient than other procedures, in the sense that the estimated factor loadings have a smaller sampling variance. It also provides a large sample test of significance for assessing the adequacy of different hypotheses about the number of common factors needed to account for the observed correlation or covariance matrix.

Under the principle of maximum likelihood, the parameter value(s) are sought which maximize the likelihood of a sample result. In its application to factor analysis, the parameter factor matrix $\mathbf{F}$ is estimated which would have the greatest likelihood, under a given hypothesis about the number of common factors, of generating the observed correlation or covariance matrix. This involves, in the case of uncorrelated factors, the minimization of a function

$G(\mathbf{F}, \boldsymbol{\psi})$ where $\mathbf{F}$ represents the matrix of factor loadings, and $\boldsymbol{\psi}$ the diagonal matrix of unique variances. When the maximum likelihood estimates of $\mathbf{F}$ and $\boldsymbol{\psi}$ have been determined, the hypothesis that the n-variable observed matrix can be accounted for by the designated number of common factors (k) can be tested for moderately large N through the χ^2 statistic with $\frac{1}{2}[(n-k)^2-(n+k)]$ degrees of freedom.

Application of the maximum likelihood factor analysis procedure to the correlation matrix in Eqn. (18) with unities in the diagonal cells showed that one factor was insufficient to account for the correlations ($\chi^2 = 69.487$, d.f. $= 5$, $p = 0.000$). Maximum likelihood loadings for an $\mathbf{F}$ matrix were then estimated on the assumption that the $\mathbf{R}$ matrix could be accounted for by two factors. This $\mathbf{F}$ matrix is

	Factor I	Factor II
Variables (j)		
$\mathbf{F}$ = English	0.934	−0.136
French	0.670	−0.033
Italian	0.657	−0.257
Physics	0.440	0.739
Chemistry	0.297	0.567
Eigenvalues	2.037	0.953

(20)

The probability that the observed correlation matrix $\mathbf{R}$ could have been generated from this $\mathbf{F}$ matrix is very high, namely 0.998 ($\chi^2 = 0$). The hypothesis that the observed correlation matrix can be accounted for by two underlying factors is therefore accepted. Following Jöreskog, the convention with empirically derived data is to accept the hypothesized number of factors as soon as the probability that the observed correlation matrix can be accounted for by that number of factors exceeds 0.10.

The significance test criterion in the maximum likelihood method tends to overestimate the number of factors when the sample size is large, and can be supplemented by other indices. The appearance of singlet factors, on which only one variable has a substantial loading, may also indicate that too many factors have been extracted. Comparison of the two-dimensional plots based on Eqn. (20) and the first two columns of the matrix in Eqn. (19) shows that the configurations from the maximum likelihood and principal factor solutions are quite similar.

Other approaches to the initial extraction of factors include the canonical factoring procedure, the Alpha factoring procedure, and image factoring. These approaches are available as options in computer packages such as SPSS and SAS.

In matrices with well-defined groupings of variables, as in the 5×5 correlation matrix in Eqn. (18), the various methods for the initial extraction of factors tend to identify the same factors even though factor loadings may differ from one solution to another. Most researchers will find that either the principal factor or maximum likelihood procedures will meet their needs, but it is often instructive to obtain both solutions.

5. Rotation of Factors

As outlined in Sect. 1, Thurstone argued that the initial factors needed to be rotated within the common factor space to arrive at a psychologically meaningful solution. He evolved the concept of simple structure to guide such rotations. As the principles of order implicit in simple structure are germane to a range of disciplines, the rotation of factors in exploratory factor analysis has continued to rely on this general concept.

In searching for new positions to which the original arbitrary orthogonal factor axes should be rotated to give substantive meaning to the factors, the investigator can choose to undertake an orthogonal or an oblique rotation. In the former case, the angles between all of the new factor axes remain at 90°, and the factors remain uncorrelated. In the latter case, the angles between the new axes can be smaller or larger than 90°, with the result that rotated factors may themselves be correlated. The difference between the two types of rotation is illustrated in Fig. 1 for the five-variable correlation problem in

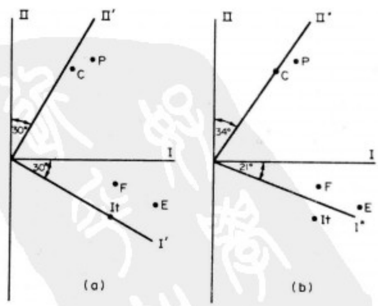

Figure 1
Orthogonal (a) and oblique (b) rotations of initial reference axes in Eqn. (21)

