# Full-Information Item Factor Analysis

**R. Darrell Bock**
**University of Chicago**

**Robert Gibbons**
**University of Illinois**

**Eiji Muraki**
**National Opinion Research Center**

A method of item factor analysis based on Thurstone's multiple-factor model and implemented by marginal maximum likelihood estimation and the EM algorithm is described. Statistical significance of successive factors added to the model is tested by the likelihood ratio criterion. Provisions for effects of guessing on multiple-choice items, and for omitted and not-reached items, are included. Bayes constraints on the factor loadings are found to be necessary to suppress Heywood cases. Numerous applications to simulated and real data are presented to substantiate the accuracy and practical utility of the method.

*Index terms: Armed Services Vocational Aptitude Battery, Beta prior,* EM *algorithm, Item factor analysis,* TESTFACT, *Tetrachoric correlation.*

Strictly speaking, any test reported in a single score should consist of items drawn from a one-dimensional universe. Only then is it a matter of indifference which items are presented to the examinee. This exchangeability of items is especially important in adaptive testing, where different examinees are presented different items.

Of the various methods that have been proposed for investigating the dimensionality of item sets, the most sensitive and informative is item factor analysis. It alone is capable of analyzing relatively large numbers of items jointly and symmetrically, and assigning items to particular dimensions when multiple factors are found. It can also reveal patterns of common item content and format that may have interesting cognitive interpretations.

Unfortunately, the methods of item factor analysis heretofore available have not been entirely satisfactory. Although conventional multiple-factor analysis of the matrix of $\phi$ coefficients is straightforward computationally, it is known to introduce spurious factors when the item difficulties are not uniform. This problem is alleviated by using tetrachoric correlations in place of $\phi$ coefficients, but this strategy also encounters difficulties. The matrix of sample tetrachoric correlation coefficients is almost never positive definite, so the common factor model does not strictly apply. Although present methods of calculating tetrachoric coefficients are fast and generally accurate (Divgi, 1979), they become unstable as the values approach $+1$ or $-1$. When an observed frequency in the four-fold table for a pair of items

*APPLIED PSYCHOLOGICAL MEASUREMENT*
*Vol. 12, No. 3, September 1988, pp. 261–280*

*0146-6216/88/030261-20$2.25*

is 0, the absolute value of an element in the item correlation matrix becomes 1, thus producing a Heywood case. These problems are exacerbated when the coefficients are corrected for guessing (Carroll, 1945).

The limitations of item factor analysis based on tetrachoric correlation coefficients have been overcome to a considerable extent by the generalized least squares (GLS) method (Cristoffersson, 1975; Muthén, 1978). Because this method allows for the large sample variance of the estimated coefficients, instabilities at the extremes are less of a problem. The GLS method requires, however, the generating and inverting of the asymptotic covariance matrix of the estimated tetrachoric coefficients; the computational burden thus becomes extremely heavy as the number of items increases. At present, its practical upper limit is about 20 items (Muthén, 1984).

It is of some interest, therefore, that Bock and Aitkin (1981) have introduced a method of item factor analysis, based directly on item response theory, that does not require calculation of inter-item correlation coefficients and is not strongly limited by the number of items. Although the computations in their method increase exponentially with the number of factors, they increase only linearly with the number of items. The practical limit of the number of factors is five, which is sufficient for most item analysis applications, while 60 to 100 items is not excessive for a fast computer.

Because the Bock-Aitkin approach uses as data the frequencies of all distinct item response vectors, it is called "full-information" item factor analysis (Bartholomew, 1980). It contrasts with the limited information methods of Cristoffersson and Muthén based on low-order joint occurrence frequencies of the item scores. The purpose of this paper is to present in more detail the derivation of full-information factor analysis, to discuss technical problems of its implementation, and to describe experience with the procedure in a number of simulated and real datasets.

## Derivation and Statistical Methods

Bock and Aitkin (1981) applied Thurstone's multiple-factor model to item response data by assuming that the *m*-factor model,

$$y_{ij} = \alpha_{j1}\theta_{1i} + \alpha_{j2}\theta_{2i} + \ldots + \alpha_{jm}\theta_{mi} + \varepsilon_{ij} \quad , \tag{1}$$

describes not a manifest variable, but an unobservable "response process." The process generates a correct response of person $i$ to item $j$ when $y_{ij}$ equals or exceeds a threshold, $\gamma_j$, and yields an incorrect response otherwise. Thus, on the assumption that $\varepsilon_{ij}$ is an unobservable random variable distributed $N(0,\sigma_j^2)$, the probability of an item score, $x_{ij} = 1$, indicating a correct response to item $j$ from person $i$, with abilities $\theta_i = (\theta_{1i}, \theta_{2i}, \ldots, \theta_{mi})$, is

$$P(x_{ij} = 1|\theta_i) = \frac{1}{(2\pi)^{1/2}\sigma} \int_{\gamma_j}^{\infty} \exp\left[ -\frac{1}{2}\left( \frac{y_j - \sum_k^m \alpha_{jk}\theta_{ki}}{\sigma_j} \right)^2 \right] dy_j = \Phi\left( -\frac{\gamma_j - \sum_k^m \alpha_{jk}\theta_{ki}}{\sigma_j} \right) = \Phi_j(\theta_i) \quad . \tag{2}$$

The conditional probability of the item score $x_i = 0$, indicating an incorrect response, is the complement, $1 - \Phi_j(\theta)$. In other words, the conditional response probability is given by a normal ogive model. Note that Equation 1 is a "compensatory" model: Greater ability on one dimension makes up for lesser ability on some other dimension. Nothing, however, prevents the methods discussed here from being applied to an "interactive" model such as

$$y_{ij} = \alpha_{j1}\theta_{1i} + \alpha_{j2}\theta_{2i} + \ldots + \alpha_{j12}\theta_{1i}\theta_{2i} + \ldots + \alpha_{jmp}\theta_{mi}\theta_{pi} + \varepsilon_{ij} \quad . \tag{3}$$

### Estimation of the Item Thresholds and Factor Loadings

Like maximum likelihood factor analysis for measured variables (Jöreskog, 1967), the Bock-Aitkin method of estimating parameters of an item response model assumes that the data have been obtained from a sample of persons drawn from a population with some multivariate distribution of ability. Provisionally, it will be assumed that the distribution is $\mathbb{\theta} \sim \mathrm{N}(\mathbb{0},\mathbb{I})$, but this assumption could be relaxed to allow for correlated factors and non-normal distributions. The convention of factor analysis that $y_j$ is distributed with mean 0 and variance 1, so that

$$\sigma_j^2 = 1 - \sum_{k=1}^{m} \alpha_{jk}^2 \quad , \tag{4}$$

is also adopted. On these assumptions, the marginal probability of the binary response pattern, $\mathbb{x}_\ell$, is given by the multiple integral,

$$\tilde{P}_\ell = P(\mathbb{x} = \mathbb{x}_\ell) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{j=1}^{n} [\Phi_j(\theta)]^{x_{\ell j}} [1 - \Phi_j(\theta)]^{1 - x_{\ell j}} g(\theta) d\theta = \int_{\theta} L_\ell(\theta) g(\theta) d\theta \quad . \tag{5}$$

Numerical approximation of this integral may be carried out by the $m$-fold Gauss-Hermite quadrature,

$$\tilde{P}_\ell = \sum_{q_m=1}^{Q} \ldots \sum_{q_2=1}^{Q} \sum_{q_1=1}^{Q} L_\ell(\mathbb{X}) A(X_{q_1}) A(X_{q_2}) \ldots A(X_{q_m}) \quad , \tag{6}$$

where $\mathbb{X}$ is a quadrature point in $m$-dimensional space and the corresponding weight is the product of weights for quadrature in the separate dimensions as shown. Equation 6 applies, of course, only to uncorrelated factors. It is an example of the so-called "product formula" for numerical integration and has the disadvantage that the number of terms in the sum is an exponential function of the number of dimensions. Fortunately, the number of points in each dimension can be reduced as the dimensionality is increased without impairing the accuracy of the approximations. Thus, factor analysis with five factors can be performed with good accuracy with as few as three points per dimension. In that case, $3^5 = 243$ quadrature points are required, and the solution is accessible to a fast computer.

