Chapter 2

A General Approach to Confirmatory Maximum Likelihood Factor Analysis

with Addendum

Karl G. Jöreskog

We describe a general procedure by which any number of parameters of the factor analytic model can be held fixed at any values and the remaining free parameters estimated by the maximum likelihood method. The generality of the approach makes it possible to deal with all kinds of solutions: orthogonal, oblique and various mixtures of these. By choosing the fixed parameters appropriately, factors can be defined to have desired properties and make subsequent rotation unnecessary. The goodness of fit of the maximum likelihood solution under the hypothesis represented by the fixed parameters is tested by a large sample χ^2 test based on the likelihood ratio technique. A by-product of the procedure is an estimate of the variance-covariance matrix of the estimated parameters. From this, approximate confidence intervals for the parameters can be obtained. Several examples illustrating the usefulness of the procedure are given.

1. Introduction and Summary

We shall describe a general procedure for performing factor analysis in the following way. Any values may be specified in advance for any number of factor loadings, factor correlations and unique variances. The remaining free parameters, if any, are estimated by the maximum likelihood method. A typical application of the procedure is in confirmatory factor studies, where the experimenter has already obtained a certain amount of knowledge about the variables measured and is therefore in a position to formulate a hypothesis that specifies some of the factors involved. For exploratory maximum likelihood factor analysis a computer program has been made available earlier [Jöreskog, 1967(a–b)]. This can be used to determine an appropriate number of factors to use and a preliminary interpretation of the data. The present procedure can then be used for a more precise analysis. We shall give examples of how a preliminary interpretation of the factors can be successively modified to determine a final solution that is acceptable from the point of view of both goodness of fit and psychological interpretation. It is highly desirable that a hypothesis that has been generated in this way should subsequently be confirmed or disproved by obtaining new data and subjecting these to a confirmatory analysis. Jöreskog and Lawley [1967], giving an expository

---

This work was supported by a grant (NSF-GB 1985) from the National Science Foundation to Educational Testing Service.

account of both exploratory and confirmatory methods, present an example where the original sample was randomly divided into two halves, where one half was used to generate a hypothesis and the other half was used to test this hypothesis.

The approach of this paper is similar to those of Howe [1955], Anderson and Rubin [1956], Lawley [1958] and Jöreskog [1966], but is more general in that it is possible to deal with all kinds of solutions: orthogonal, oblique and various mixtures of these. The fixed elements need not necessarily be zeros and the restrictions need not even be sufficient to make the solution unique. Factors can be defined to have desired properties and if a preliminary interpretation is available, the restrictions can be chosen to make any subsequent rotation unnecessary. Several examples are given in Sections 4 and 5 to illustrate the usefulness of the procedure.

The computational procedure is based on the minimization method of Fletcher and Powell [1963] and yields as a by-product an estimate of the variance-covariance matrix of the estimated parameters. From this, approximate confidence intervals for the parameters can easily be obtained.

2. *Preliminary Considerations*

In factor analysis the basic model is

(1) $$y = \Lambda x + z,$$

where y is a vector of order p of observed test scores, x is a vector of order $k < p$ of latent common factor scores, z is a vector of order p of unique scores, and $\Lambda = (\lambda_{im})$ is a $p \times k$ matrix of factor loadings. It is assumed that $E(x) = E(z) = 0$, $E(xx') = \Phi$ and $E(zz') = \Psi$, a diagonal matrix. From these assumptions one deduces that the dispersion matrix of y, $\Sigma = E(yy')$, is

(2) $$\Sigma = \Lambda\Phi\Lambda' + \Psi.$$

The basic idea of the model is that the common factors x shall account for all correlations between the y's. Once factors x have been partialed out, there shall remain no correlation between the tests.

The elements of Λ, Φ and Ψ are parameters of the model to be estimated from the data. Suppose that from a random sample of $n + 1$ observations of y we find the matrix S whose elements are the usual estimates of variances and covariances of the components of y. If y has a multivariate normal distribution, the elements of S follow a Wishart distribution with n degrees of freedom. The log-likelihood function, neglecting a function of the observations, is given by

(3) $$\log L = -\tfrac{1}{2}n[\log |\Sigma| + \operatorname{tr}(S\Sigma^{-1})].$$

Let

(4) $$F(\Lambda, \Phi, \Psi) = \log |\Sigma| + \operatorname{tr}(S\Sigma^{-1}) - \log |S| - p.$$

Then maximizing $\log L$ is equivalent to minimizing F, and n times the minimum value of F is equal to the likelihood ratio test statistic of goodness of fit [see e.g., Jöreskog, 1967b]. It should be noted, however, that if T is any nonsingular $k \times k$ matrix, then

(5) $$F(\Lambda T^{-1}, T\Phi T', \Psi) = F(\Lambda, \Phi, \Psi).$$

This means that the parameters in Λ and Φ are not independent of one another, and in order to make the maximum likelihood estimates of Λ and Φ unique, we must impose k^2 independent restrictions on Λ and Φ. In an exploratory factor analysis, where no hypotheses concerning the factors are involved, it is convenient to choose these restrictions so that $\Phi = I$ and $\Lambda'\Psi^{-1}\Lambda$ is diagonal [see e.g., Lawley & Maxwell, 1963; Jöreskog, 1967b]. In a confirmatory factor analysis, on the other hand, the investigator has certain hypotheses as to which factors are to be involved in certain tests, and it is therefore convenient to choose the restrictions by requiring that certain elements of Λ and Φ have values specified in advance. For example, $\lambda_{im} = 0$ means that the m-th factor does not enter into the i-th test and $\phi_{rs} = 0$ means that factors r and s are uncorrelated. Values other than zero could be used also. Depending on the number, values and positions of the fixed elements in Λ, Φ and Ψ, we may distinguish between two kinds of solutions: unrestricted and restricted. An *unrestricted solution* is one that does not restrict the common factor space, i.e., one that leaves $\Lambda\Phi\Lambda'$ unrestricted. All such solutions can be obtained by a rotation of an arbitrary unrestricted orthogonal maximum likelihood solution, for example, one obtained by the computational procedure of Jöreskog [1967a–b]. An unrestricted solution will usually result if the number of fixed elements in Λ and Φ is at most k^2 and if these elements are properly distributed over all factors. All unrestricted solutions for the same data will have the same communalities and the same uniquenesses, and they will all yield the same fit to the observed variances and covariances in S. In an unrestricted solution no element of Ψ can be held fixed, since clearly a restriction on Ψ is a restriction of the common factor space. A *restricted solution*, on the other hand, imposes restrictions on the whole factor space, and such a solution, therefore, cannot be obtained by a rotation of an unrestricted solution. Communalities and uniquenesses will not be the same for an unrestricted and a restricted solution for the same data. The fit to the observed variances and covariances in S, as measured by the function F, will, in general, be better for the unrestricted than for the restricted solution. However, if differences in number of estimated parameters are taken into account this may not be so.

Both unrestricted and restricted solutions may or may not be unique. This depends on the positions and values of the fixed parameters. A solution is unique if all linear transformations of the factors that leave the fixed parameters unchanged also leave the free parameters unchanged. Various sufficient conditions for a unique solution have been given by Reiersøl [1950], Howe [1955] and Anderson and Rubin [1956].