Eqn. (18), using as the initial factor plots the factor loadings from a two-factor principal factor solution for this matrix, since there were two eigenvalues greater than unity in the original $\mathbf{R}$ matrix. The principal factor matrix in this case is

$$\mathbf{F} = \begin{array}{l} \text{English} \\ \text{French} \\ \text{Italian} \\ \text{Physics} \\ \text{Chemistry} \end{array} \overset{\begin{array}{cc} \text{I} & \text{II} \end{array}}{\begin{bmatrix} 0.905 & -0.267 \\ 0.659 & -0.128 \\ 0.615 & -0.349 \\ 0.523 & 0.616 \\ 0.389 & 0.562 \end{bmatrix}} \qquad (21)$$

These loadings are plotted against the original factor axes I and II. The new positions of the axes after an orthogonal rotation are shown as I′ and II′ in Fig. 1(a). Figure 1(b) gives the new positions of the axes, I* and II*, after an oblique rotation.

In Fig. 1(a), the axes have been rotated clockwise through an angle of approximately 30°. Their placement could be subjectively determined, keeping in mind the need to have some variables with zero or near-zero loadings on each factor. The loadings of the variables on the new axes can be found from the formula $\mathbf{FT} = \mathbf{B}$, where $\mathbf{T}$ is the transformation matrix and $\mathbf{B}$ is the rotated factor matrix. In this particular rotation,

$$\mathbf{T} = \begin{array}{l} \text{I} \\ \text{II} \end{array} \overset{\begin{array}{cc} \text{I}' & \text{II}' \end{array}}{\begin{bmatrix} \cos(-29°45') & -\sin(-29°45') \\ \sin(-29°45') & \cos(-29°45') \end{bmatrix}}$$

$$= \begin{array}{l} \text{I} \\ \text{II} \end{array} \overset{\begin{array}{cc} \text{I}' & \text{II}' \end{array}}{\begin{bmatrix} 0.877 & 0.481 \\ -0.481 & 0.877 \end{bmatrix}} \qquad (22)$$

$$\mathbf{B} = \begin{array}{l} \text{English} \\ \text{French} \\ \text{Italian} \\ \text{Physics} \\ \text{Chemistry} \end{array} \overset{\begin{array}{cc} \text{I}' & \text{II}' \end{array}}{\begin{bmatrix} 0.92 & 0.20 \\ 0.64 & 0.20 \\ 0.71 & -0.01 \\ 0.16 & 0.79 \\ 0.07 & 0.68 \end{bmatrix}} \qquad (23)$$

The factor coefficients or factor loadings in Eqn. (23) could be read directly from Fig. 1(a) by measuring the orthogonal projections of the test points on axes I′ and II′ respectively. This matrix of orthogonal projections of the test points on the new axes is known as the factor structure matrix; the entries represent the correlations between the test vectors and the factors. The matrix of coordinates of the test points on the new axes, however, defines the factor pattern matrix; this is the matrix of coefficients which would be needed to estimate the standard scores of each person on the original variables, as set out in Eqn. (1). The factor pattern and factor structure are identical in an orthogonal factor solution but differ in an oblique solution.

In Fig. 1(b), the axes have been placed through the two distinct clusters of points, so that each cluster will have high loadings on one factor and zero or near-zero loadings on the other. Factor I has been rotated clockwise through 21°27′ to the new position I*, and Factor II through 34°28′ to the new position II*. Since the angle of 77° between the two new axes is not a right angle, the rotation is oblique. The orthogonal projection of the end-points of the test vectors on the new axis system is given by $\mathbf{F\Lambda} = \mathbf{S}$ where

$$\Lambda = \begin{array}{l} \text{I} \\ \text{II} \end{array} \overset{\begin{array}{cc} \text{I}^* & \text{II}^* \end{array}}{\begin{bmatrix} \cos(-21°27') & \cos 55°32' \\ \sin(-21°27') & \sin 55°32' \end{bmatrix}}$$

is the matrix of direction cosines of the new axes with respect to the original axes. In the present example,

