DOCUMENT RESUME

ED 428 100 TM 029 501

AUTHOR Beguin, Anton A.; Glas, Cees A. W.
TITLE MCMC Estimation of Multidimensional IRT Models. Research Report 98-14.
INSTITUTION Twente Univ., Enschede (Netherlands). Faculty of Educational Science and Technology.
PUB DATE 1998-00-00
NOTE 30p.
AVAILABLE FROM Faculty of Educational Science and Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.
PUB TYPE Reports - Evaluative (142)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Ability; *Bayesian Statistics; *Estimation (Mathematics); *Item Response Theory; *Markov Processes; *Monte Carlo Methods; Simulation
IDENTIFIERS Gibbs Sampling; *Multidimensionality (Tests); Two Parameter Model

ABSTRACT

A Bayesian procedure to estimate the three-parameter normal ogive model and a generalization to a model with multidimensional ability parameters are discussed. The procedure is a generalization of a procedure by J. Albert (1992) for estimating the two-parameter normal ogive model. The procedure will support multiple samples from multiple populations and restrictions on the factor matrix for testing specific hypotheses about the ability structure. The technique is illustrated using simulated and real data. A Markov chain Monte Carlo (MCMC) procedure is used to sample the posterior distributions of interest and needed chains are constructed using the Gibbs sampler (A. Gelfand and A. Smith, 1990). (Contains 8 tables and 47 references.) (Author/SLD)

MCMC Estimation of Multidimensional IRT Models

Research Report 98-14

Anton A. Béguin
Cees A.W. Glas

faculty of
EDUCATIONAL SCIENCE AND TECHNOLOGY

University of Twente

Department of
Educational Measurement and Data Analysis

MCMC estimation of multidimensional IRT models

Anton A. Béguin

Cees A.W. Glas

Department of Educational Measurement and Data Analysis
Faculty of Educational Science and Technology
University of Twente, Enschede, the Netherlands
E-mail A.A.Beguin@edte.utwente.nl

Abstract

In this paper, a Bayesian procedure to estimate the 3 parameter normal ogive model and a generalization to a model with multidimensional ability parameters will be discussed. The procedure is a generalization of a procedure by Albert (1992) for estimating the 2 parameter normal ogive model. The procedure will support multiple samples from multiple populations and restrictions on the factor matrix for testing specific hypotheses about the ability structure. The technique is illustrated using simulated and real data.

Key words: Bayes estimates, Full Information Factor Analysis, Gibbs sampler, item response theory, Markov chain Monte Carlo, multidimensional item response theory, normal ogive model.

Introduction

Item response theory (IRT) models are stochastic models for the responses of persons to items, where the influence of items and persons on the responses are modeled by disjunct sets of parameters. In the framework of educational and psychological measurement, the person parameter can usually be labeled ability, and in many instances it suffices to assume that ability is unidimensional. However, in other instances it may be a priori clear that multiple skills or abilities are involved in producing the manifest responses, or the dimensionality of the ability structure might not be clear at all. In these situations, multidimensional IRT models can serve confirmatory and explorative purposes. The focus of this paper will be dichotomously scored items. Multidimensional IRT models for dichotomously scored items were first presented by Lord and Novick (1968) and McDonald (1967). These authors use a normal ogive to describe the probability of a correct response. McDonald (1967,1997) developed an estimation procedure based on an expression for the association between pairs of items and derived from a polynomial expansion of the normal ogive. The procedure is implemented in NOHARM (Normal-Ogive Harmonic Analysis Robust Method, Fraser, 1988). An alternative using all information in the data, and therefore labeled "Full Information Factor Analysis" was developed by Bock, Gibbons, and Muraki, (1988). This approach is a generalization of the marginal maximum likelihood (MML) and Bayes model estimation procedures for unidimensional IRT models (see, Bock & Aitkin, 1981, Mislevy, 1986), and has been implemented in TESTFACT (Wilson, Wood, and Gibbons, 1991). A comparable model using a logistic rather than a normal-ogive representation has been studied by Reckase (1985, 1997) and Ackerman (1996a and 1996b).

The aim of the present paper is to present a Bayesian approach to estimating normal ogive multidimensional IRT models. The method is a generalization of the Bayesian approach to estimating the two-parameter normal ogive model (2PNO) by Albert (1992). The method will provide marginal posterior density estimates for any parameter of interest. Besides that these estimates can be used to judge the accuracy of normal approximations of the standard error of MML estimates, the method also has the advantage that restrictions on the dimensionality in the MML framework emanating from using Gaussian quadrature do not apply here (Bock, Gibbons, & Muraki, 1988).

This paper is organized as follows. In the next section the procedure will be developed for the unidimensional 3-PL normal ogive model (3PNO). In the third section this procedure will be generalized to $Q-$multidimensional normal ogive models, both with and without guessing parameters. The former model will be labeled $(Q+2)$PNO, the latter $(Q+1)$PNO. The procedure will be defined in such a way that generalization to multiple samples from multiple populations is straightforward. In the fifth section, it will be shown how restrictions can be imposed on the factor structure and how they can be used to facilitate the interpretation of the model. Then, in Section 6, two examples will be given, a simulated example to evaluate parameter recovery and an example using real data to exemplify the approach of using restrictions to produce an interpretable factor solution. Finally, in the discussion section, some suggestions for further research will be given.

Bayesian Estimation of the 3PNO

In this paper, a Markov chain Monte Carlo procedure will be used to sample the posterior distributions of interest, and the needed chains will be constructed using the Gibbs sampler (Gelfand & Smiths, 1990). To implement the Gibbs sampler, the parameter vector is divided in a number of components, and each successive component is sampled from its conditional distribution given sampled values of all other components. This sampling scheme is repeated until the sampled values produce stable posterior distributions. Albert (1992) applies Gibbs sampling to estimate the parameters of the well known two-parameter normal ogive model (Lawley, 1943, 1944, Lord, 1952, 1953a and 1953b). In this section, the procedure will be generalized to the 3PNO, in the following section, this approach will be generalized further to a multidimensional framework. In the 3PNO, the probability of a correct response of a person i on an item j, denoted by $Y_{ij} = 1$, can be written as

$$\begin{aligned} P(Y_{ij}=1;\theta_i,\alpha_j,\beta_j,\gamma_j) &= \gamma_j+(1-\gamma_i)\Phi(\eta_{ij}) \\ &= \Phi(\eta_{ij})+\gamma_j(1-\Phi(\eta_{ij})), \end{aligned} \tag{1}$$

where γ_j is called the 'pseudo-guessing parameter', Φ denotes the standard normal cumulative distribution function, and $\eta_{ij}=\alpha_j\theta_i-\beta_j$ with θ_i the ability of person i, α_j, the discrimination parameter and β_j the difficulty parameter of the item j, respectively. In (1) the usual expression for the 3PNO is rewritten to an expression that supports an interpretation of the model to be used below. In this interpretation, there is a probability $\Phi(\eta_{ij})$ that the respondent knows the item and gives a correct response with probability one, and a probability $(1-\Phi(\eta_{ij}))$ that the respondent does not know the item and guesses with γ_j as the probability of a correct response. So the probability of a correct response is a sum of a term $\Phi(\eta_{ij})$ and a term $\gamma_j(1-\Phi(\eta_{ij}))$. In line with this interpretation, it will prove convenient to introduce a vector of binary variables W_{ij} such that

$$W_{ij}=\begin{cases} 1 & \text{if person } i \text{ knows the correct answer to item } j \\ 0 & \text{if person } i \text{ doesn't know the correct answer to item } j. \end{cases} \tag{2}$$

So if $W_{ij}=0$, person i will guess the response to item j, if $W_{ij}=1$, person i will know the right answer and will give a correct response. Consequently the conditional probability of W_{ij} given Y_{ij} is given by

$$\begin{aligned} P(W_{ij}=1\,|Y_{ij}=1,\eta_{ij},\gamma_j) &\propto \Phi(\eta_{ij}) \\ P(W_{ij}=0\,|Y_{ij}=1,\eta_{ij},\gamma_j) &\propto \gamma_j(1-\Phi(\eta_{ij})) \\ P(W_{ij}=1\,|Y_{ij}=0,\eta_{ij},\gamma_j) &= 0 \\ P(W_{ij}=0\,|Y_{ij}=0,\eta_{ij},\gamma_j) &= 1. \end{aligned} \tag{3}$$

To implement the Gibbs sampler the data $\mathbf{y}$, responses of n persons to k items, will be augmented with latent data $\mathbf{W}=(W_{11},...,W_{nk})^T$. Further, following Albert (1992) the data are also augmented with latent data $\mathbf{Z}=(Z_{11},...,Z_{nk})^T$, where the variables Z_{ij} are independent and normally distributed with mean η_{ij} and a standard deviation equal to one. These variables are related to $\mathbf{W}$ by $Z_{ij}>0$ if $W_{ij}=1$ and $Z_{ij}\leq 0$ if $W_{ij}=0$. This can be written as

$$p(Z_{ij}\,|W_{ij},\eta_{ij}) \propto \phi(Z_{ij};\eta_{ij},1)\,(\mathbf{I}(Z_{ij}>0)\mathbf{I}(W_{ij}=1)+\mathbf{I}(Z_{ij}\leqslant 0)\mathbf{I}(W_{ij}=0)) \tag{4}$$

where $\phi(Z_{ij};\eta_{ij},1)$ stands for the normal density with mean η_{ij} and standard deviation one evaluated at Z_{ij}. It is assumed that θ_i is standard normally distributed, the item parameters $\boldsymbol{\xi}$, $\boldsymbol{\xi}_j = (\alpha_j,\beta_j)^T$, are given a prior $p(\boldsymbol{\xi}) = \prod_{j=1}^{k} \mathbf{I}(\alpha_j > 0)$ to insure that the discrimination parameters are positive, and the pseudo-guessing parameter γ_j has the conjugate noninformative prior Beta$(1,1)$.