Given the frequencies, $r_\ell$, of the response patterns $\mathbb{x}_\ell$ for $n$ items in a sample of $N$ persons, the number of distinct patterns is $s \leq \min(2^n, N)$, and the probability of the sample is

$$L_{\mathrm{M}} = P(X) = \frac{N!}{r_1!, r_2!, \ldots, r_s!} \tilde{P}_1^{r_1}, \tilde{P}_2^{r_2}, \ldots, \tilde{P}_s^{r_s} \quad . \tag{7}$$

Then the maximum likelihood estimates of the thresholds and factor loadings are those values that maximize Equation 7. To simplify the expression of the likelihood equations, it is convenient to express the model in terms of the intercept and slopes of the response function and to write

$$-\frac{\gamma_j - \sum_k^m \alpha_{jk}\theta_{ki}}{\sigma_j} = c_j + \sum_k^m a_{jk}\theta_{ki} \quad . \tag{8}$$

From marginal maximum likelihood (MML) estimates of $c_j$ and $a_{jk}$, MML estimates of $\gamma_j$ and $\alpha_{jk}$ may be obtained by

$$\hat{\gamma}_j = -\frac{\hat{c}_j}{\hat{d}_j} \tag{9}$$

and

$$\hat{\alpha}_{jk} = \frac{\hat{a}_{jk}}{\hat{d}_j} \quad , \tag{10}$$

where

$$\hat{d}_j = \left(1 + \sum_k^m \hat{a}_{jk}^2\right)^{1/2} \quad . \tag{11}$$

Notice that the item threshold in this model is not on an ability dimension. The invariant location parameter of the one-dimensional model does not exist in the multidimensional case: The value of one ability that corresponds to a given probability of correct response is a linear function of the other abilities.

The likelihood equation for a general item parameter, $\nu_j$, is

$$\begin{aligned}
\frac{\partial \log L_{\mathrm{M}}}{\partial \nu_j} &= \sum_{\ell=1}^{s} \frac{r_\ell}{\tilde{P}_\ell}\left(\frac{\partial \tilde{P}_\ell}{\partial \nu_j}\right) \\
&= \sum_{\ell}^{s} \frac{r_\ell}{\tilde{P}_\ell} \int_\theta \frac{L_\ell(\theta)}{[\Phi_j(\theta)]^{x_{\ell j}}[1-\Phi_j(\theta)]^{1-x_{\ell j}}} \cdot \frac{\partial\{[\Phi_j(\theta)]^{x_{\ell j}}[1-\Phi_j(\theta)]^{1-x_{\ell j}}\}}{\partial \nu_j}\, g(\theta)d\theta \\
&= \sum_{\ell}^{s} \frac{r_\ell}{\tilde{P}_\ell} \int_\theta \left(\frac{x_{\ell j} - \Phi_j(\theta)}{\Phi_j(\theta)[1-\Phi_j(\theta)]}\right) L_\ell(\theta)\, \frac{\partial \Phi_j(\theta)}{\partial \nu_j}\, g(\theta)d\theta \\
&= \int_\theta \frac{\bar{r}_j - \bar{N}\Phi_j(\theta)}{\Phi_j(\theta)[1-\Phi_j(\theta)]} \cdot \frac{\partial \Phi_j(\theta)}{\partial \nu_j}\, g(\theta)d\theta \quad ,
\end{aligned} \tag{12}$$

where

$$\bar{r}_j = \sum_{\ell=1}^{s} \frac{r_\ell x_{\ell j} L_\ell(\theta)}{\tilde{P}_\ell} \tag{13}$$

and

$$\bar{N} = \sum_{\ell=1}^{s} \frac{r_\ell L_\ell(\theta)}{\tilde{P}_\ell} \quad . \tag{14}$$

The multiple integral in this equation may be evaluated numerically by repeated Gauss-Hermite quadrature as follows:

$$\sum_{q_m}^{Q} \ldots \sum_{q_2}^{Q}\sum_{q_1}^{Q} \frac{\bar{r}_{j,q_1q_2\ldots q_m} - \bar{N}_{q_1q_2\ldots q_m}\Phi_j(\mathbb{X})}{\Phi_j(\mathbb{X})[1-\Phi_j(\mathbb{X})]}\left[\frac{\partial \Phi_j(\mathbb{X})}{\partial \nu_j} A(X_{q_1})A(X_{q_2})\ldots A(X_{q_m})\right] \quad . \tag{15}$$

The expected frequency $\bar{r}_{j,q_1q_2\ldots q_m}$ is an entry in a $Q^m$ dimensional array in which each cell corresponds to an $m$-tuple of quadrature points for a given item. The entries in this table are the numbers of examinees with abilities equal to the vector $\mathbb{X}$ who are expected to respond correctly to the item, given the sample data.

The quantity $\bar{N}_{q_1q_2\ldots q_m}$ is the margin of this array summed over items; it is the expected number of persons with ability $\mathbb{X}$ and is normalized to the sample size.

These equations correspond to the steps in the so-called EM algorithm for MML estimation as given by Dempster, Laird, and Rubin (1977). Equations 13 and 14 comprise the E-step, in which expectations of "complete data" statistics are computed conditional on the "incomplete data." Equation 12 is the M-step, in which conventional maximum likelihood estimation is carried out using the expectations in place of complete data statistics.

Because the expectations depend upon the parameters to be estimated, the calculations must be performed iteratively. Given starting values for the parameters, a $Q^m$ table of expected frequencies, $\bar{r}_{j,q_1q_2\ldots q_m}$, giving the numbers of correct responses at each point, $\mathbb{X}$, is built up for each item by distributing corresponding item scores weighted by the posterior probability of the response pattern, $\mathbb{x}_\ell$, at point $\mathbb{X}$. Similarly, $\bar{N}_{q_1q_2\ldots q_m}$ is obtained as the sum of the weights at each point. From these statistics, improved estimates of the item parameters are obtained in the M-step by applying the appropriate maximum likelihood solution to the table corresponding to the item in question. In the present case, any standard procedure for multiple-probit analysis will suffice for the M-step. But the procedure is general for any item-response model; if a logistic response model were assumed, a multiple-logit analysis would appear in the M-step. A version of this procedure based on the normal response model is implemented in the TESTFACT program of Wilson, Wood, and Gibbons (1984).

### Testing the Number of Resolvable Factors

If the sample size is sufficiently large that all $2^n$ possible response patterns have expected values greater than 1 or 2, the chi-square approximation for the likelihood ratio test of fit of the model relative to the general multinomial alternative is

$$G^2 = 2\sum r_\ell \ln\left(\frac{r_\ell}{N\tilde{P}_\ell}\right) \quad , \tag{16}$$

where $\tilde{P}_\ell$ is computed from the maximum likelihood estimates of the item parameters. The degrees of freedom are $2^n - n(m + 1) + m(m - 1)/2$. In this case, the goodness-of-fit test can be carried out after performing repeated full-information analyses, adding one factor at a time. When $G^2$ falls to insignificance, any additional factors that might be extracted could be attributed to sampling variation and should not be interpreted. More data would be required to justify the extraction and interpretation of further factors.

When the number of patterns is larger than the sample size, some of the expected frequencies may be near 0. In this case, Equation 16 (or other approximations to the likelihood ratio statistic for goodness-of-fit) becomes inaccurate and cannot be relied on. Haberman (1977) has shown, however, that the difference in these statistics for alternative models is distributed in large samples as chi-square, with degrees of freedom equal to the difference of respective degrees of freedom, even when the frequency table is sparse. Thus, the contribution of the last factor added to the model is significant if the corresponding change of chi-square is statistically significant. Properties of the change chi-square statistic are empirically investigated below.

## Implementation of Full-Information Factor Analysis

Typically, EM solutions converge so slowly that devices such as Ramsay's (1975) acceleration method must be used to speed up the computation. For the same reason, it is important that the solution begins from accurate starting values. A good strategy for obtaining starting values is to perform a principal factor analysis, with communality iteration, on the matrix of tetrachoric correlations of the items in question. The tetrachoric correlation matrix should be corrected for guessing and for missing values, and should be conditioned to be positive definite. This strategy has the advantage of allowing the investigator to dispense with the more costly full-information analysis in those cases where the maximum likelihood solution and the test of the number of factors are not essential. (For some empirical comparisons of tetrachoric and full-information solutions, see Muraki & Engelhard, 1985.)

Because the factors of the principal factor analysis are orthogonal, their loadings are suitable for the full-information solution after conversion to item intercepts and slopes. Item intercept and slope estimates based on the full-information method are then converted again into factor loadings. The resulting full-information factor pattern can be rotated orthogonally to the varimax criterion (Kaiser, 1958) and, with the varimax solution as target, rotated obliquely by the promax method (Hendrickson & White, 1964). The promax pattern is especially useful for identifying one-dimensional subsets of items into which a multidimensional set may be partitioned in order to measure abilities in the separate dimension.

## Correction For Guessing

Carroll (1945, 1983) has warned against artifacts introduced into item factor analysis by guessing on multiple-choice items. To suppress these effects, he proposed corrections to the four-fold tables from which the tetrachoric correlations are computed. In full-information analysis, a similar solution results from substituting, for the normal ogive response function, the guessing model with lower asymptote $g_j$:

$$\Phi_j^*(\theta) = g_j + (1 - g_j)\Phi_j(\theta) \quad . \tag{17}$$

If the item response model with guessing parameter is used for the full-information factor analysis, the tetrachoric correlation matrix used to produce starting values of the parameters must be corrected for guessing prior to the principal factor analysis. To express Carroll's correction method in terms of the proportions in the 2 × 2 table, let $g_i$ and $g_j$ denote the probability of chance success on items $i$ and $j$, respectively. Denote by $\pi_{ij}$ the observed proportions in the original 2 × 2 table, and by $\pi'_{ij}$ the proportions in the corrected 2 × 2 table. Thus, the original and corrected contingency tables may be expressed as in Tables 1 and 2, respectively.

The guessing parameter is the probability of observing a correct response when, given the true state of mastery for the item, the response should be failure. Thus, the observed proportion of passing is the sum of the proportion of the true state of mastery and the joint proportions of the corresponding guessing and the true failure state. Therefore the following expressions are obtained:

$$\pi_{1.} = \pi'_{1.} + g_i \pi'_{0.} \tag{18}$$

$$\pi_{.1} = \pi'_{.1} + g_j \pi'_{.0} \tag{19}$$

$$\pi_{11} = \pi'_{11} + g_i \pi'_{01} + g_j \pi'_{10} + g_i g_j \pi'_{00} \tag{20}$$

and

$$\pi'_{11} + \pi'_{01} + \pi'_{10} + \pi'_{00} = 1 \quad . \tag{21}$$

From Equations 18 through 21, the corrected proportions $\pi'$ are solved in terms of the observed proportion $\pi$ and guessing parameters $g$ as follows:

$$\pi'_{00} = \frac{\pi_{00}}{w_i w_j} \tag{22}$$

$$\pi'_{01} = \frac{w_j \pi_{01} - g_j \pi_{00}}{w_i w_j} \tag{23}$$

$$\pi'_{10} = \frac{w_i \pi_{10} - g_i \pi_{00}}{w_i w_j} \tag{24}$$

Table 1
Original Proportions of Examinees Passing and Failing Items $i$ and $j$

| | Item $j$ | | |
|---|---|---|---|
| Item $i$ | Pass | Fail | Total |
| Pass | $\pi_{11}$ | $\pi_{10}$ | $\pi_{1.}$ |
| Fail | $\pi_{01}$ | $\pi_{00}$ | $\pi_{0.}$ |
| Total | $\pi_{.0}$ | $\pi_{.0}$ | 1 |

and

$$\pi'_{11} = 1 - \pi'_{00} - \pi'_{01} - \pi'_{10} \quad , \tag{25}$$

where $w_i = 1 - g_i$ and $w_j = 1 - g_j$.

The following procedure corrects the item statistics for chance success. The conversion of the $k$th factor loading, $\alpha_{jk}$, to the provisional slope estimate, $a_{jk}$, is

$$a_{jk} = \frac{\alpha_{jk}}{\sigma_j} \quad , \tag{26}$$

where

$$\sigma_j^2 = 1 - \sum \alpha_{jk}^2 \quad . \tag{27}$$

The provisional intercept estimate, $c_j$, is computed from $\sigma_j$ and the standard difficulty, $\delta_j$, by

$$c_j = \frac{\delta_j}{\sigma_j} \quad , \tag{28}$$

because $\sigma_j$ is the reciprocal of $d_j$. The standard difficulty, $\delta_j$, is the inverse normal transform of the item facility, $\pi_j$, which is measured by the proportion of persons passing item $j$. The corrected facility, $\pi'_j$, is computed from

$$\pi'_j = 1 - \left( \frac{1 - \pi_j}{1 - g_j} \right) \quad . \tag{29}$$

## Correction For Omitted Responses

A disadvantage with Carroll's correction is that it fairly often produces a 0 or negative value in an off-diagonal element of the four-fold table. If all omitted responses are recoded as incorrect responses, the observed proportions, $\pi_{10}$, $\pi_{01}$, and $\pi_{00}$, tend to be inflated. Because the positive corrected proportions are obtained only if $\pi_{00}/\pi_{0.} \leqslant w_j$ and $\pi_{00}/\pi_{.0} \leqslant w_i$, negative corrected proportions are the likely result. This problem is almost always encountered because omitted responses are frequently found in cognitive testing. A possible solution for this problem is to allocate omitted responses to the categories of correct and incorrect responses, as shown below. This correction for omits must be made before the correction for guessing.