$$\begin{array}{l} \text{English} \\ \text{French} \\ \text{Italian} \\ \text{Physics} \\ \text{Chemistry} \end{array} \overset{\begin{array}{cc} \text{I} & \text{II} \end{array}}{\begin{bmatrix} 0.905 & -0.267 \\ 0.659 & -0.128 \\ 0.615 & -0.349 \\ 0.523 & 0.616 \\ 0.389 & 0.562 \end{bmatrix}} \overset{\begin{array}{cc} \text{I}^* & \text{II}^* \end{array}}{\begin{bmatrix} 0.9307 & 0.5659 \\ -0.3657 & 0.8245 \end{bmatrix}}$$

$$= \overset{\begin{array}{cc} \text{I}^* & \text{II}^* \end{array}}{\begin{bmatrix} 0.94 & 0.29 \\ 0.66 & 0.27 \\ 0.70 & 0.06 \\ 0.26 & 0.80 \\ 0.16 & 0.68 \end{bmatrix}} \qquad (24)$$

The matrix on the right-hand side of Eqn. (24) is the factor structure matrix, $\mathbf{S}$, which represents the correlations of the tests with the factors. The correlation between the two rotated factors is given by the off-diagonal element in

$$\boldsymbol{\phi} = \Lambda'\Lambda = \begin{bmatrix} 1.000 & 0.225 \\ 0.225 & 1.000 \end{bmatrix}$$

which represents a moderate degree of correlation. In representing the scores of persons in terms of a smaller number of factors, however, the coordinates of the test vectors with respect to the new oblique axes, which form the factor pattern matrix, are of more interest. These are given by $\mathbf{P} = \mathbf{S}\boldsymbol{\phi}^{-1}$; in the present example, the factor pattern matrix is

$$\mathbf{P} = \begin{array}{l} \text{English} \\ \text{French} \\ \text{Italian} \\ \text{Physics} \\ \text{Chemistry} \end{array} \overset{\begin{array}{cc} \text{I}^* & \text{II}^* \end{array}}{\begin{bmatrix} 0.92 & 0.08 \\ 0.63 & 0.13 \\ 0.72 & -0.10 \\ 0.08 & 0.78 \\ 0.01 & 0.68 \end{bmatrix}} \qquad (25)$$

The $\mathbf{S}$ matrix in Eqn. (24) could be read directly from Fig. 1(a) by finding the orthogonal projections of the test vectors on Factors I* and II*, and the **P** matrix in Eqn. (25) by finding their oblique projections on these two factors.

A comparison of the matrices in Eqns. (23) and (25) shows the advantages of oblique over orthogonal rotations if the factors are correlated. The factor definition is clearer in the oblique solution; zero or near-zero loadings indicate more clearly that physics and chemistry are not represented in Factor I* and that the three languages are not represented in Factor II*. Some knowledge of the nature of the variables is required to interpret the factors. Each factor must be inspected to determine what the variables with high loadings have in common which is not present in the variables with low loadings, and then named appropriately. The task is deceptively simple in the present example. Since the variables with high loadings on Factor I′/I* are language examinations, and the variables with low loadings are not, this factor can be interpreted as a language ability factor. Similarly, Factor II′/II* can be interpreted as a scientific ability/achievement factor. The task of interpretation can be much more demanding in studies involving many variables and several factors.

The new positions of the axes in Fig. 1 were obtained by analytic methods of rotation, which replaced the subjective graphical methods used prior to the 1950s, in which investigators inspected plots of each pair of factors from the **F** matrix. The first fully analytic procedures for rotation developed by Carroll (1953) and other factor analysts became known as quartimax procedures. They aimed to simplify the rows of a matrix of factor loadings, by maximizing the sum of the fourth powers of the loadings or by some equivalent criterion, with communalities held constant. The first factor defined by the quartimax procedures tends to be a general factor, and the procedures are not widely used.

The analytic criterion used to obtain the orthogonal factor solution in Fig. 1(a) was developed and subsequently refined by Kaiser (1958). It is known as the varimax criterion, and aims to simplify the columns of the factor matrix by maximizing over all factors the variance of the squared factor loadings in each column, after first dividing each factor loading by the square root of the relevant variable's communality to give equal weight to the factors in the rotation. It requires the minimization of the function

$$V = n \sum_{p=1}^{m} \sum_{j=1}^{n} \left(\frac{a_{jp}}{h_j}\right)^4 - \sum_{p=1}^{m} \left(\sum_{j=1}^{n} \frac{a_{jp}^2}{h_j^2}\right)^2 \qquad (26)$$

The varimax criterion, which is designed to generate factors on which some variables have high loadings and others have low loadings, has been found to be highly satisfactory for orthogonal rotations, and is very widely used.