The aim of the procedure is to simulate samples from the joint posterior distribution of $\boldsymbol{\xi}, \boldsymbol{\theta}, \mathbf{Z}$ and $\mathbf{W}$,given by

$$\begin{aligned} p(\boldsymbol{\xi},\boldsymbol{\theta},\mathbf{Z},\mathbf{W}\,|\mathbf{y}) &= p(\mathbf{Z},\mathbf{W}\,|\mathbf{y}\,;\boldsymbol{\xi},\boldsymbol{\gamma},\boldsymbol{\theta},)p(\boldsymbol{\theta})p(\boldsymbol{\xi})p(\boldsymbol{\gamma}) \\ &= C\prod_{i=1}^{n}\Big\{\Big(\prod_{j=1}^{k} p(Z_{ij}\,|W_{ij},\eta_{ij})p(W_{ij}\,|y_{ij},\eta_{ij},\gamma_j)\Big\} \\ &\quad \phi(\theta_i;0,1)\left(\prod_{q=1}^{Q}\mathbf{I}(\alpha_{jq}>0)\right)\prod_{j=1}^{k}p(\gamma_j) \end{aligned} \tag{5}$$

where $p(W_{ij}\,|y_{ij},\eta_{ij},\gamma_j)$ is given by (3) and $p(Z_{ij}\,|W_{ij},\eta_{ij})$ follows from (5).

Although the distribution given by (5) has an intractable form, the fully conditional distributions of $\mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\xi}$ and $\boldsymbol{\gamma}$ are each tractable and easy to simulate from. A draw from the full conditional distribution can be obtained in five steps.

Step 1 Draw from the conditional distribution of Z_{ij} given all other variables. Using (5), the distribution of Z_{ij} conditional on $\mathbf{W}$, $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ is given by

$$Z_{ij}\,|\mathbf{W},\boldsymbol{\theta},\boldsymbol{\xi},\mathbf{y}\ dist.\begin{cases} N(\eta_{ij},1) \text{ truncated at the left by } 0 & \text{if } W_{ij}=1 \\ N(\eta_{ij},1) \text{ truncated at the right by } 0 & \text{if } W_{ij}=0. \end{cases}$$

Step 2 Draw from the conditional distribution of $\boldsymbol{\theta}$ given $\mathbf{W}$, $\mathbf{Z}$ and $\boldsymbol{\xi}$. From (4) and (5) it follows that the fully conditional distribution of $\boldsymbol{\theta}$ entails a normal model for the regression of $Z_{ij} - \beta_j$ on α_j,with θ_i as a regression coefficient with a normal prior with parameters $\mu = 0$ and $\sigma = 1$. Therefor, the posterior of θ_i is normal, that is,

$$\theta_i \text{ is distributed } N\left(\frac{\hat{\theta}_i/v+\mu/\sigma^2}{1/v+1/\sigma^2},\frac{1}{(1/v+1/\sigma^2)}\right), \tag{6}$$

where $\hat{\theta}_i = \sum_j \alpha_j(Z_{ij}+\beta_j)/\sum_j \alpha_j^2$ and variance v=$1/\sum_j \alpha_j^2$.

Step 3 Draw from the conditional distribution of $\boldsymbol{\xi}_j$ given $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$. The variables $\boldsymbol{\xi}_j$ can be viewed as coefficients of the regression of $\mathbf{Z}_j = (Z_{1j},\ldots,Z_{nj})^T$ on $\mathbf{X} = (\boldsymbol{\theta}-1)$, with -1 the

n dimensional column vector with elements -1. Therefore,

$$\boldsymbol{\xi}_j \,|\boldsymbol{\theta}, \mathbf{Z}_j, \mathbf{y} \text{ is distributed } N(\hat{\boldsymbol{\xi}}_j, (\mathbf{X}^T\mathbf{X})^{-1})\mathbf{I}(\alpha_j > 0), \tag{7}$$

where $\hat{\boldsymbol{\xi}}_j = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Z}_j$.

Step 4 Draw from the distribution of $\mathbf{W}$ conditional on all other variables given by (3).

Step 5 Sample from the conditional distribution of γ_j. Let t_j be defined as $t_j = \sum_{i=1}^{n} I(W_{ij} = 0)$, that is, t_j is the number of persons who do not know the correct answer to item j and guess the response. The probability of a correct response of person i on item j given $W_{ij} = 0$ is $P(y_{ij} = 1 \,|W_{ij} = 0) = \gamma_j$. The number of correct responses obtained by guessing, say, $s_j = \sum_{i|w_{ij}=0}^{n} y_{ij}$, has a binomial distribution, $\text{Bin}(t_j, \gamma_j)$. With the noninformative conjugate Beta prior, the posterior distribution of γ_j is

$$\gamma_j \,|\mathbf{W}, \mathbf{y} \sim \text{Beta}(s_j + 1, t_j - s_j + 1). \tag{8}$$

The procedure boils down to generating a number of parameter sequences iteratively using these five steps. Convergence of the procedure can be evaluated by comparing the between and within sequences variance (see, for instance, Gelman, Carlin, Stearn & Hall, 1995). Starting points of the sequences can be provided by the Bayes modal estimates of Bilog (Zimowski, Muraki, Mislevy, & Bock, 1996).

Multidimensional normal ogive model

In a multidimensional normal ogive model with a guessing parameter (Q+2)PNO, the probability of a correct response, $Y_j = 1$, on an item j is given by

$$P(Y_{ij} = 1 \,|\boldsymbol{\theta}_i, \boldsymbol{\xi}_j, \gamma_j) = \gamma_j + (1 - \gamma_j)\Phi(\eta_{ij}), \tag{9}$$

where $\boldsymbol{\theta}_i$ is a vector of Q ability parameters θ_{qi}, $q = 1, ..., Q$, $\boldsymbol{\xi}_j = (\alpha_{j1}, ..., \alpha_{jq}, ..., \alpha_{jQ}, \beta_j)^T$ is the vector of parameters of item j, and $\eta_{ij} = \sum_{q=1}^{Q} \alpha_{jq}\theta_{qi} - \beta_j$. Further, it is assumed that all ability parameters $\boldsymbol{\theta}_i$, $i = 1, ..., n$, are independent and have identical multivariate normal distributions with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}_\theta$.

To identify the model, restrictions have to be imposed on the parameters. One approach, which, for instance, is used in the exploratory option of NOHARM (Fraser, 1988), is

setting the mean and the covariance matrix equal to zero and the identity matrix, respectively, and introducing the constraints $\alpha_{jq} = 0$, for $j = 1, ..., Q-1$ and $q = j+1, ..., Q$. An alternative approach is considering the parameters of ability distribution as unknown estimands, and to identify the model by setting Q item parameters β_j equal to zero and, for $j = 1, ..., Q$ and $q = j+1, ..., Q$, imposing the restrictions $\alpha_{jq} = 1$, if $j = q$, and $\alpha_{jq} = 0$, if $j \neq q$. Below it will be shown that these and the restrictions of the exploratory option of NOHARM are equivalent. The reason for the latter identification is that it will be easier to generalize the estimation procedure to a multiple group situation, where additional ability distributions are considered, which are labeled with unique means and covariance matrices. But to keep the presentation simple, it will first be assumed that all persons are associated with an identical ability distribution; the generalization will be returned to later, together with a generalization to incomplete test administration designs, where different persons respond to different items.