Let the observed frequencies in the 3 × 3 array whose categories are pass, fail, and omit, be expressed as in Table 3. If the proportions of correct and incorrect responses based on non-omitted responses are denoted by $p$s and $q$s respectively, they are computed by

Table 2
Corrected Proportions of
Examinees Passing and
Failing Items $i$ and $j$

| | Item $j$ | | |
|---|---|---|---|
| Item $i$ | Pass | Fail | Total |
| Pass | $\pi'_{11}$ | $\pi'_{10}$ | $\pi'_{1.}$ |
| Fail | $\pi'_{01}$ | $\pi'_{00}$ | $\pi'_{0.}$ |
| Total | $\pi'_{.1}$ | $\pi'_{.0}$ | 1 |

Table 3
Observed Frequencies of Examinees Passing, Failing, and Omitting Items *i* and *j*

| | Item *j* | | | |
|---|---|---|---|---|
| Item *i* | Pass | Fail | Omit | Total |
| Pass | $n_{11}$ | $n_{10}$ | $n_{1x}$ | $n_{1.}$ |
| Fail | $n_{01}$ | $n_{00}$ | $n_{0x}$ | $n_{0.}$ |
| Omit | $n_{x1}$ | $n_{x0}$ | $n_{xx}$ | $n_{x.}$ |
| Total | $n_{.1}$ | $n_{.0}$ | $n_{.x}$ | $n_{..}$ |

$$p_i = \frac{n_{11}+n_{10}}{N_{..}} \tag{30}$$

$$q_i = \frac{n_{01}+n_{00}}{N_{..}} \tag{31}$$

$$p_j = \frac{n_{11}+n_{01}}{N_{..}} \tag{32}$$

and

$$q_j = \frac{n_{10}+n_{00}}{N_{..}} \tag{33}$$

where $N_{..} = n_{11} + n_{10} + n_{01} + n_{00}$. If it can be assumed that omitted responses can be reallocated to correct and incorrect responses proportional to $p$ and $q$, the following corrected frequencies $n'_{ij}$ are obtained:

$$n'_{11} = n_{11} + p_j n_{1x} + p_i n_{x1} + p_i p_j n_{xx} \tag{34}$$

$$n'_{10} = n_{10} + q_j n_{1x} + p_i n_{x0} + p_i q_j n_{xx} \tag{35}$$

$$n'_{01} = n_{01} + p_j n_{0x} + q_i n_{x1} + q_i p_j n_{xx} \tag{36}$$

and

$$n'_{00} = n_{00} + q_j n_{0x} + q_i n_{x0} + q_i q_j n_{xx} \quad . \tag{37}$$

Marginal frequencies are computed by

$$n'_{1.} = n_{1.} + p_i n_{x.} \tag{38}$$

$$n'_{0.} = n_{0.} + q_i n_{x.} \tag{39}$$

$$n'_{.1} = n_{.1} + p_j n_{.x} \tag{40}$$

and

$$n'_{.0} = n_{.0} + q_j n_{.x} \quad . \tag{41}$$

Therefore,

$$n'_{1.} + n'_{0.} = n'_{.1} + n'_{.0} = n_{..} \quad , \tag{42}$$

because $p_i + q_i = p_j + q_j = 1$.

## Preliminary Smoothing of the Tetrachoric Correlation Matrix

Although the correction for omits makes the calculation of most of the tetrachoric correlations possible, there are still occasional instances in large matrices where a value close to 0 appears in the minor diagonal of the 2 × 2 table. Because no admissible coefficient can be computed from such a table, some method of imputing a value is required. A reasonable approach is to assume that the matrix of tetrachoric correlations is dominated by a single factor. In that case, Thurstone's centroid formula applied to the valid correlations can be used to estimate the item factor loadings from which the missing coefficients can be calculated. Because full-information analysis uses the tetrachoric correlations only for starting values, no bias of the solution results from these imputations.

To be analyzed by the MINRES method (Harman, 1976), the tetrachoric matrix must be positive definite. The corrected matrix obtained through the centroid method, on the other hand, may have 0 and negative roots. Therefore, a preliminary "smoothing" of the tetrachoric correlation coefficient matrix is needed before the principal factor analysis is carried out. The smoothed tetrachoric correlation matrix is produced from the eigenvectors associated with the positive roots, after renorming the sum of the roots to equal the number of items. The reproduced positive definite tetrachoric correlation matrix is then analyzed by the MINRES method to obtain good starting values for the full-information factor analysis.

## Constraints On Item Parameter Estimates

An undesirable feature of maximum likelihood factor analysis is its tendency to produce Heywood cases, that is, boundary solutions in which the uniqueness is 0 for one or more variables. These cases also occur in full-information item factor analysis, the symptom being one or more continually increasing item slopes as the EM cycles continue. Inasmuch as the existence of a nonzero component of measurement error variance for each variable is always assumed in factor analysis, these boundary solutions are inadmissible and should be constrained from occurring.

One way of handling this problem is to assume a restricted prior distribution on some of the item parameters and to employ a maximum posteriori (MAP) estimation, that is, to maximize the posterior probability density of the parameters rather than the likelihood. Martin and McDonald (1975) assumed an exponential distribution for the uniqueness, while Lee (1981) employed an inverted $\gamma$ prior. Mislevy (personal communication, April 1984) suggested that, because the uniqueness is bounded between 0 and 1, the $\beta$ prior

$$f(\sigma_j^2) = B(p,q)^{-1}(\sigma_j^2)^{p-1}(1-\sigma_j^2)^{q-1} \tag{43}$$

with $q = 1$ be used to hold $\sigma_j^2$ away from 0 without restricting its approach to 1. When $m = 2$, for example, MAP estimation with this prior adds the penalty function

$$-\frac{2(p-1)}{d_j^2}\begin{bmatrix} a_{j_1} \\ a_{j_2} \end{bmatrix} \quad , \tag{44}$$

where

$$d_j^2 = 1 + a_{j_1}^2 + a_{j_2}^2 \quad , \tag{45}$$

to the likelihood equations, and adds the ridge

$$\frac{2(p-1)}{d_j^4}\begin{bmatrix} d_j^2 - 2a_{j_1}^2 & -2a_{j_1}a_{j_2} \\ -2a_{j_1}a_{j_2} & d_j^2 - 2a_{j_2}^2 \end{bmatrix} \tag{46}$$

to the information matrix of the M-step maximum likelihood estimator. The present authors have found this method of suppressing Heywood cases to perform well in most cases. An apparent exception involves excessively difficult items for which the guessing parameter is larger than the sample item difficulty. In large datasets, such items may occasionally have to be dropped to avoid their slopes diverging despite the prior.

### Computing Times

Computing times depend upon the numbers of factors, items, examinees, quadrature points, EM cycles, M-step iterations, and the proportion of omitted items or items not presented. The preliminary steps of data input and computing starting values are not very time-consuming relative to the full-information solution. Most of the time in the latter is accounted for by the evaluation of likelihoods in the E-step; the M-step times are relatively small.

Some idea of the overall speed of the present implementation is given by a three-factor solution with 25 items, 1,178 examinees, $3^3 = 27$ quadrature points, 35 EM cycles, and a maximum of five M-step iterations. The IBM 3081 CPU time for this problem was 11 minutes and 43 seconds. Because most of the computation time is spent in the quadratures, which are readily amenable to vectorization, much faster computation would be possible on an array processing machine.

## Simulation Studies

### One-Factor Test

This simulation demonstrates the capacity of MML factor analysis to identify unidimensional item sets in the presence of guessing. To verify that the analysis has no tendency to produce difficulty factors, the item facilities were chosen to span a range larger than is typical of most tests of ability. This was done by setting the item intercepts at equally spaced points between $-2.0$ and $+2.0$. All item slopes were set at 1.0, corresponding to a factor loading of .707, and all guessing parameters (lower asymptotes) were set at .25. Responses with and without guessing were simulated for 1,000 examinees drawn randomly from a normal (0,1) distribution of ability.

Three analyses were performed: (1) no guessing assumed in the data or in the analysis; (2) guessing in the data but no guessing assumed in the analysis; and (3) guessing assumed in the data and in the analysis. In all of these analyses, the item intercepts and factor loadings were estimated from the data by an EM MML solution in which the iterations began from the principal factors of the sample tetrachoric correlation matrix (with communality iteration). Item guessing parameters, on the other hand, were set at their assumed values and not estimated.

It is instructive to examine the effects of guessing and the effect of correction for guessing on the item facilities and the item tetrachoric correlations. These relationships are shown graphically in Figures 1 and 2. Figure 1 confirms the well-known effect of guessing on item facilities. Deviation of the observed facilities from their theoretical values as a function of the true item intercepts is due entirely to sampling.

Figure 2 shows the average tetrachoric correlations for sets of three successive items ordered by facility. When guessing was not assumed or corrected for, the average coefficients are near their theoretical value of .5 at all levels of facility. When guessing was present but uncorrected, the average tetrachoric coefficients are attenuated, and the effect becomes greater as the items become more difficult. At the highest levels of difficulty, most of the correct responses are due to chance successes and the tetrachoric correlation is essentially 0. The effect of this attenuation is to increase the rank of the correlation matrix, and thus to introduce spurious factors in much the same way that variation in item difficulty introduces

**Figure 1**
Population and Sample Percent Correct as a Function of Item Threshold
(Simulated Data)

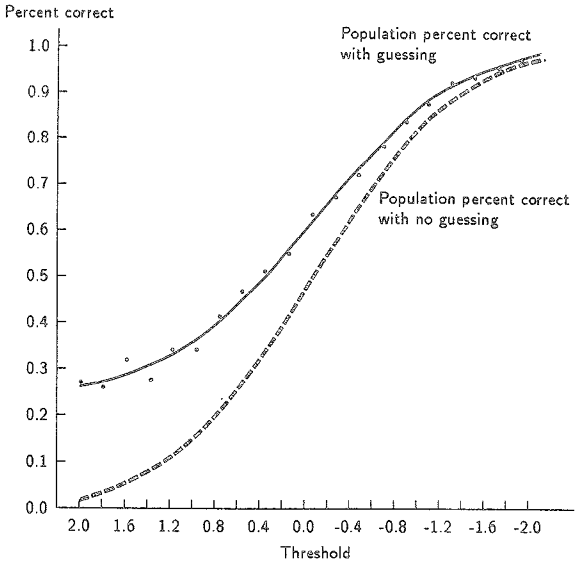
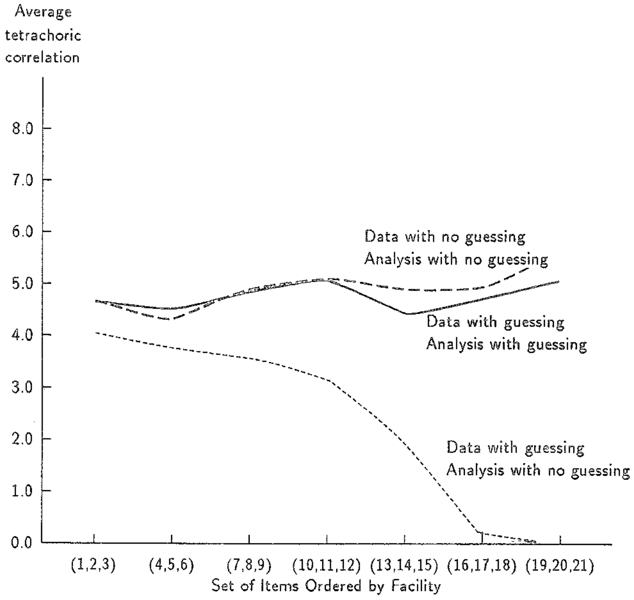


**Figure 2**
Average Tetrachoric Correlations of Sets of Three Successive Items



spurious factors in the analysis of item coefficients. Table 4 shows the distinctive pattern of loadings on the spurious second factor that results when guessing effects were ignored in the analysis: Items on either extreme of the difficulty continuum tend to have opposite signs.

When the guessing model was assumed, both in calculating the tetrachoric correlations and in the response function for the MML factor analysis, the deleterious effects of guessing were largely eliminated. As shown in Table 5, the likelihood ratio test for the addition of a second factor, which is significant when the no-guessing model was applied to guessing data in the second analysis, falls to insignificance when guessing is assumed in the third analysis. The estimated first factor loadings, which were much attenuated in the second analysis, are raised in the third analysis to near their true values.

These results illustrate the robustness of the analysis in identifying the number of factors and in estimating the factor loadings in the presence of a wide range of item difficulty and of guessing at a typical level of chance success. This relatively successful performance of the method is qualified, however, by its use of assigned rather than estimated guessing parameters. Underestimation of these parameters would certainly leave some effect of guessing in the solution and possibly produce spurious factors.

### Two-Factor Test

To demonstrate the power of MML item factor analysis to detect a second factor, a simulation study was conducted based on an analysis of the Auto and Shop Information subtest of the Armed Services

Table 4
Principal Factor Loadings From Simulated Data With True Factor Loadings of .707, With Guessing Effect Analyzed by No-Guessing and Guessing Models

| | No-Guessing Model | | Guessing Model |
|---|---|---|---|
| Item | Factor 1 | Factor 2 | Factor 1 |
| 1 | .703 | .147 | .761 |
| 2 | .719 | .046 | .724 |
| 3 | .739 | .215 | .732 |
| 4 | .654 | -.029 | .684 |
| 5 | .642 | .069 | .660 |
| 6 | .689 | .124 | .736 |
| 7 | .660 | .065 | .697 |
| 8 | .704 | .129 | .755 |
| 9 | .580 | -.032 | .697 |
| 10 | .561 | -.106 | .697 |
| 11 | .574 | -.049 | .710 |
| 12 | .583 | -.204 | .765 |
| 13 | .505 | -.102 | .715 |
| 14 | .393 | -.213 | .665 |
| 15 | .407 | -.168 | .704 |
| 16 | .329 | .003 | .716 |
| 17 | .274 | -.068 | .688 |
| 18 | .211 | -.081 | .653 |
| 19 | .148 | -.545 | .724 |
| 20 | .041 | -.068 | .594 |
| 21 | .128 | .069 | .759 |

Table 5
Analysis of Unidimensional Simulated Data: Change of the Likelihood Ratio $\chi^2$ Upon Adding a Second Factor to the Models With and Without Guessing

| Model | $\chi^2$ | df | $p$ |
|---|---|---|---|
| No Guessing | 39.166 | 20 | .006 |
| Guessing | 26.928 | 20 | .137 |

Vocational Aptitude Battery (ASVAB). This subtest was constructed from the previously separate Auto Information and Shop Information test items of the earlier Army Classification Battery. By a stepwise MML item factor analysis, three factors were extracted from the observed data for 1,178 cases from the Profile of American Youth study (Bock & Moore, 1986). As shown in Table 6, the change in the likelihood ratio chi-square due to inclusion of a second factor was significant, but that due to the third factor was not.

The resulting estimated factor loadings of the two-factor solution are plotted in Figure 3a after orthogonal rotation to the varimax criterion. The axes after oblique rotation to the promax criterion are also shown. Although items 1, 3, and 10 (and possibly item 2) are misclassified, the plot clearly separates the Auto and Shop subdivisions. Based on these loadings for the 25 items, binary scores of 1,000 simulated examinees were generated according to Equation 17 with the lower asymptote values estimated by Mislevy and Bock (1984b) using the BILOG program (Mislevy & Bock, 1984a). Factor scores were drawn randomly from a standard normal distribution.

These simulated data were then analyzed by MML item factor analysis with lower asymptotes assigned the specified values. Again two significant factors were found. Figure 3b gives the resulting varimax rotated factor loadings and promax rotated axes. The MML estimates based on the simulated responses are very similar to their generating values.

## Applications

### Analysis of the LSAT Section 7 With and Without Guessing

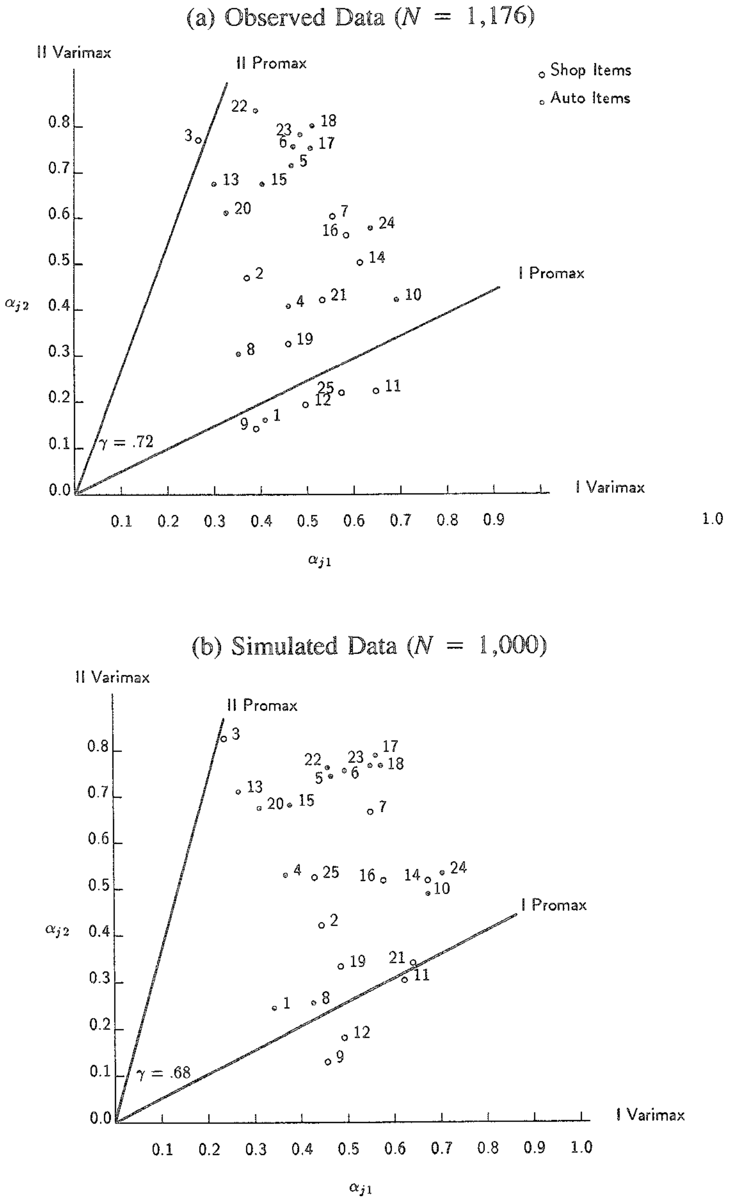
Table 7 shows the tetrachoric correlations uncorrected and corrected for guessing assuming an asymptote of .2 for all items. Note that the correction increases the magnitude of all of the coefficients.

Figure 4 shows the increase in marginal log likelihood in successive EM cycles of a two-factor solution without guessing. Even with the use of the Ramsay accelerator, the likelihood increases slowly as the solution point is approached. Twelve cycles were required for convergence.

Table 6
Change of the Likelihood Ratio $\chi^2$ With Assumed Design Effect of 2 in the Item Factor Analysis of the Auto and Shop Information Test

| Factor | $\chi^2$ | df | $p$ |
|---|---|---|---|
| 2 vs. 1 | 75.6 | 24 | .000 |
| 3 vs. 2 | 24.7 | 23 | .363 |

**Figure 3**
Factor Loadings for the Auto and Shop Information Test


(a) Observed Data ($N$ = 1,176)

(b) Simulated Data ($N$ = 1,000)

With five items and 1,000 examinees, these data permit the accurate calculation of goodness-of-fit chi-square as well as change chi-squares, as seen in Table 8. Both give evidence of a marginally significant second factor, and there is no indication that the guessing correction improves the solution. Similar conclusions are indicated by the residuals from the tetrachoric coefficients shown in Table 9.

Table 7
Tetrachoric Correlation Coefficients of the LSAT-7 Items (Coefficients Corrected for Guessing Above the Diagonal: $g$=.2; $N$=1000)

| | Item | | | | |
|---|---|---|---|---|---|
| Item | 1 | 2 | 3 | 4 | 5 |
| 1 | --- | .294 | .358 | .401 | .344 |
| 2 | .226 | --- | .567 | .288 | .174 |
| 3 | .291 | .432 | --- | .376 | .325 |
| 4 | .296 | .204 | .277 | --- | .214 |
| 5 | .286 | .135 | .265 | .161 | --- |

Figure 5 shows the principal factor starting values (open circles) and MML estimates of the factor loadings from the nonguessing solution (closed circles). It is apparent that loadings on the second factor are changed most by the full-information solution, and that the item with the most extreme correlations, item 5, is most affected. The factor axes rotated to the varimax and promax criteria show that item 2 most clearly determines the second factor.

## Quality of Life

Campbell, Converse, and Rodgers (1976) assessed 13 aspects of quality of life in 1,800 randomly selected respondents to a NORC survey. Respondents rated each quality in terms of their satisfaction with

**Figure 4**
Increase in Marginal Log Likelihood in Successive EM Cycles of a Two-Factor Solution Without Guessing: LSAT-7

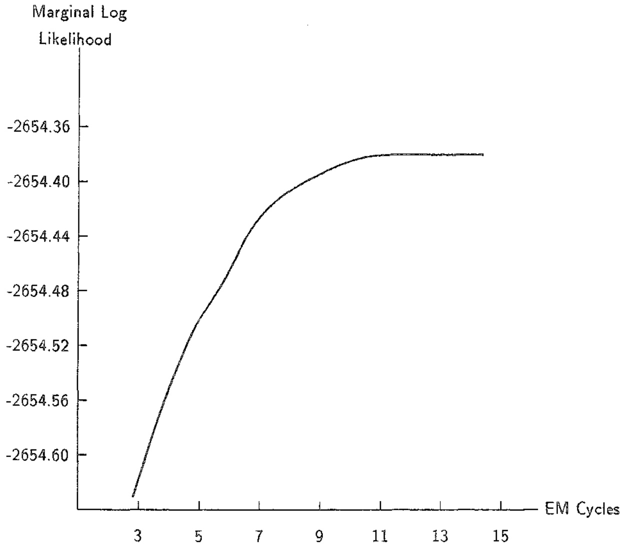


Table 8
$\chi^2$ Statistics for the Two-Factor Stepwise Analysis With and Without Guessing: LSAT-7 ($N$=1000)

| | No Guessing | | | Guessing | | |
|---|---|---|---|---|---|---|
| Analysis | $\chi^2$ | df | $p$ | $\chi^2$ | df | $p$ |
| One-Factor | 31.66 | 21 | .063 | 32.94 | 21 | .047 |
| Two-Factor | 22.86 | 17 | .154 | 24.80 | 17 | .099 |
| Change | 8.80 | 4 | .066 | 8.14 | 4 | .086 |

that aspect of their lives. For present purposes, these ratings were dichotomized at the neutral category, and a random sample of 1,000 cases was selected. A five-factor solution for these data is displayed in Table 10. Inspection of Table 10 clearly reveals five easily interpretable dimensions underlying the quality of life: (1) health, (2) satisfaction with the living environment (i.e., neighborhood and house quality), (3) satisfaction with everyday life (i.e., job, leisure, friends, family, and overall life), (4) financial satisfaction (i.e., savings and standard of living), and (5) satisfaction with self. In terms of level of satisfaction as indicated by the item thresholds in Table 10, most respondents were satisfied with their health, family, and friends; however, only the most satisfied respondents also reported satisfaction with their savings and education.

As a further verification of the factor solution, a limited-information GLS analysis was also performed (Muthén, 1978). The results of this analysis, employing Muthén's LISCOMP program, are also shown in Table 10; they correspond closely to those of the full-information solution. Parameter estimates were quite similar and the chi-square statistics for the improvement of fit with the addition of each new factor were virtually identical. The concordance between these two computationally different methods is taken as strong support for the consistency of both methods and the correctness of their implementations.

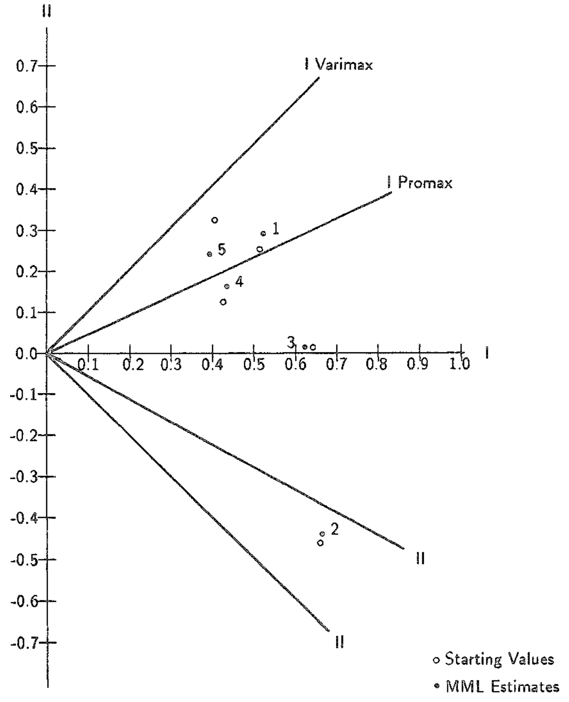
## DAT Spatial Reasoning

In a study of item features requiring spatial visualizing ability, Zimowski (1985) carried out a full-information item factor analysis of the Spatial Visualization subtest of the current edition of the Differential Aptitude Test battery (Bennett, Seashore, & Wesman, 1974). Examinees were 390 high school seniors from a suburban Chicago school system. The analysis revealed four statistically significant factors—a perhaps surprising result, considering that the test consists exclusively of pattern folding items. Upon

Table 9
LSAT-7 Residual Correlations: Guessing Above the Diagonal; No Guessing Below the Diagonal

| | Item | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| 1 | --- | .016 | −.005 | .043 | .032 |
| 2 | .009 | --- | .000 | .005 | −.048 |
| 3 | −.024 | .003 | --- | .037 | .050 |
| 4 | .026 | −.003 | .018 | --- | −.036 |
| 5 | .017 | −.015 | .034 | −.042 | --- |

**Figure 5**
Principal Factor Starting Values and MML Estimates of Factor Loadings



examining the items loading most heavily on a given factor, Zimowski found that they were constructed from basically the same stimulus pattern, but with a few additional marks and features to make them distinct items.

That these factors could represent distinct cognitive processes is an unpleasant prospect. A more plausible view is that a correct response on the first encounter with one of these items increases the probability of a correct response to later items from the similar set, while an incorrect response on the first encounter does not lead to such an increase. The resulting failures of conditional independence would produce increased associations among items that would appear as a factor.

It may be possible to distinguish this type of factor from a genuine cognitive process factor by position effects. Positively associated items should become less difficult as they are preceded by more items from the same dependent set. This sort of violation of conventional item-response theoretic assumptions could easily be corrected by avoiding repeated use of similar features among items in the same scale. Unfortunately, this strategy would rule out scales consisting of items generated from a facet design on the item content or format. It also presents difficulty for a componential theory of task difficulty on a single dimension.

## Discussion and Conclusions

Implementation of item factor analysis by marginal maximum likelihood estimation overcomes many of the problems that attend factor analysis of tetrachoric correlation coefficients: It avoids the problem of indeterminate tetrachoric coefficients of extremely easy or difficult items, it readily accommodates

Table 10
Quality of Life Data: Analysis by Marginal Maximum Likelihood (MML) and Generalized Least Squares (GLS)

| Satisfaction | Thresh- | Health | | Living Environment | | Finance | | Everyday Life | | Self | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| With | old | MML | GLS | MML | GLS | MML | GLS | MML | GLS | MML | GLS |
| 1. Neighborhood | .83 | .15 | .10 | .63 | .65 | .11 | .08 | .29 | .34 | .05 | .07 |
| 2. Education | .02 | .24 | .12 | .18 | .19 | .24 | .22 | .11 | .20 | .29 | .29 |
| 3. Job | .81 | .08 | −.03 | .19 | .18 | .33 | .32 | .57 | .62 | .07 | .08 |
| 4. Leisure | .67 | .34 | .22 | .20 | .20 | .18 | .16 | .42 | .50 | .38 | .40 |
| 5. Health | 1.07 | .64 | 1.10* | .04 | .07 | .09 | .09 | .12 | .18 | .26 | .20 |
| 6. Living Std. | .47 | .07 | .02 | .30 | .34 | .64 | .57 | .35 | .38 | .23 | .24 |
| 7. Friends | .98 | .10 | .04 | .19 | .18 | .12 | .08 | .49 | .50 | .28 | .33 |
| 8. Savings | −.17 | .17 | .11 | .18 | .16 | .72 | .83 | .20 | .20 | .13 | .17 |
| 9. House | .70 | .02 | .01 | .73 | .75 | .29 | .28 | .13 | .13 | .18 | .19 |
| 10. Family | .95 | .19 | .10 | .15 | .15 | .15 | .16 | .60 | .61 | .22 | .25 |
| 11. Life | .89 | .31 | .18 | .20 | .23 | .34 | .29 | .56 | .66 | .41 | .41 |
| 12. Life in U.S. | .85 | .43 | .25 | .14 | .17 | .12 | .11 | .40 | .48 | .02 | .10 |
| 13. Self | .90 | .30 | .15 | .11 | .11 | .21 | .20 | .35 | .32 | .74 | .86 |
| Change in $\chi^2$ | | | | 66.8 | 64.2 | 41.4 | 45.8 | 33.2 | 28.0 | 21.3 | 17.4 |
| df | | | | 12 | 12 | 11 | 11 | 10 | 10 | 9 | 9 |

*Heywood case.

effects of guessing and of omitted or not-reached items, and it provides a likelihood ratio test of the statistical significance of additional factors. Although the numerical integration used in the MML approach involves heavy computation and limits the procedure to five factors, the number of items that can be analyzed is sufficiently large (up to 100) to make the method useful in practical test development.

The applications of the procedure reported in the present paper show that, in moderately large samples (1,000 or more cases), minor factors determined by relatively few items can be detected as significant. The sensitivity of the MML method recommends it as an exploratory technique in searching for item features that are responsible for individual differences in cognitive test performance. By the same token, format attributes that may be implicated in failures of conditional independence are easily detected.

Studies of the ASVAB by Zimowski and Bock (1987) demonstrate the power of the item factor analysis in very large samples to isolate interpretable factors based on relatively few items sharing distinct content (including doublets) and accounting for only a few percent of the latent variation. In most cases, factors that are represented by item doublets would be better avoided by eliminating items that are appreciably more similar in content, form, and wording than other items in the test. Except when constructing items to appear in separate test forms, item writers should resist the temptation to create additional items by making minor variations on one idea or theme. Two such items in the same test form are very likely to turn up in a doublet defining a factor of no substantive interest.

In other cases, however, the additional factors may correspond to categories of test content that arise from potentially important dimensions of individual differences. In many cases these differences may correspond not to individual differences in distinct cognitive processes, but rather to subgroups of persons with differing educational and experiential exposure to item content. If it is important that the testing program measure the specialized proficiencies of these groups, then test development should move in the direction of better identifying the relevant classes of items and constructing separate unidimensional scales for each. This will give a more highly detailed profile of scale scores characterizing each examinee, from which more accurate predictions of job success could perhaps be obtained.

The length of the test battery would not necessarily have to increase in proportion to the number of scales added in this way; Bayesian methods could be used to draw strength from more general scales elsewhere in the battery to reduce the number of items required in the more specific scales. By this approach, batteries of vocational tests could be brought into better agreement with the assumptions of unidimensional IRT scoring without greatly increasing their length. Predictive validity should benefit as a result. Similar gains could be achieved, at the expense of greater computational complexity, by the use of multidimensional IRT models and scaling methods (see Muraki & Engelhard, 1985).

The problem of multidimensionality is more difficult in educational testing where the purpose is to measure general proficiencies in topics sampled from a conceptual domain. If, in addition to the dominant dimension (general factor), the domain includes one or more minor dimensions (group factors), the sampling of items is likely to yield subsets of items that are not conditionally independent when a one-dimensional model is assumed. In consequence, the one-dimensional form of Equation 5 for the marginal probability of $\theta$, given the answer pattern, is not valid and the corresponding estimates of the posterior mean and standard deviation may be misleading. In that case, the best course may be to assume a one-dimensional common factor model with correlated unique factors or, equivalently, correlated residuals. Then the common ability could be estimated by employing a method for computing conditional probabilities of answer patterns that does not assume conditional independence.

A good approximation for this purpose is provided by the so-called ''Clark algorithm,'' which applies when the conditional distribution of the item response processes is multivariate normal with an arbitrary covariance matrix. Calculation of such probabilities by means of the Clark algorithm, using the partial correlation matrix generated from the item factor loadings apart from the first principal factor, has been investigated by Gibbons and Bock (1987).

## References

Bartholomew, D. J. (1980). Factor analysis for categorical data. *Journal of the Royal Statistical Society, Series B, 42*, 293–321.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1974). *Manual for the Differential Aptitude Tests Forms S and T* (5th ed.). New York: The Psychological Corporation.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*, 443–459.

Bock, R. D., & Moore, E. G. J. (1986). *Advantage and disadvantage: A profile of American youth*. Hillsdale NJ: Erlbaum.

Campbell, A., Converse, P. E., & Rodgers, W. L. (1976). *The quality of American life*. New York: Russell Sage Foundation.

Carroll, J. B. (1945). The effect of difficulty and chance success on correlations between items or between tests. *Psychometrika, 10*, 1–19.

Carroll, J. B. (1983). The difficulty of a test and its factor composition revisited. In H. Wainer & S. Messick (Eds.), *Principals of modern psychological measurement* (pp. 257–282). Hillsdale NJ: Erlbaum.

Cristoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika, 40*, 5–32.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B, 39*, 1–38.

Divgi, D. R. (1979). Calculation of the tetrachoric correlation coefficient. *Psychometrika, 44*, 169–172.

Gibbons, R. D., & Bock, R. D. (1987, June). *Approximating multivariate normal orthant probabilities*. Paper presented at the annual meeting of the Psychometric Society, Montreal.

Haberman, J. S. (1977). Log-linear models and frequency tables with small expected cell counts. *Annals of Statistics, 5*, 1148–1169.

Harman, H. H. (1976). *Modern factor analysis*. Chicago: University of Chicago Press.

Hendrickson, A. E., & White, P. O. (1964). PROMAX: A quick method for rotation to oblique simple structure. *British Journal of Mathematical and Statistical Psychology, 17*, 65–70.

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

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

Lee, S.-Y. (1981). A Bayesian approach to confirmatory factor analysis. *Psychometrika, 46,* 153–160.

Martin, J. K., & McDonald, R. P. (1975). Bayesian estimation in unrestricted factor analysis: A treatment for Heywood cases. *Psychometrika, 40,* 505–517.

Mislevy, R. J., & Bock, R. D. (1984a). *BILOG: Maximum likelihood item analysis and test scoring—Logistic model.* Mooresville IN: Scientific Software, Inc.

Mislevy, R. J., & Bock, R. D. (1984b). *Item operating characteristics of the Armed Services Vocational Aptitude Battery (ASVAB) Form 8A.* Chicago: National Opinion Research Center.

Muraki, E., & Engelhard, G. (1985). Full-information item factor analysis: Applications of EAP scores. *Applied Psychological Measurement, 9,* 417–430.

Muthén, B. (1978). Contributions to factor analysis of dichotomized variables. *Psychometrika, 43,* 551–560.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical and continuous latent variable indicators. *Psychometrika, 49,* 115–132.

Ramsay, J. O. (1975). Solving implicit equations in psychometric data analysis. *Psychometrika, 40,* 337–360.

Wilson, D., Wood, R., & Gibbons, R. D. (1984). *TESTFACT: Test scoring, item statistics, and item factor analysis.* Mooresville IN: Scientific Software, Inc.

Zimowski, M. F. (1985). *Attributes of spatial test items that influence cognitive processing.* Unpublished doctoral dissertation, University of Chicago.

Zimowski, M. F., & Bock, R. D. (1987). *Full information factor analysis of items from the ASVAB CAT pool* (NORC report 87-1). Chicago: National Opinion Research Center.

## Acknowledgments

*This research was supported in part by National Science Foundation Grant #BNS-8511774 to the University of Chicago, ONR contract #N00014-83-0457 to NORC, and ONR Grant #N00014-85-K-0586 to the University of Illinois, Chicago Campus.*

## Author's Address

Send requests for reprints or further information to R. Darrell Bock, Department of Behavioral Sciences, University of Chicago, 5848 South University Avenue, Chicago IL 60637, U.S.A.