The placement of the new axes in Fig. 1(b) was determined with the aid of the most widely used oblique analytic rotation criterion, known as the direct oblimin criterion (Jennrich and Sampson 1966). This criterion has replaced oblique analytic rotational criteria based on the factor structure matrix which were developed in the late 1950s, for example, Carroll's biquartimin criterion. Following the same principles as the latter criterion, that is, the minimization over pairs of factors of the cross-products of squared factor loadings, the minimization of the covariances of these squared loadings, and the use of a coefficient to vary the relative weight given to these two components in order to control the degree of obliqueness of the factors, Jennrich and Sampson rotated the **F** matrix directly to the factor pattern matrix, **P**, by minimizing the function

$$G(\mathbf{P}) = \sum_{p<q=1}^{m} \left(\sum_{j=1}^{n} b_{jp}^2 b_{jq}^2 - \frac{\delta}{n} \sum_{j=1}^{n} b_{jp}^2 \sum_{j=1}^{n} b_{jq}^2\right) \qquad (27)$$

where b_{jp} and b_{jq} are the elements of the matrix **P** and δ is the variable quantity which controls the degree of obliqueness of the factors. Computer packages usually allow the investigator to apply a range of values of δ to facilitate the selection of a solution which best conforms to simple structure. Factors tend to be too oblique when $\delta = 0$, and become less oblique as δ becomes more negative. A delta value of approximately -0.5 has often been found to yield relatively "clean" simple structure solutions.

6. *Advances in Confirmatory Factor Analysis*

The major advance in factor analysis since the late 1960s has been the development of confirmatory factor analytic procedures. In the course of developing maximum likelihood procedures for exploratory factor analysis, Jöreskog saw their possibilities for testing hypothesized matrices. Recognizing that factor analysts generally wished to specify only some of the parameters in a hypothesized matrix, and to allow others to vary, he reformulated the factor analysis model to incorporate fixed parameters, constrained parameters (unknown in value but equal to one or more other parameters), and free parameters. He expressed the model in terms of a variance–covariance or dispersion matrix which becomes a correlation matrix if the variables are in standardized form. That is,

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\phi}\boldsymbol{\Lambda}' + \boldsymbol{\psi} \qquad (28)$$

where $\boldsymbol{\Sigma}$ is the dispersion matrix of observed scores, $\boldsymbol{\Lambda}$ is an $n \times m$ matrix of factor loadings, $\boldsymbol{\phi}$ is the factor correlation matrix, and $\boldsymbol{\psi}$ is the diagonal matrix of unique variances. In confirmatory factor analysis, the investigator is free to specify fixed values for particular parameters in $\boldsymbol{\Lambda}$, $\boldsymbol{\phi}$, and $\boldsymbol{\psi}$, given some restriction on the total number of fixed parameters. The matrix $\hat{\boldsymbol{\Sigma}}$ is then estimated by maximum likelihood procedures under these conditions, and a χ^2

test applied to determine whether the observed dispersion matrix $\mathbf{\Sigma}$ differs from the estimated matrix $\hat{\mathbf{\Sigma}}$.

The model allows great flexibility for testing a wide variety of hypothesized factor patterns, in which relationships among the factors may be orthogonal or oblique or a mixture of the two. Confirmatory factor analysis has been used, for example, to analyse data from multitrait, multimethod studies (Werts et al. 1972), to illuminate a long-standing controversy on the identification of reading comprehension skills (Spearritt 1972), and to test the simplex assumption underlying Bloom's taxonomy of educational objectives (Hill and McGaw 1981). It has also facilitated the comparison of the factorial structure of different subpopulations, allowing investigators to determine whether the factorial structure of a given set of variables varies, for example, with sex, age, ethnicity, socioeconomic status, or political affiliation, and if so, in what manner (e.g., McGaw and Jöreskog 1971)?

The model set out in Eqn. (28) forms part of a more general model for the analysis of covariance structures, which was subsequently elaborated by Jöreskog to handle a wide range of statistical models for multivariate analysis. The LISREL V suite of computer programs (Jöreskog and Sörbom 1981) which provide for the analysis of linear structural relationships by the method of maximum likelihood, has become a basic tool for studying not only exploratory and confirmatory factor analysis models, but also path analysis models and models relating to cross-sectional and longitudinal data (see *Structural Equation Models*).