The procedure to be presented will be a generalization of the above procedure. The definition of the variables, $\mathbf{W} = (W_{11}, ..., W_{nk})$ and $\mathbf{Z} = (Z_{11}, ..., Z_{nk})$ is as above, but with an alternative definition of η_{ij}. It is assumed that the ability of person n, $\boldsymbol{\theta}_n$, is Q-variate normally distributed with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$, that is, $\boldsymbol{\theta}_n \backsim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_\theta)$. Further, the prior for the item parameters $\boldsymbol{\xi}$ is given by $\pi(\boldsymbol{\xi}) = \prod_{j=1}^{k} \prod_{q=1}^{Q} \mathbf{I}(\alpha_{jq} > 0)$, that is, it is assumed that the probability of a correct response is positively related to all latent abilities involved. As above, the pseudo guessing parameter γ_j will have a Beta prior. The joint posterior of $\boldsymbol{\xi}, \boldsymbol{\theta}, \mathbf{Z}, \mathbf{W}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{\Sigma}_\theta$ is given by

$$
\begin{aligned}
p(\boldsymbol{\xi}, \boldsymbol{\theta}, \mathbf{Z}, \mathbf{W}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{\Sigma} \,|\mathbf{y}) &= p(\mathbf{Z}, \mathbf{W} \,|\mathbf{y}\,; \boldsymbol{\xi}, \boldsymbol{\gamma}, \boldsymbol{\theta},) p(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}_\theta) p(\boldsymbol{\mu} \,|\boldsymbol{\Sigma}_\theta\,) p(\boldsymbol{\Sigma}_\theta) p(\boldsymbol{\xi}) p(\boldsymbol{\gamma}) \\
&= \prod_{i=1}^{n} \prod_{j=1}^{k} p(Z_{ij} \,|W_{ij}\,; \eta_{ij}) p(W_{ij} \,|y_{ij}\,; \eta_{ij}, \gamma_j) p(\boldsymbol{\theta}_i \,|\boldsymbol{\mu}, \boldsymbol{\Sigma}_\theta) \\
&\quad p(\boldsymbol{\mu} \,|\boldsymbol{\Sigma}_\theta\,) p(\boldsymbol{\Sigma}_\theta) p(\boldsymbol{\xi}) p(\boldsymbol{\gamma})
\end{aligned}
$$

where $p(Z_{ij} \,|W_{ij}\,; \eta_{ij})$ and $p(W_{ij} \,|y_{ij}\,; \eta_{ij}, \gamma_j)$ are as defined in (4) and (5), with the proper substitution of η_{ij}. Finally, $p(\boldsymbol{\mu} \,|\boldsymbol{\Sigma}_\theta\,)$ and $p(\boldsymbol{\Sigma}_\theta)$ will be normal and inverse Wishart, respectively.

Again, the procedure consists of simulating from fully conditional posteriors, in this case, the fully posterior distributions of $\boldsymbol{\xi}, \boldsymbol{\theta}, \mathbf{Z}, \mathbf{W}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{\Sigma}_\theta$. Therefore, the procedure consists of seven steps:

(1) Draw $\boldsymbol{\Sigma}_\theta$ conditional on $\boldsymbol{\theta}$.

(2) Draw $\boldsymbol{\mu}$ conditional on $\boldsymbol{\Sigma}_\theta$ and $\boldsymbol{\theta}$.

(3) Draw $\mathbf{W}$ conditional on $\boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\gamma}$ and $\mathbf{Y}$.
(4) Draw $\mathbf{Z}$ conditional on $\mathbf{W}, \boldsymbol{\theta}$, and $\boldsymbol{\xi}$.
(5) Draw $\boldsymbol{\theta}$ conditional on $\mathbf{Z}, \boldsymbol{\xi}, \boldsymbol{\Sigma}_\theta$, and $\boldsymbol{\mu}$.
(6) Draw $\boldsymbol{\xi}$ conditional on $\mathbf{Z}$ and $\boldsymbol{\theta}$.
(7) Draw $\boldsymbol{\gamma}$ conditional on $\mathbf{W}$ and $\mathbf{Y}$.

In this section, it will be assumed that there are appropriate initial estimates of $\boldsymbol{\theta}, \boldsymbol{\xi}$ and $\boldsymbol{\gamma}$, how these initial estimates might be obtained will be discussed later. With this assumption, the seven steps can be executed in the following order.

Step 1 The conjugate prior distribution for $(\boldsymbol{\mu}, \boldsymbol{\Sigma}_\theta)$ is a product of a normal and an inverse-Wishart distribution (see, for instance, Box & Tiao, 1973). Let $\boldsymbol{\Lambda}_0$ be the scale matrix and let υ_0 be the degrees of freedom for the prior of $\boldsymbol{\Sigma}_\theta$. Further, let $\boldsymbol{\mu}_0$ be the prior mean and let κ_0 be the number of prior measurements on the $\boldsymbol{\Sigma}_\theta$ scale. The posterior is now also inverse-Wishart distributed with parameters

$$\begin{aligned}
\boldsymbol{\mu}_n &= \tfrac{\kappa_0 n}{\kappa_0+n}\boldsymbol{\mu}_0 + \tfrac{n}{\kappa_0+n}\bar{\boldsymbol{\theta}} \\
\upsilon_n &= \upsilon_0 + n \\
\kappa_n &= \kappa_0 + n \\
\boldsymbol{\Lambda}_n &= \boldsymbol{\Lambda}_o + \mathbf{S} + \tfrac{\kappa_0 n}{\kappa_0+n}(\bar{\boldsymbol{\theta}} - \boldsymbol{\mu}_0)(\bar{\boldsymbol{\theta}} - \boldsymbol{\mu}_0)^T,
\end{aligned}$$

where $\mathbf{S}$ is a matrix with elements $S_{pq} = \sum_{i=1}^{n}(\theta_{ip} - \bar{\theta}_p)(\theta_{iq} - \bar{\theta}_q)$ and $\bar{\theta}_p$ and $\bar{\theta}_q$ are the means of the ability parameters of dimension p and q, respectively. The choice of a value for $\boldsymbol{\mu}_0$ will be returned to below. With respect to the choice of κ_0, υ_0 and $\boldsymbol{\Lambda}_o$, a noninformative prior distribution is obtained if $\kappa_0 \to 0, \upsilon_0 \to -1$ and $|\boldsymbol{\Lambda}_o| \to 0$. This results in the multivariate version of Jeffreys' prior density. The corresponding posterior distribution can be written as

$$\begin{aligned}
\boldsymbol{\Sigma}_\theta \,|\boldsymbol{\theta} &\sim \text{Inverse-Wishart}_{n-1}(\mathbf{S}) \\
\boldsymbol{\mu} \,|\boldsymbol{\Sigma}_\theta, \mathbf{y} &\sim N(\bar{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_\theta/n).
\end{aligned} \tag{10}$$

So using Jeffreys' prior $\boldsymbol{\Sigma}_\theta$ can be drawn from the Inverse-Wishart distribution.

Step 2 Draw $\boldsymbol{\mu}$ from its full conditional distribution. Using Jeffreys' prior results in $\boldsymbol{\mu} \,|\boldsymbol{\Sigma}_\theta, \mathbf{y} \sim N(\bar{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_\theta/n)$.

Step 3 Sampling $\mathbf{W}$ from its full conditional distribution proceeds as in Step 4 of the previous section, with the proper substitution of η_{ij}.

Step 4 Sampling $\mathbf{Z}$ from its full conditional distribution proceeds as in Step 1 of the previous section, with the proper substitution of η_{ij}.

Step 5 To draw from the conditional distribution of $\boldsymbol{\theta}$, an orthogonal standardized ability $\boldsymbol{\theta}^o$ is defined. So the elements of $\boldsymbol{\theta}_i^o = (\theta_{i1}^o, ...\theta_{iQ}^o)^T$ have independent standard normal distributions. Let $\mathbf{L}$ be the Cholesky decomposition of $\boldsymbol{\Sigma}_\theta$, that is, $\boldsymbol{\Sigma}_\theta = \mathbf{L}\mathbf{L}^T$. Define the orthogonal ability vector for person i, $\boldsymbol{\theta}_i^o = \mathbf{L}^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\mu})$. Now η_{ij} can be written as

$$\begin{aligned}\eta_{ij} &= \sum_{q=1}^{Q}(\alpha_{iq}\theta_{qj}) - \beta_j \\ &= \sum_{q=1}^{Q}(\alpha_{iq}\sum_{h=1}^{Q} L_{hq}\theta_{qj}^o + \mu_q) - \beta_j\end{aligned}$$

and in matrix notation as

$$\boldsymbol{\eta}_i = \mathbf{A}\mathbf{L}\mathbf{L}^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\mu} + \boldsymbol{\mu}) - \boldsymbol{\beta} = \mathbf{A}(\mathbf{L}\boldsymbol{\theta}_i^o + \boldsymbol{\mu}) - \boldsymbol{\beta},$$

with $\boldsymbol{\eta}_i$ and $\boldsymbol{\beta}$ vectors of length k, $\boldsymbol{\theta}_i$, $\boldsymbol{\theta}_i^o$ and $\boldsymbol{\mu}$ vectors of length Q and $\mathbf{A}$ a $k \times Q$ matrix with elements α_{jq}. The ability parameters $\boldsymbol{\theta}_i^o$ has a full posterior density given by

$$p(\boldsymbol{\theta}_i \,|\boldsymbol{\xi}, \mathbf{Z}, \mathbf{y}, \mathbf{W}) \propto \phi(\boldsymbol{\theta}_i^o; \mathbf{0}, \mathbf{I}) \prod_{j=1}^{k} \phi(Z_{ij}; \eta_{ij}, 1).$$