Two simple sufficient conditions, as given by Howe [1955], are as follows. In the orthogonal case, let $\Phi = I$ and let the columns of Λ be arranged so that, for $s = 1, 2, \cdots, k$, column s contains at least $s - 1$ fixed elements. In the oblique case, let diag $\Phi = I$ and let each column of Λ have at least $k - 1$ fixed elements. It should be noted that in the orthogonal case there are $\frac{1}{2}k(k + 1)$ conditions on Φ and a minimum of $\frac{1}{2}k(k - 1)$ conditions on Λ. In the oblique case there are k normalizations in Φ and a minimum of $k(k - 1)$ conditions on Λ. Thus, in both cases, there is a minimum of k^2 specified elements in Λ and Φ. Let Λ be a solution under any such conditions and let $\Lambda^{(s)}$ be the submatrix of Λ, consisting of those rows of Λ, that has fixed elements in the s-th column. Then Λ is unique if for all $s = 1, 2, \cdots, k$, $\Lambda^{(s)}$ has rank equal to the smallest of the numbers m_s and k, where m_s is the number of fixed elements in the s-th column of Λ. This condition is usually fulfilled in practice. It should be noted that even a restricted solution need not be unique. For example, an orthogonal solution with no restrictions on the first two columns of Λ and with more than $\frac{1}{2}k(k - 1)$ fixed elements in the other columns is restricted but not unique. Any orthogonal rotation in the plane of the first two factors will change these and leave all the fixed elements unchanged. In general, if a solution is not unique, transformations may exist that change the free parameters while leaving the fixed parameters unchanged. To make the solution unique, additional restrictions must be imposed.

In addition to the two main kinds of restrictions on Φ, *orthogonal* and *oblique*, various mixtures of these can be used. For example, one factor may be postulated to be uncorrelated with all the others—these factors being correlated among themselves. It is convenient to refer to all such solutions as *mixed* solutions.

Examples of many different kinds of solutions are given in Sections 4 and 5.

3. Minimization Procedure

The minimization problem is that of minimizing the function $F(\Lambda, \Phi, \Psi)$ with respect to the free parameters, keeping the others fixed at the given values. A numerical procedure for obtaining the maximum likelihood estimates, under certain special cases, was first given by Howe [1955]. A very similar procedure was later proposed by Lawley [1958] and referred to subsequently by Lawley and Maxwell [1963]. In both cases the derivatives of F are equated to zero and, after some simplification, a numerical solution of the resulting equations is sought. Recent work suggests, however, that these procedures do not always converge [Jöreskog, 1966]. Even when convergence does occur, it is usually very slow. A better method, for which

ultimate convergence is assured, was given by Jöreskog [1966]. Experience with this method has revealed that it is sometimes still difficult to obtain a very accurate solution unless many iterations are performed. Efficient minimization of $F(\Lambda, \Phi, \Psi)$ seems impossible without the use of second-order derivatives.

The present procedure is based on the method of Fletcher and Powell [1963], which was used successfully in the unrestricted maximum likelihood problem [Jöreskog, 1967b]. It is a rapidly converging iterative procedure for minimizing a function of several variables when analytical expressions for the first-order derivatives are available. The efficiency of the method is obtained by the use of a matrix E, which is evaluated in each iteration. Initially, E is any positive definite matrix approximating the inverse of the matrix of second-order derivatives. In subsequent iterations E is improved, using the information built up about the function, so that ultimately E converges to the inverse of the second-order derivative matrix, evaluated at the minimum. If the number of parameters is large, the number of iterations required may still be large, but this can be decreased considerably by the provision of a good starting point and good initial estimates of second-order derivatives. Since the iteration equations have been given in detail by Jöreskog [1967b], they are not repeated here. The function F is considered only in the region where $\psi_{ii} \geq \epsilon$, $i = 1, 2, \cdots, p$, for some arbitrary small positive number ϵ. The treatment of these boundary conditions is the same as is given in the paper by Jöreskog.

The real advantage with the Fletcher and Powell method, as compared to the Newton–Raphson method, is that once an initial estimate of E has been obtained, the successive modifications of E throughout the iterations can be done very rapidly. One variant of the Newton–Raphson method computes the matrix of second-order derivatives and its inverse in each iteration, but this can be very time-consuming, especially when the number of parameters is large. Another variant of the Newton–Raphson method computes the matrix of second-order derivatives and its inverse only once and uses this inverse in all iterations. Such a procedure may require a large number of iterations to converge, especially if the starting point is not close to the solution point. The Fletcher and Powell method is a compromise between these two extremes. It improves the inverse of the matrix of second order derivatives in each iteration at very little cost.

In the unrestricted maximum likelihood problem it is possible to eliminate the parameters in Λ analytically so that the method of Fletcher and Powell is applied to a function of p variables only. Unfortunately, such a two-stage minimization procedure is not possible in this case, except under certain special conditions. The function F has, therefore, to be minimized simultaneously with respect to all free parameters. In a factor analysis of 40 tests and 10 factors, say, the number of free parameters may be almost 400 and, consequently, the matrix E, which must be evaluated in each iteration, is of order 400×400. The handling of such matrices in a computer presents many difficulties, even with present-day computers.

The function F is considered as a function of the free parameters in Λ, the free parameters in the *lower half of* Φ, *including the diagonal*, and the free parameters in the diagonal of Ψ. Expressions for the first-order derivatives of F have been given elsewhere [see e.g., Lawley & Maxwell, 1963; Jöreskog, 1966]. These expressions are

(6) $\partial F/\partial\Lambda = 2\Sigma^{-1}(\Sigma - S)\Sigma^{-1}\Lambda\Phi$

(7) $\partial F/\partial\Phi = c\Lambda'\Sigma^{-1}(\Sigma - S)\Sigma^{-1}\Lambda \left\{\begin{array}{ll} c = 1 & \text{for diagonal elements} \\ c = 2 & \text{for non-diagonal elements}\end{array}\right\}$

(8) $\partial F/\partial\Psi = \text{diag}\,[\Sigma^{-1}(\Sigma - S)\Sigma^{-1}]$,

with the understanding that elements in the matrices on the left that correspond to fixed values of Λ, Φ and Ψ are zero. The above matrices may be simplified for computational purposes by use of the identities

(9) $$\Sigma^{-1} = \Psi^{-1} - \Psi^{-1}\Lambda(I + \Phi\Lambda'\Psi^{-1}\Lambda)^{-1}\Phi\Lambda'\Psi^{-1}$$

(10) $$\Sigma^{-1}\Lambda = \Psi^{-1}\Lambda(I + \Phi\Lambda'\Psi^{-1}\Lambda)^{-1}.$$

The free parameters in Λ, Φ, Ψ are arranged in a vector θ as follows. Let θ_i, $i = 1, 2, \cdots, k$, be a vector containing the free parameters in column i of Λ and let θ_{k+1} and θ_{k+2} be vectors containing the free parameters in Φ and Ψ respectively. Then $\theta' = (\theta_1', \theta_2', \cdots, \theta_{k+2}')$. Formally we may now regard F as a function of θ and write $F(\theta)$. Similarly we may arrange all nonvanishing derivatives as a gradient vector $\partial F/\partial\theta$. If there are q free parameters all together, θ and $\partial F/\partial\theta$ are vectors of order q.

The Fletcher and Powell method requires repeated computation of function values $F(\theta)$ and gradient vectors $\partial F/\partial\theta$. For the iterative procedure to work satisfactorily it is necessary that this be done rapidly and accurately. The following method is used.