7. *Some Additional Methodological Aspects of Factor Analysis*

7.1 *Construction of Factor Scales*

When factors are identified as a result of a factor analysis, it is possible to calculate a factor score for each person on the new factors, for example a language score and a science score. With some exceptions, the calculation of factor scores has not been an important feature of educational and psychological studies, in which the emphasis has been mainly on the identification rather than the measurement of factors. In some disciplines, however, the chief concern has been to create composite factor scales to facilitate further study of a topic.

The most widely used method of calculating a person's factor scores has been to regress the factor loadings on each factor in the factor structure matrix against the original set of variables. A matrix of factor-score coefficients or regression weights can be found from the formula

$$\mathbf{W} = \mathbf{S}'\mathbf{R}^{-1} \qquad (29)$$

where **S** is the rotated factor structure matrix and **R** is the original correlation matrix. Factor scores are conventionally presented as standard scores, derived by applying the regression weights to a person's standard score on each of the original variables. Other approaches to the estimation of factor scores are outlined by Harman (1976) and Kim and Mueller (1978).

7.2 *Hierarchical Factor Solutions*

Hierarchical factor solutions were attractive to early British factorists because of their hierarchical theories of cognitive processes. Accordingly, a general factor was extracted from the correlation matrix as a first step; group factors were then extracted from the residual correlations (Burt 1950, Vernon 1961). Even without the initial assumption of a general factor, the American oblique rotational methods could still yield an hierarchical factor solution. Provided the data yielded at least three primary factors, the correlations among these factors could themselves be analysed to arrive at second-order factors; if there were sufficient primary factors to yield several second-order factors, the latter could be analysed to yield third-order factors, and so on. If the matrix of primary-factor correlations were of unit rank, a second-order general factor would emerge. If desired, such hierarchical solutions could be made orthogonal (Schmid and Leiman 1957).

7.3 *Comparison of Factors*

Coefficients of congruence designed to measure the degree of similarity between pairs of factors derived from different sets of variables in the same domain, and the degree of similarity between loadings on pairs of corresponding factors derived when the same set of variables is applied to different subpopulations, are summarized in Harman (1976). Comparisons of factor matrices can be made through confirmatory factor analysis procedures.

7.4 *Assumptions of Linearity*

It is usually assumed in factor analysis (necessarily so with maximum likelihood procedures) that the variables have a multivariate normal distribution in the population which has been sampled, and this implies that the variables are linearly related. Where this is not the case, multivalued variables may be normalized as a first step. Care needs to be taken in applying factor analysis to dichotomously scored variables such as test or scale items (Kim and Mueller 1978, Muthén 1981). Factor analysis models in which factors are not linearly related to variables have been extensively investigated by McDonald (1967).

8. *Computer Programs*

Widely available statistical packages such as SPSS, SAS, BMDP, and OSIRIS all contain factor analysis programs. For the initial extraction of factors, the researcher usually has the option of selecting the principal factor, maximum likelihood, Rao canonical, Alpha, or image method of factoring. Varimax and direct oblimin rotational solutions with nominated values of δ are available in most programs, along with other rotational methods such as, for instance, Quartimax and Equimax in SPSS and SAS, and Promax in SAS. Two-dimensional plots of the rotated factors, and the necessary matrix of coefficients for producing factor scores, are also usually obtainable.

In addition to the LISREL program mentioned in Sect. 6, another special purpose program, COFAMM, is available for confirmatory factor analysis.

9. Applications of Factor Analysis

Factor analysis has made its most direct contribution to education through its influence on the composition of test batteries used for educational or vocational guidance. Batteries of tests such as the SRA Primary Mental Abilities battery and the Psychological Corporation's Differential Aptitude Tests were designed to yield separate scores for students on aptitudes or abilities such as number computation, verbal reasoning, verbal comprehension, abstract reasoning, clerical speed and accuracy, mechanical reasoning, space relations, language usage, and word fluency. Factor analytic studies have also contributed to the selection of areas to be tested in achievement test batteries, such as reading comprehension, listening comprehension, and comprehension and interpretation in mathematics, science, and social studies. Factor analysis has served to identify skills, abilities, and areas of achievement which are relatively independent, and has thus avoided unnecessary duplication of measurement in providing a profile of a student's performance. Factor studies have also often provided the framework for personality and interest inventories used in guidance and counselling.