This can be written as

$$\mathbf{Z}_i + \boldsymbol{\beta} - \mathbf{A}\boldsymbol{\mu} = \mathbf{B}\boldsymbol{\theta}_i^o + \boldsymbol{\varepsilon}_i,$$

where $\mathbf{B} = \mathbf{A}\mathbf{L}$ and $\boldsymbol{\varepsilon}_i$ is a vectors elements ε_{ij}, which are iid $N(0,1)$.It then follows that

$$\boldsymbol{\theta}_i^o \text{ is distributed } N\left((\mathbf{I} + \boldsymbol{\Sigma}^{-1})^{-1}\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\theta}}_i^o, \quad (\mathbf{I} + \boldsymbol{\Sigma}^{-1})^{-1}\right),$$

with $\hat{\boldsymbol{\theta}}_i^o$ the common least squares estimate $\hat{\boldsymbol{\theta}}_i^o = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T(\mathbf{Z}_i + \boldsymbol{\beta} - \mathbf{A}\boldsymbol{\mu})$ and $\boldsymbol{\Sigma} = (\mathbf{B}^T\mathbf{B})^{-1}$. Now $\boldsymbol{\theta}_i$ can be obtained by the transformation $\mathbf{L}\boldsymbol{\theta}_i^o + \boldsymbol{\mu} = \boldsymbol{\theta}_i$.

Step 6 Conditional on $\mathbf{Z}$ and $\boldsymbol{\theta}$, the posterior distributions of $\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_k$ are independent with

the density of $\boldsymbol{\xi}_j$ given by

$$p(\boldsymbol{\xi}_j \,|W, \boldsymbol{\theta}, \mathbf{Z}, \mathbf{y}) \propto \prod_{i=1}^{n} \phi(Z_{ij}; \sum_{q=1}^{Q} \alpha_{jq}\theta_{iq} - \beta_j, 1) \prod_{q=1}^{Q} \mathbf{I}(\alpha_{jq} > 0).$$

Define $\mathbf{X} = (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_q, ..., \boldsymbol{\theta}_Q, -\mathbf{1})$ with $\boldsymbol{\theta}_q = (\theta_{1q}, ...\theta_{nq})^T$ and $-\mathbf{1}$ is an n dimensional column vector with elements -1. Conditional on $\boldsymbol{\theta}$, $\mathbf{Z}_j = (Z_{1j}, ..., Z_{nj})^T$ satisfies the linear model

$$\mathbf{Z}_j = \mathbf{X}\boldsymbol{\xi}_j + \varepsilon_j,$$

where $\varepsilon_j = (\varepsilon_{1j}, ..., \varepsilon_{ij}, ..., \varepsilon_{nj})^T$ and the ε_{ij} are iid $N(0, 1)$. It then follows that

$$\boldsymbol{\xi}_j \,|\boldsymbol{\theta}, \mathbf{Z}, \mathbf{y} \text{ is distributed } N(\hat{\boldsymbol{\xi}}_j, (\mathbf{X}^T\mathbf{X})^{-1}) \prod_{q=1}^{Q} \mathbf{I}(\alpha_{jq} > 0),$$

where $\hat{\boldsymbol{\xi}}_j = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Z}_j$.

Step 7 Sampling γ_j from its full conditional distribution proceeds as in Step 5 of the previous section.

As above, these steps can be used to iteratively generate a number of parameter sequences and convergence can be evaluated by comparing the between and within sequences variance. Finding starting values for the procedure may be obtained from NOHARM or TESTFACT, but an alternative approach will be sketched in the next section.

Above, it was already mentioned that there are several approaches to identify the model. Consider the model with ability parameters $\boldsymbol{\theta}_i$ and identifying restrictions $\boldsymbol{\mu} = \mathbf{0}$, $\alpha_{jq} = 1$, if $j = q$, and $\alpha_{jq} = 0$, if $j \neq q$, for $j = 1, ..., Q$ and $q = j + 1, ..., Q$, on one hand, and the model with orthogonal ability parameters $\boldsymbol{\theta}_i^o$ and identifying restrictions $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma}_\theta = \mathbf{I}$, and $\alpha_{jq}^o = 0$, for $j = 1, ..., Q - 1$ and $q = j + 1, ..., Q$, on the other hand. Let $\mathbf{A}$ and $\mathbf{A}^o$ be the matrices of factor loadings for the former and latter parameterization, respectively. Analogously to step 5, $\boldsymbol{\theta}_i$ can be transformed to $\boldsymbol{\theta}_i^o$ by $\boldsymbol{\theta}_i^o = \mathbf{L}^{-1}\boldsymbol{\theta}_i$, where $\mathbf{L}$ is the Cholesky decomposition of $\boldsymbol{\Sigma}_\theta$. Because $\mathbf{L}$ is lower triangular and $\mathbf{A}\boldsymbol{\theta}_i = \mathbf{A}\mathbf{L}\boldsymbol{\theta}_i^o = \mathbf{A}^o\boldsymbol{\theta}_i^o$, the restrictions $\alpha_{jq} = 1$, if $j = q$, and $\alpha_{jq} = 0$, if $j \neq q$, for $j = 1, ..., Q$ and $q = j + 1, ..., Q$, are transformed into restrictions $\alpha_{jq}^o = 0$, for $j = 1, ..., Q - 1$ and $q = j + 1, ..., Q$. On the other hand, defining the lower triangular matrix $\mathbf{F}$ as the first Q rows of $\mathbf{A}^o$ and applying $\boldsymbol{\theta}_i = \mathbf{F}\boldsymbol{\theta}_i^o$, results in $\boldsymbol{\Sigma}_{\theta=}\mathbf{F}^{T^{-1}}\mathbf{F}^{-1}$ and $\mathbf{A} = \mathbf{A}^o\mathbf{F}^{-1}$, which in turn produces restrictions $\alpha_{jq} = 1$, if $j = q$, and

$\alpha_{jq} = 0$, if $j \neq q$, for $j = 1, ..., Q$ and $q = j + 1, ..., Q$. So the two parameterizations of the model are easily interchanged.

As was already mentioned above, the procedure easily generalizes to the case of multiple groups and incomplete designs. In the case of multiple groups, the mean and the covariance matrix of the ability distribution of one of the groups can be set equal to zero and the identity matrix, respectively. So for this group the first two steps are not performed, these steps are only performed to generate means and covariance matrices for the remaining groups. Restricting the mean and the covariance of one of the ability distributions suffices for fixing the location and the scale of the latent continuum. However, this is not sufficient to identify the model, because the model probabilities (1) are left unaltered under a permutation of the ability dimensions. Therefore, additional constraints $\alpha_{jq} = 0$,for $j = 1, ..., Q - 1$ and $q = j + 1, ..., Q$ are needed to completely identify the model. Steps 3 and 4 are performed per item, so generalization to multiple groups and incomplete designs only entails that the variables W_{ij} and Z_{ij} are performed in the cases where person i responded to item j. Step 5 is performed per person, so here both the group to which the person belongs and the items made should be taken into account. Finally, the steps 6 and 7 for generating item parameters should take into account the persons actually confronted with the items.

Subscale Factor Analysis

In the previous section it was argued that the model can be identified by setting a mean and a covariance matrix equal to zero and the identity matrix, respectively, and introducing the constraints $\alpha_{jq} = 0$,for $j = 1, ..., Q-1$ and $q = j+1, ..., Q$. These are the constraints also used in the exploratory option of NOHARM (Fraser, 1988). In this approach, the model is identified by assuming that the responses on the first item are uniquely determined by the first ability dimension, the responses on the second items are uniquely determined by a mixture of the first and the second ability dimension, and so forth. In general, these identification restriction will be of little help in providing an interpretation for the ability dimensions. Therefore, in a exploratory factor analysis, the factor solution is usually visually or analytically rotated. Often, the rotation scheme is devised to approximate Thurstone's simple structure criterium (Thurstone, 1947), where the factor loadings are split into two groups, the elements of the one tending to zero and the elements of the other toward unity. The approach suggested in this section has much in common with this approach: the idea is to identify the dimensions with subscales of items loading on one dimension only. The idea is to either identify these $S < Q$ subscales on a priori

grounds, or to identify these subscales using an iterative search based on fitting a unidimensional IRT model by discarding items from the scale. In the latter case, there are two reasons for choosing the Rasch model as the unidimensional IRT model. Firstly, the theory evaluation of fit to the Rasch model is well founded and the asymptotic distributions of the many fit statistics are well known (Andersen, 1973, Martin Löf, 1973, 1974, Kelderman, 1984, 1989, Molenaar, 1983, Glas, 1988, Glas & Verhelst, 1989, 1995). However, this argument is not completely overwhelming, because procedures for evaluation of model fit are also available in the general MML framework (Yen, 1981, 1984, Mislevy & Bock, 1990, Reiser, 1996, Glas, 1998) and the framework of non-parametric IRT (Rosenbaum, 1984, Holland & Rosenbaum, 1986, Stout, 1987, 1990 and Junker, 1991). A second and more important reason relates to the intuitive concept of a measurement scale: a continuum where the order of the outcomes of a measurement does not depend on the objects positions on the scale. Translated to a test with dichtomous items: the probabilities of a correct response on items referring to a scale is independent of the position on the scale. This intuitive requirement is formalized by Rasch (1967) and leads to the Rasch model (Fischer, 1995).