1. Compute $\Gamma = \Lambda'\Psi^{-1}\Lambda$ and $I + \Phi\Gamma$. The last matrix is unsymmetric, nonsingular and of order $k \times k$. Invert this to give $A = (I + \Phi\Gamma)^{-1}$. This also gives $|I + \Phi\Gamma|$.
2. Compute
$$|\Sigma| = \left(\prod_{i=1}^{p} \psi_{ii}\right) |I + \Phi\Gamma|$$
3. Compute $B = A\Phi$
4. Compute $C = \Psi^{-1} - \Psi^{-1}\Lambda B\Lambda'\Psi^{-1} = \Sigma^{-1}$
5. Compute $D = SC = S\Sigma^{-1}$ and $\text{tr } D = \text{tr }(S\Sigma^{-1})$

6. Compute F from (4) using $|\Sigma|$ from step 2 and $\text{tr }(S\Sigma^{-1})$ from step 5. (The quantity $\log |S| + p$ is a constant computed before the minimization begins.)
7. Compute $E = C - CD = \Sigma^{-1}(\Sigma - S)\Sigma^{-1}$
8. Compute $G = E\Lambda$
9. Compute $2G\Phi = \partial F/\partial\Lambda$
10. Compute $c\Lambda'G = \partial F/\partial\Phi$
11. Compute $\text{diag } E = \partial F/\partial\Psi$
12. Form the vector $\partial F/\partial\theta$ from the quantities of steps 9–11.

Formulas for large-sample approximations of second-order derivatives of F were derived by Lawley [1967] and independently by Lockhart [1967]. Lawley derived them by differentiation of the elements of the first order derivative matrices ignoring contributions arising from the differentiation of elements of $\Sigma - S$. Another way to derive these formulas is to make use of the covariance structure of a Wishart matrix S [see e.g., Anderson, 1958, Theorem 4.2.4],

(11) $$nE[(s_{gh} - \sigma_{gh})(s_{ij} - \sigma_{ij})] = \sigma_{gi}\sigma_{hj} + \sigma_{gj}\sigma_{hi}.$$

As shown by Lawley [1967], the formulas for second-order derivatives can best be expressed in terms of the elements of the following matrices

(12) $$\xi = \Sigma^{-1}\Lambda$$

(13) $$\eta = \Sigma^{-1}\Lambda\Phi = \xi\Phi$$

(14) $$\alpha = \Lambda'\Sigma^{-1}\Lambda = \Lambda'\xi$$

(15) $$\beta = \Phi\Lambda'\Sigma^{-1}\Lambda = \Phi\alpha$$

(16) $$\gamma = \Phi\Lambda'\Sigma^{-1}\Lambda\Phi = \beta\Phi.$$

The formulas then become

(17) $$E(\partial^2F/\partial\lambda_{ir}\partial\lambda_{js}) = 2(\sigma^{ij}\gamma_{rs} + \eta_{is}\eta_{jr})$$

(18) $$E(\partial^2F/\partial\lambda_{ir}\partial\phi_{st}) = (2 - \delta_{st})(\xi_{is}\beta_{rt} + \xi_{it}\beta_{rs})$$

(19) $$E(\partial^2/\partial\lambda_{ir}\partial\psi_{jj}) = 2\sigma^{ij}\eta_{jr}$$

(20) $$E(\partial^2F/\partial\phi_{rs}\partial\phi_{tu}) = \tfrac{1}{2}(2 - \delta_{rs})(2 - \delta_{tu})(\alpha_{rt}\alpha_{su} + \alpha_{ru}\alpha_{st})$$

(21) $$E(\partial^2F/\partial\phi_{rs}\partial\psi_{ii}) = (2 - \delta_{rs})\xi_{ir}\xi_{is}$$

(22) $$E(\partial^2F/\partial\psi_{ii}\partial\psi_{jj}) = (\sigma^{ij})^2.$$

Here δ_{rs} is Kronecker's delta, which is 1 if $r = s$ and zero otherwise. Except for (18), (20) and (21), these formulas agree with the results obtained by Lockhart who gives these results without the factors involving the δ's.

However, these factors are necessary if some factor is not scaled to unit variance. This will usually be the case whenever there are *fixed non-zero* values in Λ.

As an example of how these formulas are derived from (11), we prove the last formula. The i-th and j-th diagonal elements of $\partial F/\partial \Psi$ are

$$\partial F/\partial \psi_{ii} = \sum_{\alpha=1}^{p}\sum_{\beta=1}^{p} \sigma^{i\alpha}(s_{\alpha\beta} - \sigma_{\alpha\beta})\sigma^{\beta i} \tag{23}$$

$$\partial F/\partial \psi_{jj} = \sum_{u=1}^{p}\sum_{v=1}^{p} \sigma^{ju}(s_{uv} - \sigma_{uv})\sigma^{vj}. \tag{24}$$

Multiplying these and using (11) then gives

$$\begin{aligned} E(\partial^2 F/\partial\psi_{ii}\partial\psi_{jj}) &= (n/2)E(\partial F/\partial\psi_{ii}\,\partial F/\partial\psi_{jj}) \\ &= (\tfrac{1}{2})\sum_{\alpha}\sum_{\beta}\sum_{u}\sum_{v} \sigma^{i\alpha}\sigma^{\beta i}\sigma^{ju}\sigma^{vj} nE[(s_{\alpha\beta} - \sigma_{\alpha\beta})(s_{uv} - \sigma_{uv})] \\ &= \sum_{\alpha}\sum_{\beta}\sum_{u}\sum_{v} \sigma^{i\alpha}\sigma^{\beta i}\sigma^{ju}\sigma^{vj}\sigma_{\alpha u}\sigma_{\beta v} \\ &= (\sigma^{ij})^2, \end{aligned}$$

where the first step follows from a general formula for likelihood functions [see e.g., Kendall & Stuart, 1961, eq. 18.60].

The computational method starts with arbitrary initial estimates of Λ, Φ, Ψ. The better these are, the fewer iterations will be required. If most of the restricted parameters are in Λ and these are sufficient to define Λ uniquely, such initial estimates can be obtained by rotating an unrotated orthogonal factor matrix using some Procrustes method, e.g. that of Lawley and Maxwell [1964]. In our computer program [Jöreskog & Gruvaeus, 1967], we generate initial estimates by a modified centroid method, if they are not provided by the user. From the initial starting point five steepest descent iterations are performed. Steepest descent iterations have been found to be very effective in the beginning but very ineffective in the neighborhood of the minimum. After these five iterations one has usually come so close to the minimum that it is worthwhile to compute the above approximations for second-order derivatives.

When suitably arranged these form a symmetric positive definite matrix G. The inverse matrix G^{-1} then serves as an initial approximation to E, and in subsequent iterations the Fletcher and Powell method is employed and the matrix E is modified accordingly.

When the number of free parameters to be estimated is large so that the order of the matrix G is large, the inversion of G is very difficult and time-consuming. In this case an approximation for G^{-1} which is sufficient to determine an initial estimate of E can be obtained as follows. Let

$$B = \begin{bmatrix} G_{11} & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & G_{kk} & & \\ 0 & \cdots & \cdot & G_{k+1,k+1} & \cdot \\ 0 & \cdots & \cdot & \cdot & G_{k+2,k+2} \end{bmatrix} \tag{25}$$

where

$$G_{ii} = \partial^2 F/\partial\theta_i\partial\theta_i', \qquad i = 1, 2, \cdots, k+2. \tag{26}$$

Then B^{-1} can be taken as an initial estimate of E. This is reasonable because earlier studies on sampling variability have indicated that a factor loading in one column of Λ correlates little with a factor loading in another column, with a factor correlation and with a unique variance. This reduces the problem to that of inverting $k + 2$ small matrices instead of one large matrix. In the Fletcher and Powell iterations the full E matrix is used.

The above method has been programmed in FORTRAN IV and tested out on the IBM 7044 [Jöreskog & Gruvaeus, 1967]. The program performs all computations in memory and is limited to at most 30 variables, 10 factors and 120 free parameters. The program is quite feasible for all the sets of data that it can handle. In computers with larger storage capacities than that of the IBM 7044, the above limits can be increased so that larger data can be handled.

When the minimum of F has been found, the minimizing values of Λ, Φ and Ψ are the maximum likelihood estimates $\hat{\Lambda}$, $\hat{\Phi}$ and $\hat{\Psi}$, and the hypothesis implied by the fixed parameters can be statistically tested. The maximum likelihood estimate of Σ under the hypothesis is

$$\hat{\Sigma} = \hat{\Lambda}\hat{\Phi}\hat{\Lambda}' + \hat{\Psi}, \tag{27}$$

and the likelihood ratio test statistic for testing the hypothesis is

$$n[\log|\hat{\Sigma}| - \log|S| + \mathrm{tr}\,(S\hat{\Sigma}^{-1}) - p]. \tag{28}$$

This is simply n times the minimum value $F(\hat{\Lambda}, \hat{\Phi}, \hat{\Psi})$ of $F(\Lambda, \Phi, \Psi)$. If n is large, this is distributed as χ^2 with

$$d = \tfrac{1}{2}p(p+1) - pk - \tfrac{1}{2}k(k+1) - p + m_\Psi + \sum_{i=1}^{k} \max\{m_i, k\} \tag{29}$$

degrees of freedom, where m_i is the number of independent restrictions on factor i, including restrictions on the ϕ_{ii} and the ϕ_{ij}, and m_Ψ is the number of fixed parameters in Ψ. In some cases the evaluation of the last term of (29) may be ambiguous. For example a restriction on ϕ_{ij} may be counted as a restriction either on factor i or on factor j. The rule to follow is to distribute

TABLE 1

Examples of Various Maximum Likelihood Solutions for Nine Psychological Variables

(Data from Holzinger & Swineford [1939]. An asterisk indicates that the parameter was fixed at this value.)

(a) Unrestricted Orthogonal Solution (m = 6)

1. Visual Perception	0.51	0.49	-0.02				0.50
2. Cubes	0.35	0.36	-0.10				0.74
3. Lozenges	0.52	0.44	-0.09				0.54
4. Paragraph Comprehension	0.81	-0.15	0.28	1.00*			0.24
5. Sentence Completion	0.74	-0.12	0.38	0.00*	1.00*		0.30
6. Word Meaning	0.77	-0.13	0.25	0.00*	0.00*	1.00*	0.32
7. Addition	0.04	0.26	0.74				0.39
8. Counting Dots	0.05	0.58	0.58				0.32
9. Straight-Curved Capitals	0.36	0.54	0.35				0.46

χ^2 = 9.77 with 12 degrees of freedom

P = 0.64

(b) Unrestricted Orthogonal Solution (m = 9)
Three-factor General Triangular Solution

1. Visual Perception	0.39	0.59	0.00*				0.50
2. Cubes	0.24	0.44	-0.07				0.74
3. Lozenges	0.38	0.56	-0.08				0.54
4. Paragraph Comprehension	0.87	0.00*	0.00*	1.00*			0.24
5. Sentence Completion	0.83	-0.01	0.12	0.00*	1.00*		0.30
6. Word Meaning	0.82	0.01	-0.01	0.00*	0.00*	1.00*	0.32
7. Addition	0.22	0.08	0.75				0.39
8. Counting Dots	0.14	0.42	0.70				0.32
9. Straight-Curved Capitals	0.36	0.51	0.39				0.46

χ^2 = 9.77 with 12 degrees of freedom

P = 0.64

(c) Unrestricted Oblique Solution (m = 9)
Reference Variables Solution

1. Visual Perception	0.71	0.00*	0.00*				0.50
2. Cubes	0.54	-0.03	-0.08				0.74
3. Lozenges	0.67	0.04	-0.09				0.54
4. Paragraph Comprehension	0.00*	0.87	0.00*	1.00*			0.24
5. Sentence Completion	-0.03	0.81	0.13	0.54	1.00*		0.30
6. Word Meaning	0.01	0.82	-0.01	0.24	0.28	1.00*	0.32
7. Addition	0.00*	0.00*	0.78				0.39
8. Counting Dots	0.42	-0.30	0.73				0.32
9. Straight-Curved Capitals	0.56	-0.06	0.41				0.46

χ^2 = 9.77 with 12 degrees of freedom

P = 0.64

(d) Restricted Oblique Solution (m = 21)
Independent Cluster Solution

1. Visual Perception	0.68	0.00*	0.00*				0.54
2. Cubes	0.52	0.00*	0.00*				0.73
3. Lozenges	0.69	0.00*	0.00*				0.52
4. Paragraph Comprehension	0.00*	0.87	0.00*	1.00*			0.25
5. Sentence Completion	0.00*	0.83	0.00*	0.54	1.00*		0.31
6. Word Meaning	0.00*	0.83	0.00*	0.52	0.34	1.00*	0.31
7. Addition	0.00*	0.00*	0.66				0.57
8. Counting Dots	0.00*	0.00*	0.80				0.37
9. Straight-Curved Capitals	0.00*	0.00*	0.70				0.51

χ^2 = 51.19 with 24 degrees of freedom

P = 0.00

such ambiguous restrictions so as to make the last term in (29) minimum. Examples of this rule are given in the next section.

The final matrix E that has been built up during the iterations is an approximation to the inverse of the matrix of second-order derivatives at the minimum. When multiplied by $(2/n)$ this gives an estimate of the variance-covariance matrix of the maximum likelihood estimates of the free parameters. Substantial experience with the method, however, has revealed that this estimate is not sufficiently accurate for most purposes. This is due to the fact that the matrix E is built up assuming the function to be exactly quadratic [see Fletcher & Powell, 1963]. Our function F is not quadratic, however, but is approximately so in a small region around the minimum. It should be noted that the matrix E is used essentially for the purpose of obtaining fast convergence and not of providing an estimate of the information matrix. Convergence is usually obtained in less than q iterations, often much less. However, at least q iterations in a quadratic region are necessary to build up numerically an accurate estimate of the inverse of the second-order derivative matrix. To obtain an accurate estimate of the variance-covariance matrix of the estimated parameters it is best, therefore, to recompute the second-order derivative matrix at the minimum and invert this. Denoting the matrix so obtained by E^*, only the diagonal elements of this, corresponding to variances of estimates, would normally be of interest. For any parameter θ_i with corresponding maximum likelihood estimate $\hat{\theta}_i$ and variance of estimate e_{ii}^*, an approximate 95% confidence interval is

$$\hat{\theta}_i - 2\sqrt{(2/n)e_{ii}^*} < \theta_i < \hat{\theta}_i + 2\sqrt{(2/n)e_{ii}^*}\,. \tag{30}$$

This formula should be used only when the restrictions are such that the solution is unique.

4. *An Analysis of Nine Mental Ability Tests*

We shall illustrate the preceding ideas and methods and indicate some possible uses of the approach on the basis of two sets of empirical data. The analysis of the first data is reported in this section and that of the second in the next section.

The first set of data consists of nine mental ability tests selected from a battery of 26 tests previously analysed by Holzinger and Swineford [1939]. The nine tests are listed in Table 1. They can be thought of as measuring essentially visualization (tests 1, 2, 3), verbal intelligence (tests 4, 5, 6) and speed (tests 7, 8, 9). A full description of the tests is given in the above reference, where miscellaneous descriptive statistics are given also. Data were obtained on seventh- and eighth-grade children from two different schools. Only the data from the Grant–White school sample of 145 children are used here. The correlations were computed directly from the raw scores given by Holzinger and Swineford.

TABLE 1 (Contd)

(e) Restricted Mixed Solution (m = 20)

1. Visual Perception	0.73	0.00*	0.00*				0.47
2. Cubes	0.50	0.00*	0.00*				0.76
3. Lozenges	0.67	0.00*	0.00*				0.56
4. Paragraph Comprehension	0.00*	0.86	0.00*	1.00*			0.25
5. Sentence Completion	0.00*	0.82	0.00*	0.49	1.00*		0.31
6. Word Meaning	0.00*	0.82	0.00*	0.00*	0.24	1.00*	0.32
7. Addition	0.00*	0.00*	0.84				0.30
8. Counting Dots	0.29	0.00*	0.67				0.44
9. Straight-Curved Capitals	0.56	0.00*	0.45				0.44

χ^2 = 26.47 with 23 degrees of freedom

P = 0.28

(f) Unrestricted Orthogonal Solution (m = 16)
Four-factor General Triangular Solution

1. Visual Perception	0.38	0.58	0.00*	0.00*					0.52
2. Cubes	0.24	0.41	0.35	0.00*					0.65
3. Lozenges	0.38	0.53	0.30	-0.03					0.48
4. Paragraph Comprehension	0.87	0.00*	0.03	0.00*	1.00*				0.25
5. Sentence Completion	0.83	0.01	-0.13	0.06	0.00*	1.00*			0.28
6. Word Meaning	0.83	0.01	0.04	-0.02	0.00*	0.00*	1.00*		0.32
7. Addition	0.24	0.02	0.00*	0.95	0.00*	0.00*	0.00*	1.00*	0.05
8. Counting Dots	0.15	0.43	-0.13	0.57					0.44
9. Straight-Curved Capitals	0.36	0.59	-0.22	0.34					0.36

χ^2 = 2.75 with 6 degrees of freedom

P = 0.84

(g) Restricted Orthogonal Solution (m = 16)

1. Visual Perception	0.38	0.62	0.00*				0.47
2. Cubes	0.24	0.42	0.00*				0.77
3. Lozenges	0.38	0.53	0.00*				0.57
4. Paragraph Comprehension	0.87	0.00*	0.00*	1.00*			0.24
5. Sentence Completion	0.83	0.00*	0.00*	0.00*	1.00*		0.32
6. Word Meaning	0.82	0.00*	0.00*	0.00*	0.00*	1.00*	0.32
7. Addition	0.25	0.00*	0.78				0.32
8. Counting Dots	0.16	0.37	0.70				0.36
9. Straight-Curved Capitals	0.37	0.49	0.42				0.45

χ^2 = 13.81 with 19 degrees of freedom

P = 0.80

Tables 1a–g show seven different maximum likelihood solutions for these data obtained under different identification conditions imposed on the factors. The fixed values of the parameters are marked with asterisks, and the number of fixed parameters is denoted m and is listed above each table. The value of χ^2 and the degrees of freedom as computed by (28) and (29) respectively are given below each solution. The probability level P of the χ^2 value is also given. This is defined as the probability of getting a χ^2 value larger than the value actually obtained, given that the hypothesized pattern is true. Thus, small values of P correspond to poor fit and large values to good fit.

In Table 1a we have postulated that the factors be uncorrelated and have unit variances. The latter, of course, is just an arbitrary scaling of the factors and will be used throughout. This gives $m = 6$ fixed parameters. Since $k = 3$, we should normally require $m = k^2 = 9$ independent restrictions to obtain a unique solution. The solution of Table 1a is therefore not unique but is just an arbitrary orthogonal solution. It happens to be the first one that the computer program found. It can be rotated orthogonally or obliquely to any other unrestricted maximum likelihood solution for the same data. To obtain an orthogonal solution, for example, we postmultiply the factor matrix by an orthogonal matrix of order 3 × 3. Since this matrix has three independent elements, it can be chosen to satisfy three independent restrictions on the factor loadings.

A particular set of such restrictions is used in Table 1b. It is seen that three factor loadings have been postulated to be zero. Test 1 (Visual Perception) is postulated not to load on factor 3, and test 4 (Paragraph Comprehension) is postulated to load only on factor 1, the idea being that factor 3 should be a speed factor, factor 2 a visualization factor and factor 1 a general factor. The solution of Table 1b represents an unrestricted orthogonal solution with nine fixed parameters. Given these fixed values, the solution is unique. It can be verified readily that the conditions for uniqueness, given in Section 2, are satisfied. The rows of the factor matrix can be permuted so that the zeros appear in the upper right triangle, hence the term general triangular solution. Table 1b serves to illustrate one important point, namely, that one has nothing to lose but may have much to gain by fixing the three zero loadings. By choosing these zero loadings appropriately, the solution will be directly interpretable, thus making subsequent rotations unnecessary. However, if further rotation is still found to be preferable, the solution can be rotated orthogonally or obliquely by any of the available methods. If the zero loadings are not too unreasonable, the rotation can probably be done by hand. We shall return to the interpretation of the solution of Table 1b after examining some other solutions.

Another unrestricted solution is shown in Table 1c. Here the factors are permitted to be correlated. To make the solution uniquely determined, we have chosen three reference variables, tests 1, 4 and 7, to represent the factors. These reference tests are postulated to be pure in their respective factors. Thus Table 1c represents an unrestricted oblique unique solution with $m = 9$ fixed elements. That the solution is unrestricted and unique is readily verified by showing that there is a unique transformation from the solution of Table 1a to that of Table 1c. Reference variables solutions are particularly useful since they impose just enough restrictions to make the solution unique, and if the reference tests are carefully chosen to tap the factors of interest, the solution will be directly interpretable in most cases. The solution in Table 1c exhibits a fairly clear simple structure. It is seen that the first seven tests are all loaded in one factor only, whereas the last two tests are more complex, being loaded not only in the speed factor but also in the visualization factor. One interpretation of this might be that the material in tests 8 and 9 consists of configurations in the plane in a way similar to tests 1 and 3. That test 9, in particular, should involve some visualization factor is evident from the fact that it is necessary here to break up a letter into straight and curved lines. Test 8, on the other hand, has its highest loading on the speed factor.

The solutions of Tables 1a, 1b and 1c are three different unrestricted solutions for the same data. This is reflected in two ways in the tables: the unique variances are the same, and the χ^2 values and the corresponding degrees of freedom are also the same for the three solutions. The probability level of 0.64, in this case, merely indicates that three factors are sufficient to account for the intercorrelations between the tests.

Let us now assume that it has been hypothesized in advance that the tests form three independent clusters so that tests 1, 2 and 3 are loaded in the visualization factor only, tests 4, 5 and 6 in the verbal factor only and tests 7, 8 and 9 in the speed factor only. This leads to the solution of Table 1d, where the free parameters have been estimated by the maximum likelihood method. This is a restricted oblique solution with $m = 21$ fixed parameters. Only nine of these restrictions are necessary to make the solution unique; the additional 12 restrictions affect the whole factor space. This is demonstrated by the differences in unique variances between Tables 1c and 1d and also by the large increase in χ^2. The χ^2 value is 51.19 which, with 24 degrees of freedom, is highly significant. The hypothesized factor structure must therefore be rejected as being untenable.

The question now arises as to what is causing the poor fit of the solution in Table 1d. In general, poor fit may be due to the fact that either the number of factors is untenable or the hypothesized structure is untenable, or both. In this case, we know from any one of the solutions of Tables 1a, 1b and 1c that three factors are tenable, so the poor fit must be due to a too restrictive structure. Which of the hypothesized zero loadings are then untenable? There are several ways to find this out. One way is to examine the residual

TABLE 2

Approximate 95% Confidence Intervals for the Parameters in the Solution of Table 1b

0.71 ± 0.17	0.00*	0.00*				0.50 ± 0.18
0.54 ± 0.25	-0.03 ± 0.14	-0.08 ± 0.19				0.74 ± 0.20
0.67 ± 0.24	0.04 ± 0.19	-0.09 ± 0.23				0.54 ± 0.19
0.00*	0.87 ± 0.18	0.00*	1.00*			0.24 ± 0.10
-0.03 ± 0.21	0.81 ± 0.15	0.13 ± 0.22	0.54 ± 0.22	1.00*		0.30 ± 0.11
0.01 ± 0.25	0.82 ± 0.18	-0.01 ± 0.22	0.24 ± 0.30	0.28 ± 0.23	1.00*	0.32 ± 0.11
0.00*	0.00*	0.78 ± 0.23				0.39 ± 0.23
0.42 ± 0.24	-0.30 ± 0.18	0.73 ± 0.22				0.32 ± 0.19
0.56 ± 0.22	-0.06 ± 0.15	0.41 ± 0.19				0.46 ± 0.14

correlations obtained after the three factors have been removed. A better, more direct, way is to compute approximate confidence intervals for all the free parameters of the solution of Table 1c. These confidence intervals are given in Table 2. It is seen that several of the loadings that were set equal to zero in Table 1d are indeed not significantly different from zero but a few of them are. Relaxing two of the zeros in the first factor of Table 1d and adding the restriction $\phi_{13} = 0$ yields the solution of Table 1e. Since one of the factor correlations is postulated to be zero, whereas the other factor correlations are free, this represents a restricted mixed solution with $m = 20$ fixed parameters. It should be noted that we have relaxed two of the zero factor loadings but have added a zero factor correlation. Thus the number of fixed parameters is only one less than before. The value of χ^2 is now 26.47 with 23 degrees of freedom. This has a probability level of 0.28. The solution is therefore acceptable, and it has also been verified that the values of all the free parameters are statistically significantly different from zero. The interpretation of the solution is very similar to that of Table 1c. In terms of the three factors, visualization, verbal and speed, the first seven tests are pure tests, whereas the last two tests are more complex. The loadings of these on the visualization factor have already been commented upon. It should be noted that factors 1 and 3 are uncorrelated but both of them correlated with factor 2. This suggests that the verbal factor is more general than the other two factors and that one might try to split this factor up into a general factor which influences all the tests and a more specific factor which is associated only with tests 4, 5 and 6.

The solution of Table 1f represents the results of such an attempt. It was postulated that there should be an orthogonal solution with a general factor and three group factors, the first one not loaded in test 4, the second one not loaded in tests 1 and 7 and the third one not loaded in tests 1, 2 and 4. It is seen that the visualization and speed factors appear as before, but the

TABLE 3

An Analysis of Two Batteries of Tests into Interbattery and Battery Specific Factors

(Original data from Thurstone & Thurstone [1941]. An asterisk indicates that the parameter was fixed at this value.)

(a) Restricted Orthogonal Solution (m = 19)

1. Prefixes	0.70	-0.12	0.15	0.00*					0.48
2. Suffixes	0.74	-0.08	0.22	0.00*					0.41
3. Vocabulary	0.39	0.81	0.33	0.00*					0.09
4. Sentences	0.37	0.75	0.08	0.00*	1.00*				0.31
5. First and Last Letters	0.65	-0.03	0.00*	0.37	0.00*	1.00*			0.44
6. First Letters	0.72	-0.05	0.00*	0.15	0.00*	0.00*	1.00*		0.46
7. Four-letter Words	0.60	0.09	0.00*	0.35	0.00*	0.00*	0.00*	1.00*	0.52
8. Completion	0.51	0.65	0.00*	0.02					0.32
9. Same or Opposite	0.48	0.67	0.00*	-0.12					0.32

χ^2 = 6.73 with 10 degrees of freedom

P = 0.75

(b) Restricted Mixed Solution (m = 20)

1. Prefixes	0.73	-0.08	0.15	0.00*					0.48
2. Suffixes	0.75	-0.04	0.22	0.00*					0.41
3. Vocabulary	0.00*	0.90	0.33	0.00*					0.09
4. Sentences	0.00	0.83	0.08	0.00*	1.00*				0.31
5. First and Last Letters	0.64	0.01	0.00*	0.37	0.38	1.00*			0.44
6. First Letters	0.72	0.00*	0.00*	0.15	0.00*	0.00*	1.00*		0.46
7. Four-letter Words	0.54	0.14	0.00*	0.35	0.00*	0.00*	0.00*	1.00*	0.52
8. Completion	0.19	0.74	0.00*	0.02					0.32
9. Same or Opposite	0.15	0.75	0.00*	-0.12					0.32

χ^2 = 6.73 with 10 degrees of freedom

P = 0.75

(c) Restricted Mixed Solution (m = 24)

1. Prefixes	0.67	0.00*	0.21	0.00*					0.51
2. Suffixes	0.71	0.00*	0.39	0.00*					0.35
3. Vocabulary	0.00*	0.91	0.17	0.00*					0.15
4. Sentences	0.00*	0.84	0.03	0.00*	1.00*				0.29
5. First and Last Letters	0.68	0.00*	0.00*	0.33	0.35	1.00*			0.43
6. First Letters	0.73	0.00*	0.00*	0.05	0.00*	0.00*	1.00*		0.44
7. Four-letter words	0.56	0.14	0.00*	0.29	0.00*	0.00*	0.00*	1.00*	0.52
8. Completion	0.20	0.73	0.00*	-0.01					0.32
9. Same or opposite	0.14	0.75	0.00*	-0.13					0.33

χ^2 = 9.43 with 14 degrees of freedom

P = 0.80

(d) Restricted Mixed Solution (m = 27)

1. Prefixes	0.70	0.00*	-0.15	0.00*					0.49
2. Suffixes	0.75	0.00*	-0.20	0.00*					0.39
3. Vocabulary	0.00*	0.89	0.12	0.00*					0.20
4. Sentences	0.00*	0.82	0.36	0.00*	1.00*				0.20
5. First and Last Letters	0.63	0.00*	0.00*	0.34	0.47	1.00*			0.48
6. First Letters	0.68	0.00*	0.00*	0.26	0.00*	0.00*	1.00*		0.46
7. Four-letter words	0.58	0.00*	0.00*	0.35	0.00*	0.00*	0.00*	1.00*	0.53
8. Completion	0.00*	0.81	0.00*	0.25					0.27
9. Same or opposite	0.00*	0.81	0.00*	0.06					0.33

χ^2 = 34.26 with 17 degrees of freedom

P = 0.01

attempt to isolate one general and one verbal factor was not successful. Tests 4, 5 and 6 load only in the general factor, and the second group factor is a very weak one having no psychological interpretation. This solution, therefore, is an example of overfactoring. If the second group factor is omitted and zero loadings postulated to identify the other two group factors as before, we arrive at the unrestricted orthogonal three-factor general triangular solution as in Table 1b. In this solution, all the small loadings are not significant. If these are set to zero, we get the restricted orthogonal solution of Table 1g.

The solutions of Tables 1e and 1g are two alternative final solutions for these data. Both solutions are final in the sense that all nonzero parameters are significant. The solution of Table 1e is slightly more restrictive than that of Table 1g, but the latter has a much better fit. The choice between the solutions, of course, is a matter of psychological theory [e.g., cf. Thurstone, 1947; Vernon, 1951].

The above examples show how the procedure of Section 3 can be used in an exploratory way to determine a solution that is reasonable from the point of view of both goodness of fit and psychological interpretation.

5. *Interbattery Analysis*

The second example serves to illustrate how our method can be used to analyse two batteries of tests into interbattery factors and battery specific factors. Tucker [1958] developed a method for determining factors that is common to the two batteries. These factors, called interbattery factors, account for the correlations between batteries but may not account for correlations within batteries. This example shows how factors that are specific to each battery can also be determined.

We shall use the same data as were used by Tucker. The data consist of nine tests from Thurstone and Thurstone [1941] listed in Tables 3a–d. The first four tests constitute battery 1 and the last five battery 2. Within and between correlations are given in Tucker's Table 2. The correlations are based on a sample of 710. Tucker found two interbattery factors that account for the correlations between batteries. This was also supported by a statistical test. However, two factors do not adequately account for all the correlations in the whole 9×9 correlation matrix. We checked this by performing an unrestricted maximum likelihood analysis with two factors. This gave χ^2 = 50.10 with 19 degrees of freedom which is highly significant. The hypothesis of only two common factors in the two batteries must therefore be rejected. This suggests that there are factors specific to each battery but common to two or more tests within the battery. It is therefore postulated that there are four factors in the two batteries, that the first two are interbattery factors, that the third is specific to battery 1 and that the fourth is specific to battery 2. The two battery specific factors are postulated to be uncorrelated and uncorrelated with the two interbattery factors. Otherwise they would contribute to correlations between tests and hence not be battery specific. In addition we arbitrarily set the correlation between factors 1 and 2 equal to zero. The maximum likelihood solution under the above restrictions is given in Table 3a. This solution has acceptable fit, thus confirming the hypothesis that correlations between and within the two batteries can be accounted for by four factors with the above structure. It should be noted that this solution, regarded as a solution for all the nine tests, is a restricted orthogonal nonunique solution. It is restricted because too many restrictions have been imposed on factors 3 and 4, and it is not unique because too few restrictions have been imposed on factors 1 and 2. The latter can be rotated, orthogonally or obliquely, without changing the former. For example, such a rotation can be done to give two correlated interbattery factors with one zero loading for each factor. An example of this kind of solution is given in Table 3b, where test 3 (vocabulary) and test 6 (first letters) have been used as reference variables. A refinement of this solution, obtained by setting to zero the small loadings for the interbattery factors, is given in Table 3c. This imposes additional restrictions, and, as a consequence, factor loadings for the battery specific factors and the unique variances are changed. The solution of Table 3c fits the data well, and the two interbattery factors can be interpreted as a word-fluency and a verbal factor. Considering these two factors only, the first six tests are loaded on only one of them, whereas the last three are loaded on both. The three loadings $\lambda_{72} = 0.14$, $\lambda_{81} = 0.20$ and $\lambda_{91} = 0.14$, though small, appear to be significant. If all three are postulated to be zero, the solution of Table 3d is obtained.

The interbattery factors of Table 3d agree fairly well with the rotated solution obtained by Tucker [1958], but there are some differences for some of the loadings. Although battery specific factors may not be of direct interest, this example shows that they may be important in determining the interbattery factors.

6. *The Question of Goodness of Fit*

It was stated in Section 3 that n times the minimum value of F can be used as a large sample χ^2 statistic to test the hypothesis that the population variance-covariance matrix Σ is of the form (2) with specified values for certain parameters in Λ, Φ, Ψ. Such a hypothesis may be quite unrealistic in most empirical work with test data. If a sufficiently large sample were obtained this χ^2 statistic would, no doubt, indicate that any such non-trivial hypothesis is statistically untenable. The hypothesis of the experimenter rather is that (2) represents the variance-covariance matrix of the major factors that the experimenter is interested in, but that there are also a lot of minor factors which influence the test scores and which the experimenter has little or no control over [Tucker *et al.*, 1968]. These minor factors cause the lack of agreement between the formal mathematical model (2) and the

variance-covariance matrix of the entire population. From this point of view the statistical problem is not one of testing a given hypothesis but rather one of fitting models with different numbers of parameters and of deciding when to stop fitting. The meaning and use of χ^2 in such problems are as follows. If a value of χ^2 is obtained, which is large compared to the number of degrees of freedom, this is an indication that more information can be extracted from the data. One may then try to relax the model somewhat by introducing more parameters. This can be done by relaxing some restrictions on the common factor space or by introducing additional factors or both. If, on the other hand, a value of χ^2 is obtained which is close to the number of degrees of freedom, this is an indication that the model "fits too well". Such a model is not likely to remain stable in future samples and all parameters may not have real meaning. When to stop fitting additional parameters cannot be decided on a purely statistical basis. This is largely a matter of the experimenter's interpretations of the data based on substantive theoretical and conceptual considerations. Ultimately the criteria for goodness of the model depends on the usefulness of it and the results it produces.

Examining the solutions of Table 1 from this point of view, we may say that for these data the solutions of Tables 1a, b, c, f, g are examples of overfitting whereas the solution of Table 1d is too restrictive. The solution of Table 1e represents a reasonable compromise. For the other data, the solution of Table 3d is the most reasonable, the other three solutions being fitted too well.

REFERENCES

Anderson, T. W. *An introduction to multivariate statistical analysis.* New York: Wiley, 1958.

Anderson, T. W., & Rubin, H. Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability,* Vol. V. Berkeley: University of California Press, 1956. pp. 111–150.

Fletcher, R., & Powell, M. J. D. A rapidly convergent descent method for minimization. *Computer Journal,* 1963, **2**, 163–168.

Holzinger, K. J., & Swineford, F. *A study in factor analysis: The stability of a bi-factor solution.* University of Chicago: Supplementary Educational Monographs, No. 48, 1939.

Howe, W. G. Some contributions to factor analysis. Report No. ORNL-1919, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 1955.

Jöreskog, K. G. Testing a simple structure hypothesis in factor analysis. *Psychometrika,* 1966, **31**, 165–178.

Jöreskog, K. G. UMLFA - A computer program for unrestricted maximum likelihood factor analysis. Research Memorandum 66-20. Princeton, N. J.: Educational Testing Service. Revised Edition, 1967. (a)

Jöreskog, K. G. Some contributions to maximum likelihood factor analysis. *Psychometrika,* 1967, **32**, 443–482. (b)

Jöreskog, K. G., & Gruvaeus, G. RMLFA - A computer program for restricted maximum likelihood factor analysis. Research Memorandum 67-21. Princeton, N. J.: Educational Testing Service, 1967.

Jöreskog, K. G., & Lawley, D. N. New methods in maximum likelihood factor analysis. *British Journal of Mathematical and Statistical Psychology,* 1968, **21**, in press.

Kendall, M. G., & Stuart, A. *The advanced theory of statistics*, vol. 2. London: Charles Griffin and Co., 1961.
Lawley, D. N. Estimation in factor analysis under various initial assumptions. *British Journal of Statistical Psychology*, 1958, **11**, 1–12.
Lawley, D. N. Some new results in maximum likelihood factor analysis. *Proceedings of the Royal Society of Edinburgh*, Section A, 1967, **67**, 256–264.
Lawley, D. N., & Maxwell, A. E. *Factor analysis as a statistical method*. London: Butterworth, 1963.
Lawley, D. N., & Maxwell, A. E. Factor transformation methods. *British Journal of Statistical Psychology*, 1964, **17**, 97–103.
Lockhart, R. S. Asymptotic sampling variances for factor analytic models identified by specified zero parameters. *Psychometrika*, 1967, **32**, 265–277.
Reiersøl, O. On the identifiability of parameters in Thurstone's multiple factor analysis. *Psychometrika*, 1950, **15**, 121–149.
Thurstone, L. L. *Multiple-factor analysis*. Chicago: University of Chicago Press, 1947.
Thurstone, L. L., & Thurstone, T. G. Factorial studies of intelligence. *Psychometric Monographs No. 2*, 1941. Chicago: University of Chicago Press.
Tucker, L. R. An inter-battery method of factor analysis. *Psychometrika*, 1958, **23**, 111–136.
Tucker, L. R, Koopman, R. F., & Linn, R. L. Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, 1968, in press.
Vernon, P. E. *The structure of human abilities*. London: Methuen, 1951.

Reprinted from: *Psychometrika*, June 1969, Vol. 34, No. 2.

Author's Addendum
February, 1979

In this article, written in 1967, I quoted Howe (1955) and gave two conditions for the uniqueness, under factor rotation, of a factor matrix Λ containing specified fixed elements. Howe (1955) gave these conditions for the case of fixed *zero* elements, but I was assuming that they would be valid also for the case of nonzero fixed elements. However, this has been shown to be incorrect by Jennrich (1978). Also, in trying to formulate the conditions for orthogonal and oblique solutions at the same time, I made a mistake, so that Howe's formulation in the oblique case with fixed zero elements is incorrectly stated in my article. To clarify the issues it seems best to consider the four cases separately: (i) orthogonal solution with fixed zero elements, (ii) orthogonal solution with arbitrary fixed elements, (iii) oblique solution with fixed zero elements, (iv) oblique solution with arbitrary fixed elements.

Case (i) was considered by Dunn (1973) who gave a counterexample to show that the original condition, given by Howe (1955) and correctly quoted by me, is not sufficient. He also stated and proved a substitute condition for sufficiency. In Dunn's counterexample there are two columns of Λ with fixed zeroes in the same rows. Such a Λ cannot be unique, since it can be rotated orthogonally in the plane of these two columns without affecting the fixed zero elements.

Case (ii) was considered by Jennrich (1978) who gave an example of two orthogonally equivalent Λ-matrices with $k(k-1)/2$ fixed nonzero elements. These two Λ-matrices have different elements in the nonfixed positions and hence are transparently different. In case (i) Dunn's conditions are sufficient for uniqueness up to column sign changes. Thus the specification of fixed zero elements does not lead to uniqueness but to a transparent form of uniqueness (column sign changes). However, when the specified values are not necessarily zero, one may be led to much less transparent forms of nonuniqueness. In both cases the specification of $k(k-1)/2$ loadings reduces the indeterminacy of Λ considerably, from an infinite number of solutions to 2^k solutions. This suggests that while one may not obtain unique solutions using $k(k-1)/2$ specified values, one will probably obtain solutions that are at least locally unique; indeed, one usually sees computer confirmation of this in the form of a positive definite information matrix.

Case (iii) is by far the most interesting case in practice, and I shall therefore restate and prove the original sufficiency conditions given by Howe (1955).

(α) Let Φ be a symmetric positive definite matrix with diag $\Phi = \mathbf{I}$.

(β) Let Λ have at least $k-1$ fixed zeroes in each column.

(γ) Let Λ_s have rank $k-1$, where Λ_s, $s = 1, 2, \ldots, k$, is the submatrix of Λ, consisting of the rows of Λ which have fixed zero elements in the sth column.

Then conditions α, β, and γ are sufficient for uniqueness of Λ.

The fixed unities in the diagonal of Φ merely fix the unit of measurement of the factors. An alternative way of doing this is to fix one nonzero value in each column of Λ instead. Conditions α and β are therefore equivalent to

(δ) Λ has at least $k-1$ fixed zeroes in each column and one fixed nonzero value in each column, the latter values being in different rows.

I shall prove that conditions δ and γ are sufficient to define Λ uniquely. Let $\mathbf{B} = \Lambda\mathbf{T}$, where $\mathbf{T}$ is an arbitrary nonsingular matrix of order $k \times k$. I shall prove that if $\mathbf{B}$ has the same fixed elements as Λ and if conditions δ and γ hold, then $\mathbf{T}$ must be an identity matrix. Let Λ_s^* be the submatrix of Λ consisting of the rows of Λ that have fixed elements (including the fixed nonzero value) in the sth column. This is of the order $m_s + 1$ by k, where m_s is the number of fixed zeroes in column s. With a suitable permutation of the rows of Λ, Λ_s^* will be of the form

$$\Lambda_s^* = \begin{bmatrix} a & \gamma' \\ 0 & \\ 0 & \\ \cdot & \Lambda_{(s)} \\ \cdot & \\ \cdot & \\ 0 & \end{bmatrix}$$

where a is the fixed nonzero value in column s, λ' is a row vector of the remaining $k - 1$ values in the same row as a and $\Lambda_{(s)}$ of order m_s by $k - 1$ is the matrix Λ_s with the zero column omitted. Since **B** has the same fixed values as Λ, we must have

$$\Lambda_s^* \mathbf{t}_s = \begin{bmatrix} a \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \quad (1.41)$$

where $\mathbf{t}_s = (\mathbf{t}_{1s}' \ \mathbf{t}_{2s}')'$ is the sth column of **T**. Equation (1) is equivalent to

$$a t_{1s} + \lambda' \mathbf{t}_{2s} = a \quad (1.42)$$

$$\Lambda_{(s)} \mathbf{t}_{2s} = 0 \quad (1.43)$$

Since $m_s \geqslant k - 1$ and Λ_s has rank $k - 1$, the omission of the zero column of Λ_s will not change the rank. So the rank of $\Lambda_{(s)}$ is also $k - 1$. Therefore the only solution for $\mathbf{t}_{2s}$ satisfying (2a) is $\mathbf{t}_{2s} = \mathbf{0}$. With $\mathbf{t}_{2s} = \mathbf{0}$ and $a \neq 0$, (2a) implies that $\mathbf{t}_{1s} = 1$. Hence $\mathbf{t}_s$ is equal to a column of the identity matrix. The condition that the fixed nonzero values in Λ are in different rows will guarantee that a different column $\mathbf{t}_s$ will be obtained for $s = 1, 2, \ldots, k$. Hence $\mathbf{T} = \mathbf{I}$.

It is obvious that the same conclusion does not follow if the fixed zeroes in (1) are replaced by fixed nonzero values. In view of the results in case (ii)

it is clear that much more research need to be done in case (iv) in order to clarify the issues and resolve the identification problem.

References

Dunn, J. E. 1973. A note on a sufficiency condition for uniqueness of a restricted factor matrix. *Psychometrika* 38: 141-143.

Howe, W. G. 1955. Some contributions to factor analysis. Report No. ORNL-1919 (Oak Ridge, Tenn.: Oak Ridge National Laboratory).

Jennrich, R, I. 1978. Rotational equivalence of factor loading matrices with specified values. *Psychometrika* 43: 421-426.