The major impact of factor analysis has been in the area in which it was first employed, that is, in the study of intellectual or cognitive abilities. It has been the chief technique for exploring the structure of human abilities. It has been used to map the broad areas of human abilities which are needed to account for the variation which occurs in the performance of subjects on a great variety of mental tasks. A test kit of confirmed factors of cognitive abilities was prepared at the Educational Testing Service (French 1954) and was revised and extended in 1963 and 1976; the kit has been of great value in defining factors for use in further exploratory studies. Abilities isolated at one level, such as reasoning ability or memory, have also been subjected to detailed factor analyses of their infrastructure. Studies of these kinds, supplemented by the very extensive factor studies undertaken in connection with Guilford's Structure of Intellect model (Guilford 1967), and Cattell's theory of "crystallized" and "fluid" intelligence (Cattell 1971), have produced a very considerable body of knowledge about the structure of human abilities.

In applications in education, factor analytic studies have been undertaken in such diverse areas as prose style, administrative behaviour, occupational classification, attitudes and belief systems, and the economics of education. The technique is still in extensive use in the exploration of abilities, in the refining of tests and scales, and in the development of composite variables for use in research studies. Its most promising applications in recent years, however, have been concerned with the testing of explicit hypotheses about the structure of sets of variables, as in the study of growth models and other models mentioned in Sect. 6.

Factor analysis will remain an important technique for reducing and classifying sets of variables as a means of improving theoretical understanding in various disciplines, and for testing hypotheses about structural relationships among sets of variables. Confirmatory factor analysis procedures should assist in the formulation of more precise theories about such structural relationships; current theories about the structure of educational or psychological domains have rarely been formulated in sufficiently explicit terms to attract support from these procedures. In the search for explanations about how and why such structural relationships take the form they do, closer links can be expected to be developed between factor analysis and path analysis models. Methodological developments might be expected in the application of factor analysis to dichotomously scored variables and to categorical variables, and in the development of nonlinear models where linear models prove to be inadequate. Considerable scope remains for research on the emergence of factors, involving neurological, general environmental, and schooling influences; relationships between factorial and information processing models (Sternberg 1977) also need investigation. Finally, comprehensive factor studies of the abilities tested in different school subjects would be highly relevant to the design of school curricula.

See also: Path Analysis; Statistical Analysis in Educational Research; Factorial Modeling

Bibliography

Burt C 1950 Group factor analysis. *Br. J. Psychol. Stat. Sect.* 3: 40–75

Carroll J B 1953 Approximating simple structure in factor analysis. *Psychometrika* 18: 23–38

Cattell R B 1966 The Scree test for the number of factors. *Mult. Behav. Res.* 1: 245–76

Cattell R B 1971 *Abilities, Their Structure, Growth and Action.* Houghton Mifflin, Boston, Massachusetts

French J W (ed.) 1954 *Manual for Kit of Selected Tests for Reference Aptitude and Achievement Factors*. Educational Testing Service, Princeton, New Jersey
Guilford J P 1967 *The Nature of Human Intelligence*. McGraw-Hill, New York
Harman H H 1976 *Modern Factor Analysis*, 3rd edn. University of Chicago Press, Chicago, Illinois
Hill P W, McGaw B 1981 Testing the simplex assumption underlying Bloom's taxonomy. *Am. Educ. Res. J.* 18: 93–101
Jennrich R I, Sampson P F 1966 Rotation for simple loadings. *Psychometrika* 31: 313–23
Jöreskog K G 1966 Testing a simple structure hypothesis in factor analysis. *Psychometrika* 31: 165–78
Jöreskog K G 1969 A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika* 34: 183–202
Jöreskog K G, Lawley D N 1968 New methods in maximum likelihood factor analysis. *Br. J. Math. Stat. Psychol.* 21: 85–96
Jöreskog K G, Sörbom D 1981 *Lisrel V (User's Guide)*. National Educational Resources, Chicago, Illinois
Kaiser H F 1958 The varimax criterion for analytic rotation in factor analysis. *Psychometrika* 23: 187–200
Kim J O, Mueller C W 1978 *Factor Analysis: Statistical Methods and Practical Issues*. Sage, London
Lawley D N, Maxwell A E 1971 *Factor Analysis as a Statistical Method*, 2nd edn. Butterworth, London
McDonald R P 1967 *Nonlinear Factor Analysis*. Psychometric Society, Richmond
McGaw B, Jöreskog K G 1971 Factorial invariance of ability measures in groups differing in intelligence and socio-economic status. *Br. J. Math. Stat. Psychol.* 24: 154–68
Mulaik S A 1972 *The Foundations of Factor Analysis*. McGraw-Hill, New York
Muthén B 1981 Factor analysis of dichotomous variables: American attitudes toward abortion. In: Jackson D J, Borgatta E F (eds.) 1981 *Factor Analysis and Measurement in Sociological Research: A Multi-dimensional Perspective*. Sage, London
Schmid J, Leiman J M 1957 The development of hierarchical factor solutions. *Psychometrika* 22: 53–61
Spearman C 1904 "General intelligence" objectively determined and measured. *Am. J. Psychol.* 15(2): 201–93
Spearritt D 1972 Identification of subskills of reading comprehension by maximum likelihood factor analysis. *Read. Res. Q.* 8: 92–111
Sternberg R J 1977 *Intelligence, Information Processing, and Analogical Reasoning: The Componential Analysis of Human Abilities*. Erlbaum, Hillsdale, New Jersey
Thomson G H 1951 *The Factorial Analysis of Human Ability*, 5th edn. Houghton Mifflin, Boston, Massachusetts
Thurstone L L 1931 Multiple factor analysis. *Psychol. Rev.* 38: 406–27
Thurstone L L 1947 *Multiple-factor Analysis: A Development and Expansion of the Vectors of the Mind*. University of Chicago Press, Chicago, Illinois
Vernon P E 1961 *The Structure of Human Abilities*, 2nd edn. Methuen, London
Werts C E, Jöreskog K G, Linn R L 1972 A multitrait–multimethod model for studying growth. *Educ. Psychol. Meas.* 32: 655–78