These considerations lead to the following heuristic:

(1) Set counter $s = 1$
(2) Estimate the parameters of the items in the target set;
(3) Compute fit statistics and remove poorly fitting items from the target set to the set "removed items";
(4) If the overall model fit is unacceptable, goto 2;
(5) If $s = S$ then stop else "removed items" becomes the target set, $s = s + 1$ and goto 2.

In the first iteration of the selection process ($s = 1$), the target distribution can be set to the complete test, and the inner loop (steps (2)-(4)) is repeated until an acceptable unidimensional subscale is found. In the next iteration ($s = 2$), the nest subscale is selected from the remaining items, and so forth, until $s = S$. Detailing this procedure further, the item parameters estimates of step (2) can, for instance, be computed using conditional maximum likelihood (see, for instance, Molenaar, 1995), which has the advantage that these estimates do not depend on an assumption with respect to the ability distribution. These estimates can be computed using RSP (Glas & Ellis, 1993), which program also computes the so-called R_1 and R_0 statistics (Glas, 1988) that can be used in step (3) and (4). The R_1-test can be used for evaluation of the form of the ICC's and the R_2-test for evaluation of local independence.

After identification of S sets of scaled items, there will usually be a set of remaining

items that may load an all ability dimensions. Of course, the number of items in this set depends on the size of the critical regions used in step (3) and (4). Next, restrictions will be imposed on the matrix of factor loadings A to reflect the structure of the subscales found, that is, if item j belongs to subscale q, $\alpha_{jq} = 1$, and $\alpha_{jq'} = 0$, for $q' = 1, ..., Q,\ q' \neq q$. Finally the remaining dimensions $q = S + 1, , ..., Q$ are identified using $\alpha_{j'q} = 1$,and $\alpha_{j'q'} = 0$, for some item j' not belonging to a subscale, and $q' = 1, ..., Q,\ q' \neq q$.

Examples

In this section, two studies of the performance of the MCMC method will be presented, one for the case of the 3PNO and one for the case of the (3+2)PNO. For the first study, data were simulated using the 3PNO parameter values given in Table 1. These parameters are a subset of the items parameters of a real, and therefore realistic data set (ACT, 1997). Using these parameters, ability parameters for $n = 1000$ simulees were drawn from a standard normal ability distribution and for every simulee a response patterns was generated using the 3PNO. Next, the items parameters were estimated using the MCMC procedure described above and Bilog-MG. In both procedures the same Beta(a, b) prior on γ was used. The values for a and b were set equal to $20 * \gamma_{true} + 1$ and $20 * (1 - \gamma_{true}) + 1$, respectively, the values γ_{true} are listed in Table 1. Further, in the Bilog analysis a log normal (0,0.5) prior on α was used. The MCMC procedure had a run length of 30000 iterations. The starting values used in the MCMC procedure were, $\alpha_i = 1, \beta_i = 0, \gamma_i = \gamma_{i,true}$ for all items i, and $\theta_j = 0$ for all persons j. In a real estimation situation those starting values can be improved by the use of, for example, Bilog-MG estimated parameter values. From examining the plots of sampled parameter values, it was concluded that a burn-in period of 1000 iterations was sufficient. These first 1000 iterations were discarded.

Insert Table 1 about here

The mean of the generated posterior distribution and posterior standard deviation of the parameters issued from the MCMC procedure are given in Table 2. In Table 3, the MML estimated parameter values and their standard errors issued from Bilog-MG are given. The β parameters in Table 3 are transformed $\beta = \alpha\beta_{\text{bilog}}$, because in Bilog-MG the parameterization $\Phi(\alpha(\theta - \beta))$ is used. Comparing the Tables 1, 2 and 3, it can be seen that parameter recovery of the MCMC and Bilog-MG procedures are comparable.

Insert Table 2 and Table 3 about here

To evaluate the performance of the MCMC procedure in combination with the 3-dimensional (3+2)PNO model, 2000 response patterns were simulated with the item parameters given in Table 4, and a multivariate normal ability distribution with $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}$. Note that 25 of the 28 items only load on a single dimension. The items loading on the same dimension form unidimensional subscales. Again, the MCMC procedure was performed with a run length of 30000. As with the 3PNO model, from examination of the plots of sampled parameter values it could be seen that a burn-in period of length 1000 was sufficient. In the MCMC procedure $\boldsymbol{\alpha}$ is generally restricted to imply $\boldsymbol{\alpha} > 0$ for all items. In the current run this prior was dropped for $\boldsymbol{\alpha}_i$, with $i \leqslant 3$. Above it was shown that these parameters play a role in the change between a model representation with and without covariance. A restriction on the values of $\boldsymbol{\alpha}_i$, $i \leqslant 3$ would imply an undesirable restriction on the covariance. Because no additional priors were used on α and β an informative prior on γ proved necessary to support convergence of the procedure. Therefore, a Beta(a, b) prior with $a = 1000 * \gamma_{true} + 1$ and $b = 1000 * (1 - \gamma_{true}) + 1$ was used. The true parameter values γ_{true} are listed in Table 4. Further, the starting values for the item parameters were $\alpha_{iq} = 1/i$ for $i \leqslant Q$ and $\alpha_{iq} = 1/Q$ otherwise, and $\beta_i = 0$, and $\gamma_i = \gamma_{i,true}$.The starting values for the person parameters were $\theta_{jq} = 0$.

Insert Table 4 to 8 about here

In Table 5 the results of the MCMC procedure are summarized. For each item the mean posterior and posterior standard error are presented. In Table 6 the result are given of the same MCMC procedure as described above only now the parameterization with a covariance matrix is chosen and the prior $\alpha > 0$ is dropped. Those modifications enable comparison of the results with the results of NOHARM (Table 7) and TESTFACT (Table 8). Both NOHARM and TESTFACT use fixed values of γ. the results of TESTFACT are transformed to make them comparable with the results in Table 6 and Table 7.From inspection of these tables it can be seen that recovery of the factor structure is quite good.

References

Ackerman, T.A. (1996a). Developments in multidimensional item response theory. *Applied Psychological Measurement 20,* 309-310.

Ackerman, T.A. (1996b). Graphical representation of multidimensional item response theory analyses. *Applied Psychological Measurement 20,* 311-329.

ACT, Inc. (1997). *ACT Assessment Technical Manual.* Iowa City, IA: ACT, Inc.

Albert, J.H. (1992). Bayesian estimation of normal ogive item response functions using Gibbs sampling. *Journal of Educational Statistics, 17,* 251-269.

Andersen, E.B. (1973). A goodness of for test for the Rasch model. *Psychometrika, 38,* 123-140.

Bock, R.D., Gibbons, R.D. & Muraki, E. (1988). Full-information factor analysis. *Applied Psychological Measurement 12,* 261-280.

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

Box, G. and Tiao, G. (1973). *Bayesian Inference in Statistical Analysis.* Addison-Wesley Publishing Company, Reading, Massachusetts.

Fischer, G.H. (1995). Derivations of the Rasch model. In: G.H.Fischer & I.W.Molenaar (eds.). *Rasch models: foundations, recent developments and applications.* (pp.39-52). New York: Springer.

Fraser, C. (1988). NOHARM: *A Computer Program for Fitting Both Unidimensional and Multidimensional Normal Ogive Models of Latent Trait Theory.* NSW: University. of New England.

Gelfand, A.E. & Smith, A.F.M.(1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association 85,* 398-409.

Gelman, A, Carlin, J.B., Stern, H.S. & Rubin, D.B. (1995). *Bayesian Data Analysis.* London: Chapman and Hall.

Glas, C.A.W. (1988). The derivation of some tests for the Rasch model from the multinomial distribution. *Psychometrika, 53,* 525-546.

Glas, C.A.W. (1998) Detection of differential item functioning using Lagrange multiplier tests. *Statistica Sinica, 8, vol. 1.* 647-667.

Glas C.A.W. & Ellis, J.L. (1993) *RSP, Rasch Scaling Program, computer program and user's manual.* Groningen: ProGAMMA.

Glas, C.A.W., and Verhelst, N.D. (1989). Extensions of the partial credit model.

Psychometrika, 54, 635-659.

Glas, C.A.W. & Verhelst, N.D. (1995). Tests of fit for polytomous Rasch models. In G.H.Fischer & I.W.Molenaar (eds.). *Rasch models: foundations, recent developments and applications.* (pp.325-352). New York: Springer.