D. Spearritt

Factorial Modeling

Factorial modeling (FaM) provides researchers with a simple method for constructing latent structural variables linking theory with correlational data. Structural variables emphasize construct validity rather than predictive power in the linear components fitted by data analysis to vector variables. At the research design stage, the researcher who is planning to use FaM is required to name the latent variables (hereafter called factors) and to specify each of them by means of an exclusive subset of the independent variates (i.e., the measurements). The researcher is also required to specify an order of extraction of the factors, because the mathematical simplicity and computational efficiency of FaM are purchased at the expense of order dependence of the factors. The FaM data analysis yields a structural equation for every observational variate and attributes all explained variance unambiguously among the factors. This is possible because the factors are constructed to be mutually uncorrelated. Partitioning of criterion variance into independent contributions from uncorrelated factors is particularly useful in educational policy studies, where intervention decisions have to be justified by straightforward causal inferences.

The method was named factorial modeling to encourage comparison with factorial analysis of variance. In the design of experiments, a balanced sampling scheme creates uncorrelatedness of the causal factors to permit unambiguous causal inferences. In observational studies, FaM creates uncorrelated causal factors by analysis where they cannot be created by the sampling scheme. Causal inferences from FaM do not possess the high internal validity of inferences from a randomized factorial experiment, but if the data represent an important natural system in situ, the external validity of the factorial model may be much higher than that of any possible experiment. When the constructs of a theory represent interdependent attributes, the corresponding factors of an FaM model for data represent the partial contents of their constructs which are nonoverlapping with the contents of previously extracted factors. Thus FaM is a method which resembles multiple partial correlation.

FaM does not employ an algorithm derived by the differential calculus of linear systems. The method may appeal to researchers who fear that multivariate regression methods overpower many of the data collections available in education. The rationale for FaM is that using noncalculus multivariate mathematics on data will produce loose fitting models that may not be readily transferred to new situations.

1. Mathematics

Factorial modeling extracts ordered orthogonal factors by the method of matrix exhaustion. Each fact is specified by assignment to it of an exclusive subset of the independent variates. It is weighted by the covariances of those specifying variates with a selected dependent variate in the residual matrix prevailing at the start of that factoring step. For the first factor the weights are simply the bivariate correlations of the specifying variates with the selected criterion, or simple predictive validities. For all later factors, the weights are residual covariances representing residual predictive validities. It is not necessary that every independent variate be assigned to the specification of one of the factors, but it is highly desirable that at least two variates specify each factor. Otherwise, the total variance for a single variate specifying a factor will be swept out, resulting in $h^2 = 1$ for that variate and a degenerate structural equation for it. The object of modeling is to regress criteria on latent variables, not on observed variates.

Let an idempotent matrix containing ones on the main diagonal in the positions corresponding to the positions in the correlation matrix of the specifying variates for the kth factor and zeros everywhere else be identified as $\mathbf{I}_k$. Let

$$\mathbf{v}_k = \mathbf{I}_k \, \mathbf{r}_c \tag{1}$$

where $\mathbf{r}_c$ is the column of the residual matrix $\mathbf{C}_k$ which belongs to the selected criterion. Then $\mathbf{v}_k$ is a vector of the order of $\mathbf{R}$ (which may be designated p for the count of all the measurement variates, independent plus dependent), but the only nonzero elements of $\mathbf{v}_k$ are predictive validities of the specification variates for the kth factor. For the following equation

$$\mathbf{h}_k = (1/\sqrt{\mathbf{v}_k' \, \mathbf{C}_k \, \mathbf{v}_k}) \, \mathbf{v}_k \tag{2}$$

when $k = 1$, $\mathbf{C}_1 = \mathbf{R}$, the correlation matrix for the measurements. When $k = 2$, $\mathbf{C}_2$ is the covariance matrix remaining after the first factor has been exhausted from $\mathbf{R}$. In general, $\mathbf{C}_k$ is the residual covariance matrix after $k - 1$ factors have been exhausted.

Then the structural coefficients for the kth factor are

$$\mathbf{s}_k = \mathbf{C}_k \, \mathbf{h}_k \tag{3}$$

and $\mathbf{C}_k$ is exhausted of $\mathbf{s}_k$.

$$\mathbf{C}_{k+1} = \mathbf{C}_k - \mathbf{s}_k \, \mathbf{s}_k' \tag{4}$$

When all n planned factors have been computed, their column vectors of structural coefficients are assembled in a $p \times n$ matrix $\mathbf{S}$. Now the theory plus error partition of $\mathbf{R}$ is given by

$$\mathbf{R} = \mathbf{S}\,\mathbf{S}' + \mathbf{C}_{n+1} \tag{5}$$

The elements of the main diagonal of the theory matrix $\mathbf{SS}'$ are the proportions of the variate variances explained by the theory for the data, called the communalities, h_j^2. The square roots of the elements of the main diagonal of $\mathbf{C}_{n+1}$ are the disturbance weights, d_j, which apply to the combined unknown sources of variance in each of the variates. From these results a structural equation can be written for each variate

$$z_j = s_{j1} f_1 + s_{j2} f_2 + \ldots + s_{jn} f_n + d_j \, u_j \tag{6}$$

In this equation the f_k are the factor scores and the u_j is a uniqueness score, that is, a score for the combination of all other sources of variance in z_j. Dropping the final addend gives the multiple regression equation

$$\hat{z}_j = s_{j1} f_1 + s_{j2} f_2 + \ldots + s_{jn} f_n \tag{7}$$

This shows that any structural coefficient s_{jk}, besides being a product–moment correlation between a variate and a factor, is also a standardized multiple regression weight for the regression of the jth variate on the kth factor, and its square, s_{jk}^2, is the contribution of the kth factor to the explanation of the variance in the jth variate. The squared multiple correlation coefficient is the communality

$$R_j^2 = h_j^2 = s_{j1}^2 + s_{j2}^2 + \ldots + s_{jn}^2 \tag{8}$$

Thus the canon of unambiguous attribution of variance is satisfied.

Two salient facts emerge from this mathematics. The exact definition achieved for each factor beyond the first is order dependent, in the sense that it depends in part on the order in which the factors are extracted. Also, as factoring continues, the degrees of freedom for arbitrary location of a factor are reduced and the disciplinary force of the uncorrelatedness requirement over the hypothetical location of the factor becomes stronger.

As originally proposed by Lohnes (1979), FaM required the designation of a key criterion toward which all the factors were oriented. Lohnes has modified the algorithm so that when there are multiple criteria, it searches the vectors of residual predictive validities of the specification variates for a factor to find the largest sum of squares of those validities, and orients the factor to that criterion for which it has, in this sum of squares sense, the largest predictive validity. The program provides for the user to override this feature by designating the criterion variate toward which each factor is to be oriented, and it is permissible to designate the same criterion for all the factors, thus restoring the original emphasis on a key criterion.

The current program for FaM also incorporates an improved algebra for computing coefficients defining the latent variables as linear functions of some of the variates. The new algebra supplies true zero coefficients in every possible place. Only the specification variates for the first factor enter its operational definition. Only the specification variates for the first two factors enter the operational definition of the second factor. For any other factor, only