Holland, P.W. & Rosenbaum, P.R. (1986). Conditional association and unidimensionality in monotone latent variable models. *Annals of Statistics, 14,* 1523-1543.

Junker, B. (1991). Essential independence and likelihood-based ability estimation for polytomous items. *Psychometrika, 56,* 255-278.

Kelderman, H. (1984). Loglinear RM tests. *Psychometrika,49,* 223-245.

Kelderman, H. (1989). Item bias detection using loglinear IRT. *Psychometrika,54,* 681-697.

Lawley, D.N. (1943). On problems connected with item selection and test construction. *Proceedings of the Royal Society of Edinburgh, 61,* 273-287.

Lawley, D.N. (1944). The factorial analysis of multiple test items. *Proceedings of the Royal Society of Edinburgh, 62-A,* 74-82.

Lord, F.M. (1952). A theory of test scores. *Psychometric Monograph 7.*

Lord, F.M. (1953a). An application of confidence intervals and of maximum likelihood to the estimation of an examinee's ability. *Psychometrika, 18,* 57-75.

Lord, F.M. (1953b). The relation of test score to the trait underlying the test. *Educational and Psychological Measurement, 13,* 517-548.

Lord, F.M. and Novick, M.R. (1968). *Statistical theories of mental test scores.* Reading: Addison-Wesley.

Martin-Löf, P. (1973). *Statistika Modeller Anteckningar från seminarier Lasåret 1969-1970, utardeltade av Rolf Sunberg. Obetydligt ändrat nytryck, oktober 1973.* Stockholm: Institutet för Försäkringsmatematik och Matematisk Statistik vid Stockholms Universitet.

Martin Löf, P. (1974). The notion of redundancy and its use as a quantitative measure if the discrepancy between a statistical hypothesis and a set of observational data. *Scandinavian Journal of Statistics, 1,* 3-18.

McDonald, R.P. (1967). Nonlinear factor analysis. *Psychometric monographs, No.15.*

McDonald, R.P. (1997). Normal-ogive multidimensional model. In: W.J.van der Linden and R.K.Hambleton (eds.). *Handbook of Modern Item Response Theory.* (pp.257-269). New York: Springer.

Mislevy, R.J. (1986). Bayes modal estimation in item response models. *Psychometrika, 51,* 177-195.

Mislevy, R.J. & Bock, R.D. (1990). *PC-BILOG. Item analysis and test scoring with binary logistic models.* Scientific Software: Mooresville.

Molenaar, I.W. (1983). Some improved diagnostics for failure in the Rasch model. *Psychometrika, 48,* 49-72.

Molenaar, I.W. (1995). Estimation of Item Parameters. In: G.H.Fischer & I.W.Molenaar (eds.). *Rasch models: foundations, recent developments and applications.* New York: Springer.

Reckase, M.D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement, 9,* 401-412.

Reckase, M.D. (1997). A linear logistic multidimensional model for dichotomous item response data. In: W.J.van der Linden and R.K.Hambleton (eds.). *Handbook of Modern Item Response Theory.* (pp.271-286). New York: Springer.

Rasch, G. (1977). *On specific objectivity: an attempt at formalizing the request for generality and validity of scientific statements.* In M. Blegvad (Ed.) *The Danish Yearbook of Philosophy* (pp. 58-94). Copenhagen: Munksgaard.

Reiser, M. (1996). Analysis of residuals for the multinomial item response model. *Psychometrika, 61,* 509-528.

Rosenbaum, P.R. (1984). Testing the conditional independence and monotonicity assumptions of item response theory. *Psychometrika, 49,* 425-436.

Stout, W.F. (1987). A nonparametric approach for assessing latent trait dimensionality. *Psychometrika, 52,* 589-617.

Stout, W.F. (1990). A new item response theory modeling approach with applications to unidimensional assessment and ability estimation. *Psychometrika, 55,* 293-326.

Thurstone, L.L. (1947). *Multiple Factor Analysis.* Chicago: University of Chicago Press.

Wilson, D.T., Wood, R., & Gibbons, R. (1991) *TESTFACT: Test scoring, Item statistics, and Item Factor Analysis.* (Computer Software). Chicago: Scientific Software International, Inc.

Yen, W.M. (1981). Using simultaneous results to choose a latent trait model. *Applied Psychological Measurement, 5,* 245-262.

Yen, W.M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 8,* 125-145.

Zimowski, M.F., Muraki, E., Mislevy, R.J. & Bock, R.D. (1996). *Bilog MG: Multiple-Group IRT Analysis and Test Maintenance for Binary Items.* Chicago: Scientific Software International, Inc.

Table 1
True item parameter values

ITEM	α	β	γ
1	0.642	-1.619	0.19
2	0.806	-1.533	0.15
3	0.956	-1.292	0.11
4	0.972	-1.061	0.14
5	1.045	-0.245	0.37
6	0.834	-0.264	0.14
7	0.614	0.023	0.17
8	0.796	0.213	0.10
9	1.171	-0.669	0.19
10	1.514	0.480	0.31

Table 2
Recovery of 3PNO parameters by MCMC

ITEM	α	β	γ	$\sigma(\alpha)$	$\sigma(\beta)$	$\sigma(\gamma)$
1	0.564	-1.607	0.220	0.111	0.114	0.087
2	0.818	-1.500	0.187	0.125	0.119	0.081
3	0.928	-1.185	0.161	0.142	0.112	0.079
4	0.826	-0.829	0.173	0.125	0.111	0.080
5	0.927	-0.399	0.325	0.202	0.166	0.079
6	0.975	-0.369	0.113	0.130	0.095	0.049
7	0.629	0.038	0.214	0.151	0.200	0.082
8	1.374	0.682	0.221	0.444	0.332	0.048
9	1.093	-0.691	0.160	0.173	0.107	0.061
10	1.027	0.273	0.279	0.234	0.198	0.057

Table 3
Recovery of 3PNO parameters by BILOG-MG

ITEM	α	β	γ	$\sigma(\alpha)$	$\sigma(\beta)$	$\sigma(\gamma)$
1	0.653	-1.736	0.191	0.113	0.165	0.088
2	0.865	-1.600	0.150	0.125	0.161	0.080
3	0.919	-1.249	0.119	0.122	0.143	0.074
4	0.809	-0.869	0.145	0.108	0.121	0.077
5	0.879	-0.445	0.311	0.152	0.172	0.082
6	0.930	-0.416	0.084	0.116	0.099	0.047
7	0.582	-0.024	0.195	0.106	0.088	0.081
8	1.139	0.499	0.199	0.279	0.150	0.049
9	1.029	-0.730	0.131	0.135	0.140	0.061
10	0.942	0.203	0.267	0.209	0.150	0.063

Table 4
True item parameters (3+2)PNO

ITEM	α_1	α_2	α_3	β	γ
1	1.000	.000	.000	.000	0.19
2	.000	1.000	.000	.000	0.15
3	.000	.000	1.000	.000	0.11
4	1.000	.000	.000	.500	0.14
5	1.000	.000	.000	1.000	0.37
6	1.000	.000	.000	-1.000	0.14
7	1.000	.000	.000	-1.000	0.17
8	1.000	.000	.000	.500	0.10
9	1.000	.000	.000	1.000	0.19
10	1.000	.000	.000	-1.000	0.31
11	1.000	.000	.000	-1.000	0.20
12	.000	1.000	.000	.500	0.20
13	.000	1.000	.000	.250	0.20
14	.000	1.000	.000	-.250	0.20
15	.000	1.000	.000	-.500	0.20
16	.000	1.000	.000	.500	0.20
17	.000	1.000	.000	.250	0.20
18	.000	1.000	.000	-.250	0.20
19	.000	1.000	.000	-.500	0.20
20	.000	.000	1.000	1.000	0.20
21	.000	.000	1.000	.000	0.20
22	.000	.000	1.000	1.000	0.20
23	.000	.000	1.000	1.000	0.20
24	.000	.000	1.000	.000	0.20
25	.000	.000	1.000	1.000	0.20
26	.200	.800	.000	.000	0.20
27	.200	.200	.600	1.000	0.20
28	.800	.400	.200	-1.000	0.20

Table 5
Recovery of 3PNO parameters by MCMC procedure with prior on α

ITEM	α_1	α_2	α_3	β	γ	$\sigma(\alpha_1)$	$\sigma(\alpha_2)$	$\sigma(\alpha_3)$	$\sigma(\beta)$	$\sigma(\gamma)$
1	1.148	0.000	0.000	0.023	0.188	0.109			0.059	0.012
2	0.138	0.927	0.000	-0.011	0.150	0.050	0.074		0.049	0.011
3	-0.027	0.046	1.007	0.046	0.111	0.053	0.051	0.090	0.050	0.010
4	0.919	0.034	0.046	0.504	0.144	0.093	0.027	0.035	0.068	0.011
5	0.854	0.088	0.110	0.923	0.378	0.162	0.061	0.072	0.165	0.015
6	0.881	0.048	0.029	-0.923	0.142	0.079	0.034	0.025	0.055	0.011
7	1.061	0.040	0.062	-1.072	0.172	0.098	0.032	0.043	0.067	0.012
8	0.989	0.066	0.028	0.463	0.105	0.098	0.042	0.024	0.066	0.010
9	1.272	0.075	0.129	1.258	0.190	0.213	0.053	0.074	0.197	0.011
10	0.982	0.031	0.057	-1.067	0.313	0.110	0.027	0.042	0.073	0.015
11	0.977	0.047	0.102	-1.139	0.201	0.093	0.036	0.055	0.068	0.013
12	0.034	1.083	0.132	0.652	0.202	0.029	0.109	0.064	0.086	0.012
13	0.040	1.015	0.102	0.294	0.202	0.032	0.096	0.055	0.066	0.013
14	0.043	1.034	0.035	-0.270	0.203	0.033	0.088	0.029	0.051	0.013
15	0.078	0.943	0.050	-0.477	0.201	0.045	0.080	0.037	0.051	0.013
16	0.094	1.088	0.130	0.556	0.203	0.053	0.114	0.062	0.086	0.012
17	0.038	1.089	0.180	0.334	0.205	0.031	0.106	0.065	0.073	0.013
18	0.061	0.984	0.076	-0.226	0.203	0.041	0.081	0.047	0.052	0.013
19	0.035	1.020	0.028	-0.459	0.202	0.028	0.085	0.024	0.051	0.013
20	0.062	0.102	1.025	1.041	0.206	0.047	0.062	0.141	0.139	0.013
21	0.078	0.054	1.330	0.039	0.203	0.051	0.041	0.143	0.068	0.013
22	0.039	0.062	1.135	1.146	0.212	0.035	0.048	0.209	0.208	0.014
23	0.056	0.086	1.207	1.221	0.209	0.046	0.059	0.207	0.203	0.013
24	0.105	0.040	1.032	0.009	0.204	0.053	0.032	0.102	0.058	0.013
25	0.070	0.061	1.356	1.355	0.213	0.055	0.049	0.286	0.276	0.013
26	0.253	0.782	0.040	-0.015	0.202	0.054	0.071	0.031	0.051	0.013
27	0.096	0.182	0.607	1.069	0.200	0.057	0.068	0.089	0.102	0.012
28	0.716	0.391	0.201	-0.933	0.201	0.069	0.057	0.056	0.053	0.013

Table 6

Recovery of 3PNO parameters by MCMC procedure without prior on α

ITEM	α_1	α_2	α_3	β	γ	$\sigma(\alpha_1)$	$\sigma(\alpha_2)$	$\sigma(\alpha_3)$	$\sigma(\beta)$	$\sigma(\gamma)$
1	1.000	0.000	0.000	0.004	0.184				0.052	0.012
2	0.000	1.000	0.000	-0.006	0.152				0.049	0.011
3	0.000	0.000	1.000	0.045	0.111				0.049	0.010
4	0.923	0.025	-0.042	0.506	0.144	0.100	0.064	0.067	0.068	0.011
5	0.890	0.104	0.056	0.956	0.380	0.186	0.095	0.098	0.182	0.015
6	0.896	0.064	-0.093	-0.930	0.142	0.084	0.062	0.064	0.054	0.011
7	1.071	0.043	-0.017	-1.071	0.172	0.105	0.070	0.073	0.068	0.012
8	0.991	0.103	-0.098	0.465	0.105	0.105	0.065	0.067	0.064	0.010
9	1.310	0.104	0.066	1.287	0.190	0.238	0.096	0.099	0.210	0.012
10	1.020	-0.010	-0.035	-1.079	0.313	0.119	0.072	0.077	0.074	0.015
11	0.985	0.052	0.046	-1.138	0.201	0.098	0.069	0.072	0.068	0.012
12	-0.182	1.137	0.092	0.673	0.204	0.082	0.134	0.080	0.099	0.012
13	-0.130	1.036	0.070	0.294	0.202	0.069	0.100	0.071	0.066	0.012
14	-0.124	1.059	-0.059	-0.271	0.202	0.067	0.095	0.070	0.052	0.013
15	-0.044	0.967	-0.009	-0.474	0.202	0.062	0.084	0.066	0.050	0.013
16	-0.047	1.113	0.096	0.552	0.203	0.074	0.120	0.078	0.084	0.012
17	-0.156	1.111	0.152	0.333	0.205	0.073	0.119	0.078	0.073	0.013
18	-0.076	1.018	0.034	-0.224	0.203	0.064	0.092	0.069	0.053	0.013
19	-0.145	1.058	-0.087	-0.458	0.203	0.069	0.095	0.071	0.052	0.013
20	0.024	0.052	1.030	1.042	0.206	0.080	0.084	0.151	0.144	0.013
21	0.082	-0.018	1.341	0.037	0.203	0.079	0.081	0.154	0.069	0.013
22	-0.052	-0.041	1.139	1.132	0.211	0.083	0.086	0.166	0.159	0.012
23	-0.004	0.007	1.226	1.228	0.209	0.094	0.099	0.204	0.198	0.013
24	0.125	-0.037	1.038	0.010	0.204	0.067	0.069	0.109	0.059	0.013
25	0.016	-0.062	1.320	1.310	0.212	0.098	0.102	0.193	0.183	0.012
26	0.159	0.822	-0.033	-0.014	0.202	0.061	0.077	0.064	0.051	0.013
27	0.066	0.166	0.599	1.062	0.200	0.071	0.074	0.091	0.104	0.012
28	0.692	0.433	0.159	-0.938	0.201	0.075	0.066	0.064	0.053	0.013

Covariance			Standard error covariance		
0.975			0.056		
0.016	0.938		0.036	0.054	
-0.005	0.017	0.968	0.036	0.036	0.058

Table 7
NOHARM estimates (3+2)PNO

ITEM	α_1	α_2	α_3	β
1	1.000	0.000	0.000	-0.038
2	0.000	1.000	0.000	0.014
3	0.000	0.000	1.000	-0.042
4	0.686	0.007	-0.014	-0.480
5	0.577	0.075	-0.006	-0.801
6	0.671	0.042	-0.056	0.922
7	0.783	0.015	0.007	1.049
8	0.734	0.064	-0.081	-0.436
9	0.958	0.021	0.130	-1.230
10	0.686	-0.040	0.036	1.025
11	0.765	0.010	0.081	1.143
12	-0.117	1.128	0.113	-0.634
13	-0.073	1.068	0.023	-0.284
14	-0.114	1.070	-0.061	0.280
15	-0.042	0.978	-0.032	0.478
16	-0.037	1.108	0.066	-0.526
17	-0.096	1.082	0.125	-0.297
18	-0.069	1.033	0.032	0.237
19	-0.107	1.055	-0.085	0.463
20	0.032	0.019	0.956	-0.962
21	0.072	-0.070	1.293	-0.020
22	-0.040	-0.056	0.922	-0.942
23	-0.033	0.020	1.061	-1.066
24	0.098	-0.062	0.968	0.007
25	0.014	-0.024	1.022	-1.043
26	0.127	0.818	-0.019	0.024
27	0.065	0.161	0.645	-1.081
28	0.498	0.389	0.195	0.918
Covariance				
1.638				
0.062	0.886			
-0.070	0.042	0.995		

Table 8
TESTFACT estimates (3+2)PNO

ITEM	α_1	α_2	α_3	β
1	1.000	0.000	0.000	0.013
2	0.000	1.000	0.000	-0.013
3	0.000	0.000	1.000	0.046
4	0.782	-0.004	-0.041	0.467
5	0.678	0.082	0.022	0.809
6	0.760	0.029	-0.083	-0.928
7	0.883	-0.005	-0.021	-1.051
8	0.840	0.071	-0.106	0.429
9	1.035	0.033	0.073	1.158
10	0.880	-0.060	-0.027	-1.086
11	0.818	-0.005	0.042	-1.116
12	-0.181	1.203	0.097	0.651
13	-0.131	1.084	0.040	0.283
14	-0.132	1.099	-0.069	-0.271
15	-0.066	1.014	-0.010	-0.470
16	-0.061	1.146	0.112	0.529
17	-0.156	1.130	0.142	0.307
18	-0.085	1.024	0.023	-0.225
19	-0.158	1.097	-0.090	-0.455
20	0.037	0.063	1.008	0.991
21	0.095	-0.032	1.251	0.022
22	-0.033	-0.069	1.047	1.011
23	0.004	0.016	1.031	1.035
24	0.135	-0.053	0.972	-0.004
25	0.046	-0.061	1.060	1.066
26	0.109	0.836	-0.041	-0.025
27	0.076	0.142	0.587	1.039
28	0.572	0.409	0.155	-0.924
Covariance				
1.175				
0.081	0.754			
-0.032	0.037	0.890		

Titles of Recent Research Reports from the Department of Educational Measurement and Data Analysis. University of Twente, Enschede, The Netherlands.

RR-98-14 A.A. Béguin & C.A.W. Glas, *MCMC Estimation of Multidimensional IRT Models*

RR-98-13 E.M.L.A. van Krimpen-Stoop & R.R. Meijer, *Person Fit based on Statistical Process Control in an AdaptiveTesting Environment*

RR-98-12 W.J. van der Linden, *Optimal Assembly of Tests with Item Sets*

RR-98-11 W.J. van der Linden, B.P. Veldkamp & L.M. Reese, *An Integer Programming Approach to Item Pool Design*

RR-98-10 W.J. van der Linden, *A Discussion of Some Methodological Issues in International Assessments*

RR-98-09 B.P. Veldkamp, *Multiple Objective Test Assembly Problems*

RR-98-08 B.P. Veldkamp, *Multidimensional Test Assembly Based on Lagrangian Relaxation Techniques*

RR-98-07 W.J. van der Linden & C.A.W. Glas, *Capitalization on Item Calibration Error in Adaptive Testing*

RR-98-06 W.J. van der Linden, D.J. Scrams & D.L.Schnipke, *Using Response-Time Constraints in Item Selection to Control for Differential Speededness in Computerized Adaptive Testing*

RR-98-05 W.J. van der Linden, *Optimal Assembly of Educational and Psychological Tests, with a Bibliography*

RR-98-04 C.A.W. Glas, *Modification Indices for the 2-PL and the Nominal Response Model*

RR-98-03 C.A.W. Glas, *Quality Control of On-line Calibration in Computerized Assessment*

RR-98-02 R.R. Meijer & E.M.L.A. van Krimpen-Stoop, *Simulating the Null Distribution of Person-Fit Statistics for Conventional and Adaptive Tests*

RR-98-01 C.A.W. Glas, R.R. Meijer, E.M.L.A. van Krimpen-Stoop, *Statistical Tests for Person Misfit in Computerized Adaptive Testing*

RR-97-07 H.J. Vos, *A Minimax Sequential Procedure in the Context of Computerized Adaptive Mastery Testing*

RR-97-06 H.J. Vos, *Applications of Bayesian Decision Theory to Sequential Mastery Testing*

RR-97-05 W.J. van der Linden & Richard M. Luecht, *Observed-Score Equating as a Test Assembly Problem*

RR-97-04 W.J. van der Linden & J.J. Adema, *Simultaneous Assembly of Multiple Test Forms*

RR-97-03 W.J. van der Linden, *Multidimensional Adaptive Testing with a Minimum Error-Variance Criterion*

RR-97-02 W.J. van der Linden, *A Procedure for Empirical Initialization of Adaptive Testing Algorithms*

RR-97-01 W.J. van der Linden & Lynda M. Reese, *A Model for Optimal Constrained Adaptive Testing*

RR-96-04 C.A.W. Glas & A.A. Béguin, *Appropriateness of IRT Observed Score Equating*

RR-96-03 C.A.W. Glas, *Testing the Generalized Partial Credit Model*

RR-96-02 C.A.W. Glas, *Detection of Differential Item Functioning using Lagrange Multiplier Tests*

RR-96-01 W.J. van der Linden, *Bayesian Item Selection Criteria for Adaptive Testing*

RR-95-03 W.J. van der Linden, *Assembling Tests for the Measurement of Multiple Abilities*

RR-95-02 W.J. van der Linden, *Stochastic Order in Dichotomous Item Response Models for Fixed Tests, Adaptive Tests, or Multiple Abilities*

RR-95-01 W.J. van der Linden, *Some decision theory for course placement*

RR-94-17 H.J. Vos, *A compensatory model for simultaneously setting cutting scores for selection-placement-mastery decisions*

RR-94-16 H.J. Vos, *Applications of Bayesian decision theory to intelligent tutoring systems*

RR-94-15 H.J. Vos, *An intelligent tutoring system for classifying students into Instructional treatments with mastery scores*

RR-94-13 W.J.J. Veerkamp & M.P.F. Berger, *A simple and fast item selection procedure for adaptive testing*

RR-94-12 R.R. Meijer, *Nonparametric and group-based person-fit statistics: A validity study and an empirical example*

RR-94-10 W.J. van der Linden & M.A. Zwarts, *Robustness of judgments in evaluation research*

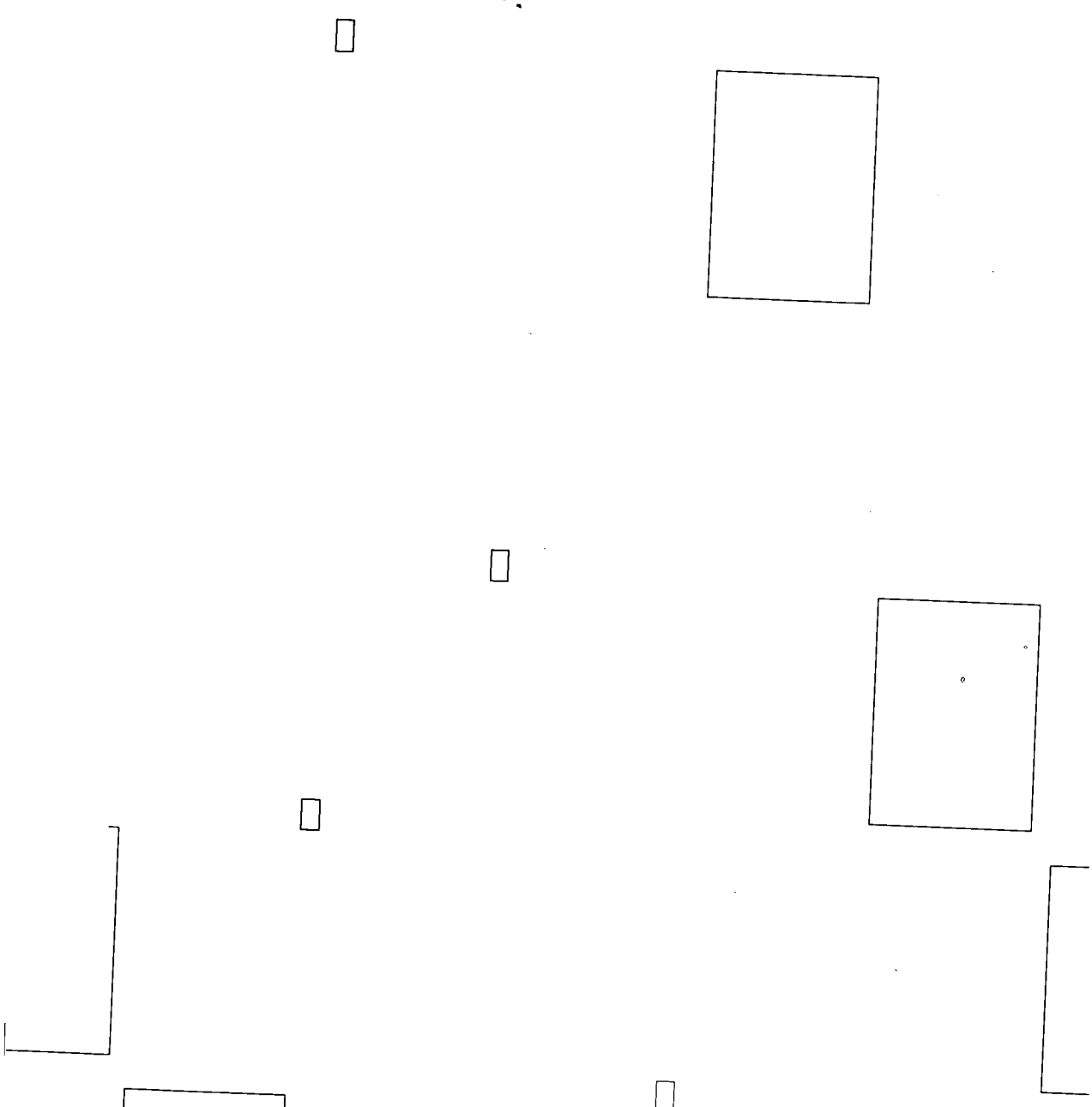
RR-94-9 L.M.W. Akkermans, *Monte Carlo estimation of the conditional Rasch model*

RR-94-8 R.R. Meijer & K. Sijtsma, *Detection of aberrant item score patterns: A review of recent developments*

RR-94-7 W.J. van der Linden & R.M. Luecht, *An optimization model for test assembly to match observed-score distributions*

...

Research Reports can be obtained at costs, Faculty of Educational Science and Technology, University of Twente, Mr. J.M.J. Nelissen, P.O. Box 217, 7500 AE Enschede, The Netherlands.



faculty of

EDUCATIONAL SCIENCE AND TECHNOLOGY

A publication by
The Faculty of Educational Science and Technology of the University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands