THE ESTIMATION OF POLYTOMOUS ITEM RESPONSE MODELS WITH MANY DIMENSIONS

Nikolai Volodin
Australia and New Zealand Banking Group

and

Raymond J. Adams
University of Melbourne and Australian Council for Educational Research

December 2002

The first author was funded by U. S. Department of Education Contract No. RN 93005001 and the second author was funded by NSF Grant RED-9255272.

Correspondence should be sent to Correspondence should be sent to Ray Adams, Department of Learning and Educational Development, Faculty of Education, University of Melbourne, Grattan Street, Parkville, Victoria, Australia, 3052. e-mail: r.adams@unimelb.edu.au

THE ESTIMATION OF POLYTOMOUS ITEM RESPONSE MODELS WITH MANY DIMENSIONS

ABSTRACT

Identification conditions and an improved estimation method for a *D*-dimensional mixed coefficients multinomial logit model are discussed. This model is a generalisation of the Adams and Wilson (1997) random coefficients multinomial logit and it can be used to fit multdimensional forms of a wide range of Rasch measurement models. The computational demands of the numerical integration required in fitting such models have limited previous implementations to three and perhaps four-dimensional problems (Glas, 1992; Adams, Wilson and Wang, 1997). This paper illustrates a Monte Carlo integration method that permits the estimation of models with much higher dimensionality. The example in this paper fits models of six dimensions.

1. INTRODUCTION.

The development of multidimensional item response models that can be practically employed in applied contexts is becoming an increasingly important psychometric problem. The increased popularity of performances assessment and the demands of a wide coverage of testing material within large scale testing programs are just two circumstances that are motivating the use and further development of item response models.

In the case of performance assessments; much of their motivation lies with the desire to obtain a richer array of information about student performances than can be typically obtained from multiple choice tests. Analysing such data with models that hypothesise one underlying latent ability is likely to be unsatisfactory both from substantive and statistical perspectives. Multidimensional item response models that can be routinely applied to complex performance data are clearly an important psychometric requirement. In the case of large scale testing programs, even the multiple choice components of such programs, there is a growing demand for student performances to be reported on many sub-scales. The National Assessment of Educational Progress (NAEP; Beaton, 1987; Zwick, 1992), the Third International Mathematics and Science Study (TIMSS) (Martin and Kelly, 1996) and the OECD Programme for International Student Assessment (PISA) (Adams and Wu, 2002) are good examples of this. In TIMSS, single test booklets of 70 minutes duration are used to provide information on as many as 12 latent ability dimensions. A set of 12 unidimensional analyses of these data will not provide reliable student ability estimates for subsequent use in analysis or reporting. In PISA test booklets of 120 minutes were used to provide information on five dimensions.

Multidimensional item response models have been presented and investigated by many authors (for example; Ackerman, 1992; Andersen, 1985; Batley and Boss, 1993; Camilli, 1992; Embretson, 1991; Glas, 1992; Kelderman and Rijkes, 1994; Luecht and Miller, 1992; Reckase, 1985; Reckase & McKinley, 1991), there application has, however, been limited to a few isolated examples. Even in NAEP a two step estimation procedure is used to avoid the use of a fully multidimensional item response model (see Beaton, 1987).

In this paper we address two concerns with regard the application of one multidimensional model that has been proposed—the multidimensional random coefficients multinomial logit that was introduced and discussed by Adams, Wilson

and Wang (1997), and Wang (1994). The conditions under which the model is identified and the application of the model in high dimensions is considered.

In the next section of the paper the model is described. In section three identification restrictions are considered, in section four a Monte Carlo EM method is described for fitting the model in high dimensions, in section five some simulations that explore the properties of the estimation method are reported and in section six the application of the model to some real data is shown.

2. FORMAL DEFINITION OF THE MODEL.

Suppose there is a set of items, **I**, indexed $i = 1, \ldots, I$ and each item i has $K_i + 1$ response categories with index $j = 0, \ldots, K_i$. Further, there are also N individuals, indexed $n = 1, \ldots, N$ and each individual responds to some subset $\mathbf{I}_n$ of items from **I**. The vector valued random variable $\mathbf{X}_i$ is applied to indicate the $K_i + 1$ responses to item i. That is $\mathbf{X}_{ni} = \left(X_{ni1}, X_{ni2}, \ldots, X_{niK_i}\right)^T$, where

$$X_{nij} = \begin{cases} 1 & \text{if person } n \text{ scores in category } j \text{ on item } i \\ 0 & \text{otherwise} \end{cases}$$

for $j = 1, \ldots, K_i$. Note that it follows that if the individual responds in the category $j = 0$, or if item i was not in the set $\mathbf{I}_n$ then $\mathbf{X}_{ni} = \mathbf{0}$.

The esponse pattern is the vector valued random variable $\mathbf{X}_n = (\mathbf{X}_{n1}^T, \mathbf{X}_{n2}^T, \ldots, \mathbf{X}_{nI}^T)^T$ which was created by collecting the $\mathbf{X}_{ni}$ together into a single vector. Particular instances of each of the random variables are indicated by using their lower case equivalents. That is $\mathbf{x}_n, \mathbf{x}_{ni}$ and $\mathbf{x}_{nik}$.

Individuals are modelled through a *D*-dimensional latent attribute parameter $\boldsymbol{\theta} = \left(\theta_1, \theta_2, \ldots, \theta_D\right)^T$, which is seen as random with a population distribution given by the multivariate normal probability density function (pdf):

$$g(\boldsymbol{\theta}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left\{ -\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}) \right\}, \qquad (1)$$

where $\boldsymbol{\mu}$ is the vector of means and $\boldsymbol{\Sigma}$ is the matrix of covariance of the random vector $\boldsymbol{\theta}$.

An additional feature of the model is the introduction of a scoring function that allows the description of the score or *performance level* that is assigned to each

response category on each of the *D*-dimensions. Using this approach it is possible to do if to introduce the notion of a response score, which gives the score level for dimension d of the observed response in category k of item i. If a response category for a particular item does not relate to a particular latent dimension then the score on that latent dimension is set to zero. The b_{ikd} can be collected in a vector as $\mathbf{b}_{ik} = (b_{ik1}, b_{ik2}, \ldots, b_{ikD})^T$, and the vectors can be collected into the matrix $\mathbf{B} = (\mathbf{b}_{11}^T, \mathbf{b}_{12}^T, \ldots, \mathbf{b}_{1K_1}^T, \mathbf{b}_{21}^T, \ldots, \mathbf{b}_{2K_2}^T, \ldots, \mathbf{b}_{IK_I}^T)^T$.

The item parameters are given by the vector $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_P)^T$. Linear combinations of these P parameters are used in the response probability model to describe the behaviour of the response categories to each item. These linear combinations are defined by design vectors $\mathbf{a}_{ik}^T \quad (\text{for } i = 1, \ldots, I; \text{and } k = 1, \ldots, K_i)$, so that if to define $K = \sum_{i=1}^{I} K_i$, they can be denoted collectively by the $K \times P$ design matrix $\mathbf{A} = \left(\mathbf{a}_{11}^T, \mathbf{a}_{12}^T, \ldots, \mathbf{a}_{1K_1}^T, \mathbf{a}_{21}^T, \ldots, \mathbf{a}_{2K_2}^T, \ldots, \mathbf{a}_{IK_I}^T\right)^T$.

Through the introduction of the scoring matrix **B** and the design matrix **A** it is possible to write a general mixed multinomial logit regression model that includes as special cases a wide class of existing Rasch models and perhaps more importantly provides a general context in which to develop and test Rasch measurement models (references)

The item response probability model is:

$$f(\mathbf{x}_{nik}; \mathbf{A}, \mathbf{B}, \boldsymbol{\xi} | \boldsymbol{\theta}) = \frac{\exp \mathbf{x}_{nik}\left(\mathbf{b}_{ik}^T \boldsymbol{\theta} + \mathbf{a}_{ik}^T \boldsymbol{\xi}\right)}{\sum_{u=1}^{K_i} \exp \mathbf{x}_{niu}\left(\mathbf{b}_{iu}^T \boldsymbol{\theta} + \mathbf{a}_{iu}^T \boldsymbol{\xi}\right)}. \tag{2}$$

The assumption of conditional independence allows us to write the probability of a vector of responses conditioned on the random quantities for n-th person as

$$\boldsymbol{\varphi}(\mathbf{x}_n; \mathbf{A}, \mathbf{B}, \boldsymbol{\xi} | \boldsymbol{\theta}) = \frac{\exp \mathbf{x}_n^T\left(\mathbf{B}\boldsymbol{\theta} + \mathbf{A}\boldsymbol{\xi}\right)}{\boldsymbol{\Psi}_n(\boldsymbol{\theta}, \boldsymbol{\xi})}, \tag{3}$$

where

$$\boldsymbol{\Psi}_n(\boldsymbol{\theta}, \boldsymbol{\xi}) = \prod_{i \in \mathbf{I}_n} \sum_{j=1}^{K_i} \exp \mathbf{b}_{ij}^T \boldsymbol{\theta} + \mathbf{a}_{ij}^T \boldsymbol{\xi}\ , \tag{4}$$

or equivalently

$$\Psi_n(\theta,\xi) = \sum_{\mathbf{z}\in\Gamma_n} \exp\ \mathbf{z}^T\left(\mathbf{B}\theta + \mathbf{A}\xi\right)\ , \tag{5}$$

and Γ_n is the set of all possible response vectors for the items in $\mathbf{I}$ with zeros for all categories for all items that were not in $\mathbf{I}_n$, the sub-set of items responded to by person n.

Under the marginal formulation the probability of a response vector $\mathbf{x}_n$ is:

$$f\left(\mathbf{x}_n;\mathbf{A},\mathbf{B},\xi,\mu,\Sigma\right) = \int_{R^D} \varphi\left(\mathbf{x}_n;\mathbf{A},\mathbf{B},\xi|\theta\right) g(\theta;\mu,\Sigma)d\theta \tag{6}$$

or equivalently

$$f\left(\mathbf{x}_n;\mathbf{A},\mathbf{B},\xi,\mu,\Sigma\right) = \int_{R^D} \frac{\exp\ \mathbf{x}_n^T\left(\mathbf{B}\theta + \mathbf{B}\mu + \mathbf{A}\xi\right)}{\Psi_n\left(\theta+\mu,\xi\right)} g(\theta;0,\Sigma)d\theta$$

$$= \int_{R^D} \frac{\exp\ \mathbf{x}_n^T\left(\mathbf{B}\theta + \ \mathbf{B}\quad\mathbf{A}\quad\mu^T\quad\xi^{T\ T}\right)}{\Psi_n\left(\theta+\mu,\xi\right)} g(\theta;0,\Sigma)d\theta. \tag{7}$$

Through appropriate choices of $\mathbf{A}$ and $\mathbf{B}$, (7) can be shown to encompass models such as Rasch's (1960) simple logistic model, Fischer's (1973) linear logistic latent trait model, Andrich's (1978) rating scale model, Masters' (1982) partial credit model, Linacre's many-faceted model (1989), Wilson's (1992) ordered partition model and a range of multidimensional models such as Whitely's (1980) multicomponent latent trait model, Andersen's (1985) Rasch model for repeated testing and Embretson's (1991) multidimensional Rasch model for learning and change. The reader is referred to Adams and Wilson (1997) and Adams, Wilson and Wang (1997) for examples of how this is accomplished.

As an illustration the form of the $\mathbf{A}$ and $\mathbf{B}$ matrices necessary to give a unidimensional version of the simple logistic model, the partial credit model and the rating scale model are provided.

For the simple logistic model as applied to dichotomies, $\mathbf{A}$ is simply a diagonal matrix with ones on the diagonal and $\mathbf{B}$ is a vector of ones. So that if a test has four items the matrices $\mathbf{A}$ and $\mathbf{B}$ have the following form

$$\mathbf{A}_{SLM} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{B}_{SLM} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}. \tag{8}$$

For the partial credit model and using the Masters' parameterisation the **A** matrix is made up of a sequence of blocks, one for each item. The size of each block is equal to K_i and the elements of the matrix are zeros if they are over the main diagonal and one if they are on or below this diagonal. **B** is a vector made up of sequences of successive integers, one sequence for each item. For example, if there are three items with 3, 4 and 5 categories respectively then the form of complete matrices **A** and **B** are:

$$\mathbf{A}_{PCM} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{B}_{PCM} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \\ 3 \\ 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \tag{9}$$

For Andrich's (1978) rating scale model, again using the Wright & Masters (1982) parameterisation a test of I items each with $K+1$ categories would be modelled with the following **A** and **B**.

$$\mathbf{A}_{RSM} = \left[\begin{array}{cccccccc} 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 2 & 0 & \cdots & 0 & 1 & 1 & \cdots & 0 \\ \cdots & 0 & \cdots & 0 & \cdots & \cdots & \cdots & \cdots \\ K-1 & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 \\ K & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 2 & \cdots & 0 & 1 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & 0 & \cdots & \cdots & \cdots & \cdots \\ 0 & K-1 & \cdots & 0 & 1 & 1 & \cdots & 1 \\ 0 & K & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 2 & 1 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & K-1 & 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & K & 0 & 0 & \cdots & 0 \end{array}\right] \quad \text{and} \quad \mathbf{B}_{RSM} = \left[\begin{array}{c} 1 \\ 2 \\ \cdots \\ K-1 \\ K \\ 1 \\ 2 \\ \cdots \\ K-1 \\ K \\ \cdots \\ 1 \\ 2 \\ \cdots \\ K-1 \\ K \end{array}\right] \tag{10}$$

3. IDENTIFICATION OF THE MODEL.

While the matrices given in (8), (9) and (10) provide the simple logistic, rating scale and partial credit models, as they are usually described, they cannot be used directly in (7) because they will result in an unidentified model. This is easily recognised for the simple logistic model by noting that if $\boldsymbol{\mu}^* = \boldsymbol{\mu} + c$ and $\boldsymbol{\xi}^* = \boldsymbol{\xi} - c\mathbf{1}_P$ where c is a constant and $\mathbf{1}_P$ is a P vector of ones then

$$f(\mathbf{x}_n; \mathbf{A}_{SL}, \mathbf{B}_{SL}, \boldsymbol{\xi}^*, \boldsymbol{\mu}^*, \Sigma) = f(\mathbf{x}_n; \mathbf{A}_{SL}, \mathbf{B}_{SL}, \boldsymbol{\xi}, \boldsymbol{\mu}, \Sigma)$$

It can be similarly shown that, $\mathbf{A}_{PCM}$ and $\mathbf{B}_{PCM}$, and $\mathbf{A}_{RSM}$ and $\mathbf{B}_{RSM}$, result in unidentified models.

More generally, from (7), it immediately follows that the model is not identified for arbitrary choices of $\mathbf{A}$ and $\mathbf{B}$. In addressing the problem of identification the goal is to determine the conditions that must be satisfied by $\mathbf{A}$ and $\mathbf{B}$ to ensure that if $\mathbf{x}^T(\mathbf{B}\boldsymbol{\mu} + \mathbf{A}\boldsymbol{\xi}) = \mathbf{x}^T(\mathbf{B}\boldsymbol{\mu}^* + \mathbf{A}\boldsymbol{\xi}^*)$ for every possible response vector $\mathbf{x}$, then $\boldsymbol{\mu} \equiv \boldsymbol{\mu}^*$ and $\boldsymbol{\xi} \equiv \boldsymbol{\xi}^*$.

In past practice this problem has normally been solved for the simple logistic model by applying the constraint $\boldsymbol{\mu} \equiv 0$, or by considering one of the item parameters as a linear combination of the others. Wright and Masters (1982), for example, constrain

one parameter to be the negative sum of all other parameters. However, in the case of this more general model the identification becomes a more difficult issue.

The following three propositions give us a set of conditions with regard the identification of the model.

Proposition One: If D is the number of latent dimensions, P is the length of the parameter vector, ξ, and K_i+1 is the number of response categories for item i and $K=\sum_{i\in\mathbf{I}} K_i$, then model (7) if applied to the set of items **I** can only be identified if $P+D\leq K$.

Proof: The proof of this proposition follows directly from the fact that the rank of $\mathbf{B}\ \ \mathbf{A}$ must be less than or equal to K, since it contains at most K non-zero rows. If the length of the vector $\mu^T\ \ \xi^T\ ^T$ (which equals $P+D$) exceeds K then $\mathbf{B}\ \ \mathbf{A}$ will not be of full column rank and single unique solution for $\mu^T\ \ \xi^T\ ^T$ will not exist.

Proposition Two: If D is the number of latent dimensions, P is the length of the parameter vector, ξ, then model (7) can only be identified if $rank(\mathbf{A})=P$, $rank(\mathbf{B})=D$ and $rank(\mathbf{B}\ \ \mathbf{A})=P+D$.

Proof: Since **A** must be conformable with ξ it must have $rank(\mathbf{A})\leq P$ Suppose that $rank(\mathbf{A})<P$ then a fixed value of the product $\mathbf{A}\xi$ does not provide a unique solution for ξ and the model is not identified. An identical argument can be provided for **B** and for $\mathbf{B}\ \ \mathbf{A}$.

Proposition Three: If D is the number of latent dimensions, P is the length of the parameter vector, ξ, and K_i+1 is the number of response categories for item i and $K=\sum_{i\in\mathbf{I}} K_i$, then model (7) if applied to the set of items **I** can only be identified if and only if $rank(\mathbf{B}\ \ \mathbf{A})=P+D\leq K$.

Proof: The necessary conditions of this proposition follow directly from propositions one and two. For proving sufficiency let $\mathbf{C}=\mathbf{B},\mathbf{A}$, $\eta=\mu^T\ \ \xi^T\ ^T$ and Γ be the set of all possible response vectors for the items in **I** so that the identification requirement can be written as,

$$\mathbf{x}^T\mathbf{C}\left(\boldsymbol{\eta}-\boldsymbol{\eta}^*\right)=\mathbf{0} \quad \forall\mathbf{x}\in\boldsymbol{\Gamma} \;\; \textit{iff} \;\; \boldsymbol{\eta}=\boldsymbol{\eta}^*.$$

Since $\mathbf{C}$ is a $K\times(P+D)$ matrix with rank $P+D$ therefore $K-(P+D)$ rows could be removed from $\mathbf{C}$ to produce a square sub-matrix $\widetilde{\mathbf{C}}$ with rank $P+D$. Now let $\mathbf{x}_k$ be the response vector that has a one in the *k*-th position and zeros elsewhere and let $\widetilde{\mathbf{x}}_k$ be the corresponding vector with the same rows removed as where removed from $\mathbf{C}$ to construct $\widetilde{\mathbf{C}}$. Finally let $\boldsymbol{\Omega}$ be the set of $P+D$ such vectors that are not all zeros.

Now, $\mathbf{x}_k^T\mathbf{C}\left(\boldsymbol{\eta}-\boldsymbol{\eta}^*\right)=\widetilde{\mathbf{x}}_k^T\widetilde{\mathbf{C}}\left(\boldsymbol{\eta}-\boldsymbol{\eta}^*\right)=0$ for $\widetilde{\mathbf{x}}_k\in\boldsymbol{\Omega}$ is equivalent to $\widetilde{\mathbf{C}}\left(\boldsymbol{\eta}-\boldsymbol{\eta}^*\right)=\mathbf{0}$ which can hold if and only if $\boldsymbol{\eta}=\boldsymbol{\eta}^*$.

Note, that it follows directly from these propositions that the models given in (8), (9) and (10) are not identified. For each of these models identification can be conveniently achieved by imposing the constraint $\boldsymbol{\mu}\equiv\mathbf{0}$. There are however, practical circumstances where it may be appropriate to modify the $\mathbf{A}$ matrix so that identification can be provided. For example, a constraint of this type may not be suitable when the population model (1), is extended to include collateral variables (see Mislevy, 1985; Adams, Wilson and Wu, 1997). If the original design matrix is called the complete matrix and denoted $\mathbf{A}_c$ then in the next section a procedure for producing a "reduced matrix" $\mathbf{A}$ that ensures the identification of (7) yet maintains the basic structure and intention of a specified $\mathbf{A}$ matrix is given. In describing the procedure the cases of a multidimensional dichotomous model, a multidimensional partial credit model and a multidimensional rating scale model are considered separately.

The Dichotomous Case

The construction of a reduced matrix $\mathbf{A}$ can proceed by defining *D* subsets of items $\mathbf{J}_k=\left\{i_{1k},i_{2k},\ldots,i_{n_kk}\right\}$ for $k=1,\ldots,D$ each of size n_k. It is not necessary that $\mathbf{J}_k$ and $\mathbf{J}_l$ have an empty intersection when $k\neq l$. For every $\mathbf{J}_k$ a vector $\mathbf{e}_k=\sum_{i\in\mathbf{J}_k}\mathbf{b}_{i1}$ can be built and the matrix $\mathbf{E}=\left(\mathbf{e}_1^T,\mathbf{e}_2^T,\ldots,\mathbf{e}_D^T\right)^T$ is considered.

Proposition Four Suppose that in every $\mathbf{J}_k$, there is $n_k > 1$, $k = 1,\ldots,D$ and it is possible to construct a set F which contains D different items with $\mathbf{J}_k \cap F \neq \varnothing$ for any *k*. Further, assume that none of the $\mathbf{J}_k$ is a subset of any other $\mathbf{J}_l$, $l \neq k$ and none of the $\mathbf{J}_k$ is a subset of F, $k = 1,\ldots,D$. Then, if the determinant of $\mathbf{E}$ is not zero, the multidimensional dichotomous model can be fully specified by reducing complete matrix $\mathbf{A}_c$ by D columns so that the rank of reduced matrix $\mathbf{A}$ will be $P - D$.

Proof. Without loss of generality $i_{n_k k}$ can be assumed to be an element of $\mathbf{J}_k$ ($k = 1,\ldots,D$), which belongs to F (if there is more than one such element one which has not been previously used should be chosen). Now suppose that the $i_{n_k k}$-th item parameter is the negative sum of the other item parameters from $\mathbf{J}_k$, then in the $i_{n_k k}$-th row of matrix $\mathbf{A}_c$ in columns $i_{jk}, j = 1,\ldots,n_k - 1$ place -1 and in column n_k place 0. Repeat this procedure D times so there will be D columns in the matrix that contain only zeros. Deleting these zeros columns from $\mathbf{A}_c$ results in the reduced matrix $\mathbf{A}$.

Suppose that there exists non-zero vector $\mathbf{c}$, such that $\mathbf{x}^T(\mathbf{B}\boldsymbol{\mu} + \mathbf{A}\boldsymbol{\xi}) = \mathbf{x}^T(\mathbf{B}\boldsymbol{\mu}^* + \mathbf{A}\boldsymbol{\xi}^*)$, for all $\mathbf{x}$. Now choose a vector $\mathbf{x}$ with 1 in the positions $i_{jk}, j = 1,\ldots,n_k$ and 0 elsewhere. It follows that $\mathbf{x}^T\mathbf{A}\boldsymbol{\xi} = \mathbf{x}^T\mathbf{A}\boldsymbol{\xi}^* = \mathbf{0}$ and $\mathbf{x}^T\mathbf{B}(\boldsymbol{\mu}^* - \boldsymbol{\mu}) = \mathbf{x}^T\mathbf{B}\mathbf{c} = \mathbf{e}_k^T\mathbf{c} = \mathbf{0}$ for $k = 1,\ldots,D$ and if the determinant of $\mathbf{E}$ is non-zero, then vector $\mathbf{c}$ must be equal to zero. It follows that $\boldsymbol{\mu} \equiv \boldsymbol{\mu}^*$ and $\boldsymbol{\xi} \equiv \boldsymbol{\xi}^*$.

The following example illustrates proposition four.

Example 1. Consider four dichotomous items with $D = 3$ and matrices $\mathbf{A}$ and $\mathbf{B}$ as follows

$$\mathbf{A}_c = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Now, choose $\mathbf{J}_1 = \{1,2\}$, $\mathbf{J}_2 = \{2,3\}$, and $\mathbf{J}_3 = \{2,4\}$ then

$$\mathbf{E} = \begin{bmatrix} 2 & 1 & 1 \\ 2 & 2 & 2 \\ 1 & 1 & 2 \end{bmatrix},$$

and it is obvious that all the conditions of Proposition four hold. Following the procedure given in the proposition the reduced matrix could be constructed

$$\mathbf{A} = \begin{bmatrix} -1 \\ 1 \\ -1 \\ -1 \end{bmatrix}.$$

At this point it is sufficient to apply proposition three by noting that

$$rank\begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix} = rank\begin{bmatrix} -1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 \\ -1 & 0 & 0 & 1 \end{bmatrix} = 4 = P + D.$$

The Partial Credit Case

Proposition four can be extended to the partial credit case through a modification of the procedure for constructing the matrix $\mathbf{E}$.

Let $\mathbf{J}_k$, $k = 1, \ldots, D$ be D subsets of all of the item response categories and the categories are sequentially labelled so that each $\mathbf{J}_k$ contains the labels of n_k categories. That is $\mathbf{J}_k = \{i_{1k}, i_{2k}, \ldots, i_{n_k k}\}$. Again, there is no requirement that $\mathbf{J}_k$ and $\mathbf{J}_l$ have an empty intersection when $k \neq l$.

Now, for every J_k a vector $\mathbf{e}_k = \sum_{m=1}^{n_k} \mathbf{b}_{j_{mk} i_{mk}}$ can be built and the matrix $\mathbf{E} = \left(\mathbf{e}_1^T, \mathbf{e}_2^T, \ldots, \mathbf{e}_D^T\right)^T$ is considered.

Proposition Five. Suppose that in every $\mathbf{J}_k$, there is $n_k > 1$, $k = 1, \ldots, D$ and it is possible to construct a set F which contains D different categories with $\mathbf{J}_k \cap F \neq \varnothing$ for any *k*. Assume that none of the $\mathbf{J}_k$ is a subset of any other $\mathbf{J}_l$, $l \neq k$ and none of the $\mathbf{J}_k$ is a subset of F, $k = 1, \ldots, D$. Then if the determinant of $\mathbf{E}$ is not zero, the multidimensional partial credit model can be fully

specified by reducing complete matrix $\mathbf{A}_c$ by D columns so that the rank of the reduced matrix $\mathbf{A}$ will be $P-D$.

The proof of this proposition follows the proof of proposition four.

To illustrate consider three polytomous items with $D=4$ and matrices $\mathbf{A}_c$ and $\mathbf{B}$ as follows

$$\mathbf{A}_c = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ 4 & 2 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 2 & 2 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 4 \end{bmatrix}.$$

Now, choose $\mathbf{J}_1 = \{1, 2, 3\}$, $\mathbf{J}_2 = \{3, 4, 5\}$, $\mathbf{J}_3 = \{5, 6, 7\}$, $\mathbf{J}_4 = \{7, 8, 9, 10, 11\}$ and $F = \{1, 4, 6, 11\}$ then

$$\mathbf{E} = \begin{bmatrix} 6 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 \\ 0 & 1 & 6 & 1 \\ 0 & 0 & 3 & 11 \end{bmatrix},$$

it is easy to see that the conditions of proposition five hold. The reduced matrix is constructed by following the procedure given in proposition five.

$$\mathbf{A} = \begin{bmatrix} -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

If the item parameter vector for the complete model was $\boldsymbol{\xi}^T = (\xi_1, \xi_2, \ldots, \xi_{11})$ then the use of the reduced $\mathbf{A}$ given above is equivalent to imposing the constraints

$$\left.\begin{aligned} \xi_1 &= -\xi_2 - \xi_3 - \xi_4 \\ \xi_4 &= -\xi_3 - \xi_5 - \xi_6 \\ \xi_7 &= -\xi_5 - \xi_4 \\ \xi_{11} &= -\xi_8 - \xi_9 - \xi_{10} \end{aligned}\right\}$$

At this point it is sufficient to note that $rank\ \mathbf{A} \quad \mathbf{B} = 11 = P + D$.

The Rating Scale Case

If following the Andrich formulation of the rating scale model the complete $\mathbf{A}_c$ matrix takes the form given in (10), the parameter vector, $\boldsymbol{\xi}$, consists of a set of **I**, item difficulty parameters and $K-1$ threshold parameters. A key feature of this model, is that the threshold parameters describe a response structure that is constant across items, it will therefore be a requirement of the construction of the reduced $\mathbf{A}$ that this property of the threshold parameters be maintained.

If $\mathcal{K} = (1, 2, \ldots, K)^T$ then the $\mathbf{B}$ matrix for a rating scale model is defined as $\mathbf{B} = \mathbf{B}^* \otimes \mathcal{K}$ where $\mathbf{B}^*$ is an $I \times D$ matrix of zeros and ones which indicates the assignment of items to dimensions and $\otimes$ - denotes the Kronecker product. This matrix corresponds to the scoring matrix for a D–dimensional dichotomous model. Similarly the complete $\mathbf{A}_c$ matrix could be presented as $\mathbf{A}_c = \mathbf{A}_c^* \otimes \mathcal{K}$ where $\mathbf{A}_c^*$ is an identity matrix of order I and it is precisely the complete design matrix for a D-dimensional dichotomous model. So, the operator $\otimes$ provides a bijective mapping between the D-dimensional dichotomous model and the D-dimensional rating scale

model. It follows that to reducing $\mathbf{A}_c^*$ to $\mathbf{A}^*$ can follow the procedure described for the dichotomous model.

4. ESTIMATION.

A maximum likelihood technique for estimating the parameters, ξ and μ and Σ is fully described in Adams, Wilson and Wang (1997). They show that the likelihood of a response matrix $\mathbf{X}$ for N persons drawn at random from normal population is

$$\Lambda(\xi,\mu,\Sigma|\mathbf{X}) = \prod_{n=1}^{N} f(\mathbf{x}_n;\mathbf{A},\mathbf{B},\xi,\mu,\Sigma) \tag{11}$$

and that following directly the procedures developed by Bock and Aitken (1981) this likelihood can be maximised with an EM algorithm that iteratively solves,

$$\mathbf{A}^T \sum_{n=1}^{N} \left\{ \mathbf{x}_n - \int_{\theta} E_{\mathbf{z}}(\mathbf{z}|\theta) dH(\theta;\xi,\alpha|\mathbf{x}_n) \right\} = \mathbf{0}, \tag{12}$$

$$\mu = \frac{1}{N} \sum_{n=1}^{N} \int_{\theta} \theta dH(\theta;\xi,\mu,\Sigma|\mathbf{x}_n), \tag{13}$$

and

$$\Sigma = \frac{1}{N} \sum_{n=1}^{N} \int_{\theta} (\theta-\mu)(\theta-\mu)^T dH(\theta;\xi,\mu,\Sigma|\mathbf{x}_n) \,. \tag{14}$$

Where

$$E_{\mathbf{z}}(\mathbf{z}|\theta) \;=\; \Psi(\theta,\xi) \sum_{\mathbf{z}\in\Omega} \mathbf{z} \exp\left\{ \mathbf{z}^T (\mathbf{B}\theta + \mathbf{A}\xi) \right\},$$

is the expected response pattern and $H(\theta;\xi,\alpha|\mathbf{x}_n)$ is the distribution function of the marginal density of θ given $\mathbf{x}_n$. The density of which is given by

$$h(\theta;\xi,\alpha|\mathbf{x}_n) = \frac{f(\mathbf{x}_n;\xi|\theta) g(\theta;\alpha)}{f(\mathbf{x}_n;\xi)} \tag{15}$$

Under the Adams *et al.* (1997) approach the integration required to solve the system (12), (13), (14) is computed using a straightforward application of quadrature over a fixed uniform grid of nodes that is specified a-priori. They have found that while

this is a practicable approach in low dimension the computing time becomes excessive for models of four or more dimensions. As an alternative Glas (1992) proposes the use of Gauss-Hermite quadrature (see, Ralston & Rabinowitz, 1978) in conjunction with a transformation of the multivariate normal density to ensure that the latent dimensions are orthogonal.

In the case of a single latent dimension and a normal population density it is clear that using Gauss-Hermite quadrature for the integration is desirable since under these circumstances the integrals take the form

$$\begin{aligned} f(\mathbf{x}_n;\mathbf{A},\mathbf{B},\xi,\mu,\sigma) &= \int \varphi(\mathbf{x}_n;\mathbf{A},\mathbf{B},\xi\,|\,\theta)\, g(\theta;\mu,\sigma)\, d\theta \\ &= \int \varphi(\mathbf{x}_n;\mathbf{A},\mathbf{B},\sigma\cdot\xi+\mu\,|\,\theta)\, g(\theta;0,1)\, d\theta, \end{aligned} \tag{16}$$

where $g(\boldsymbol{\theta};\boldsymbol{\mu},\boldsymbol{\sigma})$ is the pdf of the normal distribution.

Our exploration of this method suggests that an approximation that uses eight nodes will generally be adequate. Glas' suggestion of using the same procedure for the multidimensional case appears to work well for two dimensions but is not as efficient in higher dimensions because, as in the Adams *et al.* approach, the number of nodes increases exponentially with dimensionality. This method also suffers from a theoretical disadvantages — for a fixed number of nodes the remainder term in the approximation tends to zero slowly with increasing of number of dimensions.

An alternative approach is to use a Monte Carlo method where the integral (16) is considered as the mathematical expectation of a function of a normally distributed random variable (see Kalos and Whitlock, 1986; Tanner, 1993); that is

$$f(\mathbf{x}_n;\mathbf{A},\mathbf{B},\boldsymbol{\xi},\boldsymbol{\mu},\boldsymbol{\sigma}) = E\boldsymbol{\varphi}(\mathbf{x}_n;\mathbf{A},\mathbf{B},\boldsymbol{\sigma}\cdot\boldsymbol{\xi}+\boldsymbol{\mu}|\boldsymbol{\eta}). \tag{17}$$

For the multidimensional case the decomposition of the covariance matrix $\boldsymbol{\Sigma}$ in the form $\boldsymbol{\Sigma}=\mathbf{V}\mathbf{D}\mathbf{V}^T$, where $\mathbf{V}$ is an orthogonal matrix and D is a diagonal matrix of eigenvalues of the matrix $\boldsymbol{\Sigma}$ and $\mathbf{V}^{-1}=\mathbf{V}^T$, so that

$$\boldsymbol{\Sigma}^{-1}=\mathbf{V}\mathbf{D}^{-1}\mathbf{V}^T=\left(\mathbf{V}\mathbf{D}^{-1/2}\right)\left(\mathbf{D}^{-1/2}\mathbf{V}^T\right),$$

is found. Then,

$$\int_{\Re^D} \varphi(\mathbf{x}_n;\mathbf{A},\mathbf{B},\xi\,|\,\theta)g(\theta;\mu,\Sigma)d\theta = \int_{\Re^D} \varphi\left(\mathbf{x}_n;\mathbf{A},\mathbf{B},(\mathbf{V}\mathbf{D}^{1/2})^T\xi+\mu\,|\,\theta\right)g\left(\theta;\mathbf{0},\mathbf{I}\right)d\theta$$
$$= E\varphi\left(\mathbf{x}_n;\mathbf{A},\mathbf{B},(\mathbf{V}\mathbf{D}^{1/2})^T\xi+\mu\,|\,\eta_D\right), \tag{18}$$

where η_D is distributed as a *D*-dimensional standard normal distribution. For estimation of (18) a sample of normally distributed random vectors $\eta_{Di},\ i=1,\ldots,M$ is simulated and the expectation is approximated by

$$\frac{1}{M}\sum_{i=1}^{M}\varphi\left(\mathbf{x}_n;\mathbf{A},\mathbf{B},(\mathbf{V}\mathbf{D}^{1/2})^T\xi+\mu\,|\,\eta_{Di}\right) \tag{19}$$

It is well known (see, for example Kalos and Whitlock, 1986) that the rate of convergence of (19) in this case has the order $M^{-1/2}$ and this rate does not depend on number of dimensions.

5. STATISTICAL SIMULATIONS.

With a model as complex as the one being considered here an exhaustive array of simulations that explore and illustrate the properties of the model is obviously not feasible. In the following the results of six sets of simulations that were each designed to explore one set of issues in relation to the fitting of the model are reported.

In set one, the influence that the number of nodes has on the recovery of the item parameters and the covariance matrix when a six dimensional partial credit model is fit with samples of 1000 is considered. In set two, these analyses are repeated using samples of 2000. In set three the influence of using a convergence criteria based on the change in the likelihood, rather than the change in the parameter estimates, which is used for all other simulations is examined. In set four the reduction of the number of items on each dimension to just four dichotomous items is studied. In set five the number of dimensions was reduced to four dimensions and the covariance matrix was changed to one that gives a wider range of correlations between the level variables — including negative values. Finally, in set six, a four-dimensional model using multidimensional Gauss-Hermite quadrature is fitted.

All of the simulations that were undertaken involved the following procedure

a) Using the Monte Carlo method, 200 samples were drawn from a hypothetical population in which the latent ability dimensions followed a muiltivariate normal

distribution with a specified covariance matrix and zero means. In these simulations only two different covariance matrices were used. For sets one to four the following six dimensional covariance matrix was used.

$$\Sigma = \begin{bmatrix} 1.0 & 0.7 & 0.7 & 0.7 & 0.7 & 0.7 \\ 0.7 & 1.0 & 0.7 & 0.7 & 0.7 & 0.7 \\ 0.7 & 0.7 & 1.0 & 0.7 & 0.7 & 0.7 \\ 0.7 & 0.7 & 0.7 & 1.0 & 0.7 & 0.7 \\ 0.7 & 0.7 & 0.7 & 0.7 & 1.0 & 0.7 \\ 0.7 & 0.7 & 0.7 & 0.7 & 0.7 & 1.0 \end{bmatrix}. \qquad (20)$$

For sets five and six the four-dimensional covariance matrix

$$\Sigma = \begin{bmatrix} 1.000 & 0.678 & -0.348 & -0.430 \\ 0.678 & 1.234 & 0.990 & -0.780 \\ -0.348 & 0.990 & 2.100 & -0.500 \\ -0.430 & -0.780 & -0.500 & 1.978 \end{bmatrix}, \qquad (21)$$

which yields the correlation matrix

$$corr = \begin{bmatrix} 1.000 & 0.610 & -0.240 & -0.306 \\ 0.610 & 1.000 & 0.615 & -0.499 \\ -0.240 & 0.615 & 1.000 & -0.245 \\ -0.306 & -0.499 & -0.245 & 1.000 \end{bmatrix},$$

was used.

When the data is generated with the matrix (20) it will be referred to as the six-dimensional population of model. If it was drawn with (21) it will be referred to as the four-dimensional population model.

b) Item parameters were randomly selected from fixed intervals which are different for each dimension.

c) For each student in the simulated samples response vectors that conformed to the multidimensional Rasch model were generated using the Monte Carlo method. For each individual the response vector was generated assuming their generated ability vector and the fixed assumed, item parameters.

SET ONE: The Influence of Number of Nodes.

The first set of simulations used the six dimensional population model with four partial credit items in each dimension. One item with each of 2, 3, 4 and 5 response categories. Standard integer scoring was used so that the possible scores on each dimension ran from 0 to 10. Item parameters, ξ, and population parameters, Σ, were estimated using the Monte Carlo Method with 500, 1000, 1500 and 2000 nodes for each of 200 generated data sets.

Table 1 summarises the effectiveness of the parameter recovery for the four nodes numbers.

Table 1. Summary Results for Simulations Set One.

No. of nodes	Largest absolute value of *t*-statistics (*t*) and corresponding quantile (*p*) for these statistics				Hotelling's T^2-statistics	Average number of iterations
	t for ξ	*p* for ξ	*t* for Σ	*p* for Σ		
500	1.056	0.854	1.358	0.913	39.30	95
1000	0.626	0.734	1.019	0.846	11.61	111
1500	0.445	0.672	0.776	0.781	6.89	109
2000	0.306	0.620	0.411	0.660	2.51	114

The values in Table 1 were computed as follows. For each item parameter i we have the generating value ξ_i and 200 estimated values $\hat{\xi}_{i1},\ldots,\hat{\xi}_{i200}$, one for each simulated sample. It follows from design that number of parameters vary from one to four for different items. If $\hat{m}_i$ and $\hat{s}_i$ are the mean and standard deviation of the estimated values then the statistics $t_i = (\hat{m}_i - \xi_i)/\hat{s}_i$ have a *t*-distribution with 199 degrees of freedom (which can be approximated by the standard normal distribution). The maximum absolute value among all t_i is reported in the second column in Table 1. In the third column the probability from the standard normal distribution which corresponds to the quantile $\max_i |t_i|$ is reported. The same procedure is applied to the variances and covariances to provide columns four and five. In column six of Table 1 the statistics $T^2(N-p)/((N-1)p)$ are reported, where T^2 is the Hotelling's statistic which was calculated to test the hypothesis that the means of all of the estimated parameters are equal to the generating values. The statistics $T^2(N-p)/((N-1)p)$ have an F distribution with $N-p$ and p degrees of freedom. In the last column of the table the average number of iterations which were needed to achieve convergence is reported.

Table 1 shows that increasing the number of nodes improves the estimation. If the hypotheses that the estimated parameters are the same as the generating values are tested for each component separately, it follows from Table 1 that for 500 nodes the hypotheses are not rejected with $\alpha = 0.15$, for 1000 nodes with $\alpha = 0.30$, for 1500 nodes with $\alpha = 0.40$ and for 2000 nodes with $\alpha = 0.65$. The results are even better if the covariance parameters are excluded.

The values of T^2 decline with each increase in the number of nodes but even at 2000 nodes Hotelling's test rejects the null hypothesis that the estimated values are unbiased. A possible explanation for this fact is the fact that the vectors are only asymptotically normal while Hotelling's test assumes normality. Dependence of the rate of convergence in the multidimensional central limit theorem on the number of dimensions could well also be a factor that affects the validity of the application of the Hotelling's test.

The adequacy of the recovery is further illustrated in the three plots in Figure 1.

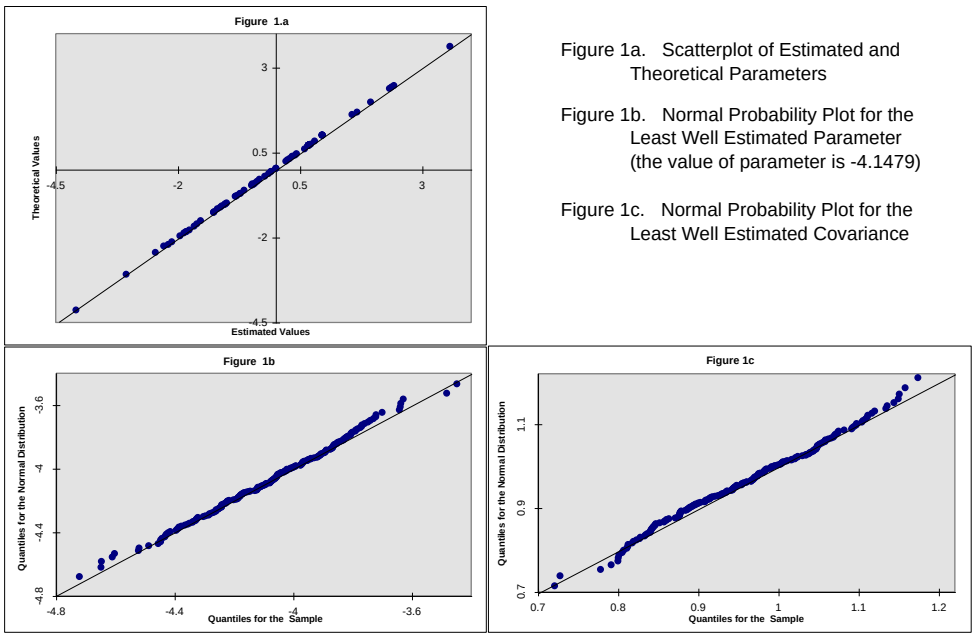

Figure 1a. Scatterplot of Estimated and Theoretical Parameters

Figure 1b. Normal Probability Plot for the Least Well Estimated Parameter (the value of parameter is -4.1479)

Figure 1c. Normal Probability Plot for the Least Well Estimated Covariance

Figure 1. Set 1 (2000 nodes)

Figure 1a is a scatterplot of the means of estimated values for each parameter against the generating values of the corresponding parameters for the 2000 nodes estimation. For reference an identity line has been shown. Figure 1b is a normal probability plot of the estimated values for the single item parameter that was estimated least well and Figure 1c is the normal probability plot for the covariance that was least well estimated. For these least well cases the hypothesis of normality of data was tested and they were not rejected by both Anderson-Darling (the A-

Squared statistics are equal to 0.203 for the item parameters and 0.448 for the covariances) and Kolmogorov-Smirnov ($D_{199} = 0.029$ for parameter of difficulty and $D_{199} = 0.04$ for covariance) tests.

SET TWO: The Influence of Size of Sample.

In set two the six dimensional model was again considered but with 2000 observations. An analysis with 2500 nodes was also added. The results for this set are given in the Table 2.

The results in Table 2 are consistent with those reported in Table 1. It is necessary to note again that none of the means of the estimated parameters is significantly different from theirs generating values. However, as with the previous simulation set while the Hotelling's test decreases as the number of nodes increases it is significant for all numbers of nodes. Interestingly, the T^2 is generally larger for these analyses than in the analyses reported in Table 1.

Table 2. Summary Results Simulations Set Two.

No. of nodes	Largest absolute value of *t*-statistics (*t*) and corresponding quantile (p) for these statistics				Hotelling's T^2-statistics	Average number of iterations
	t for ξ	*p* for ξ	*t* for Σ	*p* for Σ		
500	1.152	0.875	1.683	0.954	43.33	93
1000	0.763	0.777	1.100	0.864	20.74	103
1500	0.606	0.728	0.847	0.801	9.93	96
2000	0.496	0.690	0.666	0.747	5.92	105
2500	0.437	0.669	0.542	0.706	4.41	102

This suggests the larger sample is providing greater power to reject the null hypothesis. Scatterplots and normal plots for these data show results that are essentially identical to those shown in Figure 1 and are therefore not reported.

SET THREE: The Influence of Convergence Criteria.

In set one and two the estimation was terminated when changes in the parameter estimates from one iteration to the next became less than 0.001. In this set the alternative criteria of the change in the likelihood is considered, and the estimation was terminated when the change in the loglikelihood was less than 0.001. Samples

of size 2000 were used in this simulation set, and a simulation with 3000 nodes was added.

The results reported in Table 3 illustrate that the termination criterion based upon estimate change gives slightly better results.

Table 3. Summary Results for Simulations Set Three.

No. of nodes	Largest absolute values of *t*-statistics (*t*) and corresponding quantiles (*p*) for them				Hotelling's T^2 - statistics	Average number of iterations
	t for ξ	*p* for ξ	*t* for Σ	*p* for Σ		
500	1.092	0.862	1.585	0.943	37.68	96
1000	0.760	0.776	1.060	0.856	21.99	97
1500	0.511	0.695	0.770	0.779	10.20	101
2000	0.411	0.659	0.627	0.735	6.72	107
2500	0.330	0.629	0.524	0.700	4.64	111
3000	0.264	0.604	0.444	0.672	3.62	113

In the following simulations the change in parameter estimate criterion is used for set four and five and the change in likelihood criterion is used for set six.

SET FOUR: The Influence of Number of Categories by each Dimension.

In set four two sets of 200 samples were generated, one set containing samples of 1000 observations and one set containing samples of 2000 observations. The six dimensional covariance matrix was used with four dichotomous items on each dimension. Therefore on each dimension scores could only range from 0 to 4. The estimation was undertaken using 500, 1000, 1500, 2000, 2500 and 3000 nodes. The results are reported in Table 4.

Table 4. Summary Results for Simulation Set Four.

No. of observ.	No. of nodes	Largest absolute values of *t*-statistics (t) and corresponding quantiles (*p*) for them				Hotelling's T^2 - statistics	Average number of iterat.
		t for ξ	*p* for ξ	*t* for Σ	*p* for Σ		
1000	500	0.234	0.593	0.532	0.703	14.09	94
1000	1000	0.287	0.613	0.784	0.784	36.00	27
1000	1500	0.307	0.620	0.955	0.830	70.93	22
1000	2000	0.323	0.627	0.925	0.823	68.50	22
1000	2500	0.311	0.622	0.972	0.835	83.75	21
1000	3000	0.321	0.626	1.017	0.845	80.36	21
2000	500	0.298	0.617	0.836	0.798	20.51	31
2000	1000	0.347	0.636	1.104	0.865	42.66	25
2000	1500	0.371	0.645	1.300	0.903	70.42	30
2000	2000	0.379	0.647	1.385	0.917	116.98	21
2000	2500	0.389	0.651	1.457	0.927	135.32	21
2000	3000	0.388	0.651	1.446	0.926	137.03	21

Interestingly Table 4 suggests that in this case the parameter recovery becomes worse as the number of nodes is increased. A possible explanation for this is that there is a small bias in the estimation and when the estimation is made more accurate (by increasing the number of nodes) the bias becomes more evident. Note, that if the number of observations in each sample is increased from 1000 to 2000 this effect becomes even clearer.

In Figure 2 scatterplots of the generating and mean estimated value's are reported for the 500 and 2000 nodes simulations. In both cases the variances are slightly (but systematically) underestimated. It is likely that with only four possible score points on each dimension that the variance of the latent variable could not be well estimated. It is also interesting to note from the last column of Table 4 that the average number of iterations has changed its order of magnitude when compared to the same column in Table 1.

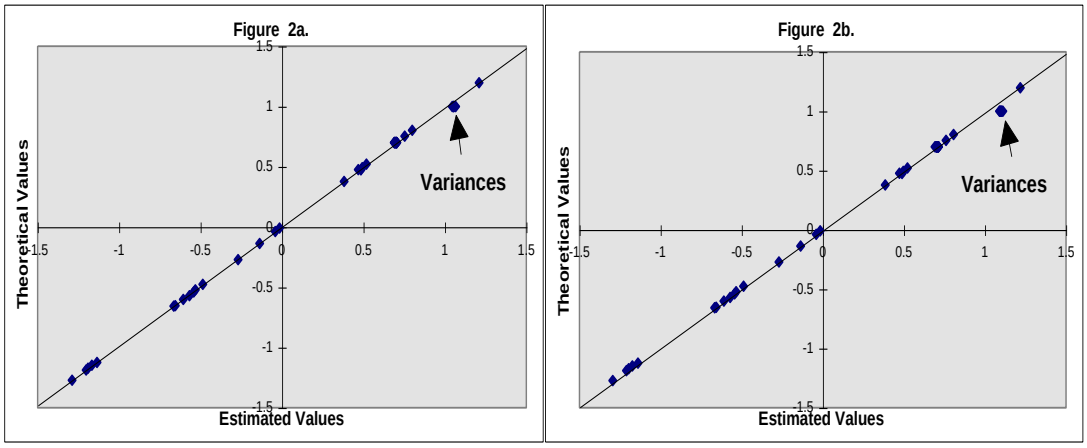

Figure 2 . Set Four - Scatterplots of Estimated and Theoretical Parameters (Figure 2a - for 500 nodes; Figure 2b - for 2000 nodes).

Normal probability plots for the worst estimated of the item parameters and the worst estimated of the covariance parameters for 500 and 2000 nodes respectively look similar to Figures 1b and 1c. In each case the estimates are shown to be close to normal. Further, the results in Table 4 would fail to reject the null hypothesis with $\alpha = 0.7$ for the item difficulty parameter and $\alpha = 0.25$ for the covariances.

SET FIVE: Case with Negative Covariances

In this fifth set, 200 samples of 2000 observations were simulated using the four-dimensional covariance matrix, (21). Twelve items of three categories each were allocated to each dimension, giving a score range from 0 to 24 on each dimension.

The covariance matrix used was deliberately chosen to contain a diverse range of values, including negative values. While such values are unlikely in educational outcomes, they may not be surprising in other applications.

Table 5. Summary Results Simulations Set Five.

No. of nodes	Largest absolute value of t-statistics (t) and corresponding quantile (p) for this statistics				Hotelling's T^2-statistics	Average number of iterations
	t for ξ	p for ξ	t for Σ	p for Σ		
500	1.568	0.942	2.251	0.988	46.70	54
1000	0.978	0.836	0.812	0.791	9.01	59
1500	0.486	0.687	0.933	0.825	11.81	73
2000	0.439	0.670	0.905	0.817	5.51	63

The results for this simulation are reported in Table 5. They show the same patterns as have been reported previously.

SET SIX: Estimations with Gauss Quadrature

As a point of comparison for the Monte Carlo method it is worth undertaking a simulation with multidimensional Gauss-Hermite quadrature. This comparison was chosen because it is well known that in one dimension Gauss-Hermite quadrature provides the most accurate results. In set six Gauss-Hermite quadrature with 2401 nodes (7 nodes in each dimension) was applied using the same specifications as in set five. The results of these analyses are shown in Table 6.

Table 6. Summary Results Simulations Set Six.

No. of nodes	Largest absolute value of *t*-statistics (*t*) and corresponding quantile (*p*) for this statistics				Hotelling's T^2-statistics	Average number of iterations
	t for ξ	*p* for ξ	*t* for Σ	*p* for Σ		
2401	0.945	0.828	3.125	0.999	47.06	579

The results as reported clearly show that this method is inferior to the Monte Carlo method that we have used earlier. In Figure 3 scatterplots that show the relationship between the generating values and the estimated values using the Monte Carlo method with 2000 nodes (Figure 3a) and the Gauss-Hermite quadrature with 2401 nodes (Figure 3b) are reported.

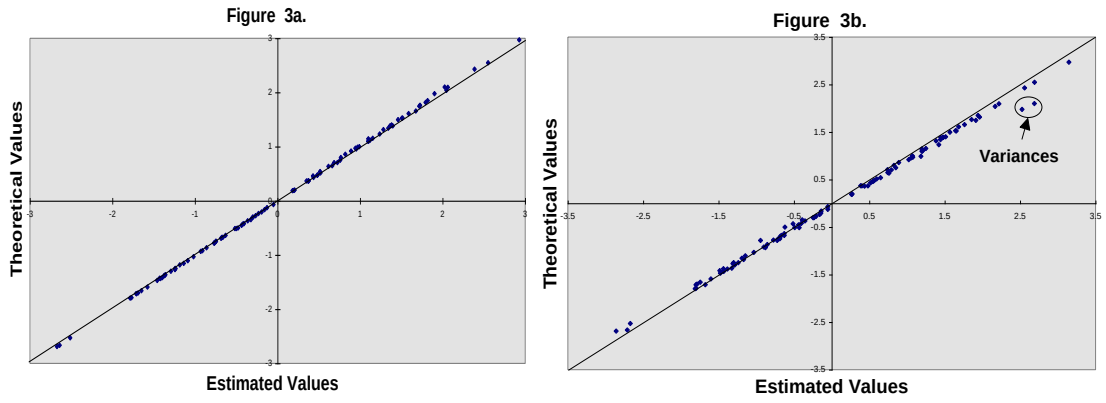

Figure 3. Scatterplots of Estimated and Theoretical Parameters (Figure 3a - Set Five with 2000 nodes; Figure 3b - Set Six with 2401 nodes)

There is a noticeable bias in the estimation of two variance parameters and greater variance in the item parameter estimates, when using the Gauss-Hermite quadrature method. At the same time normal probability plots shown in Figure 4 look reasonable for both 2000 nodes and for 2401 nodes.

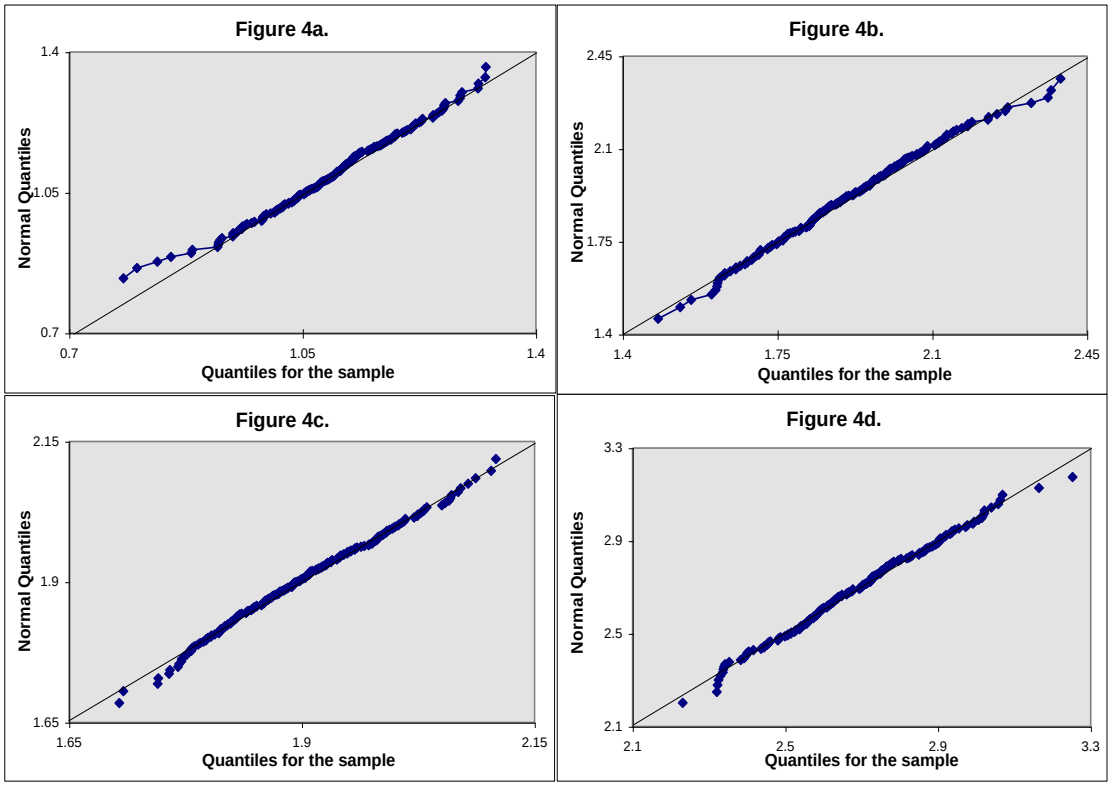

Figure 4. Normal Probability Plots for the worst Estimated Parameters and Covariances (Figure 4a - for parameter with value 1.144 in Set Five; Figure 4b - for parameter with value 1.746 in Set Six; Figure 4c - for variance with value 1.978 in Set Five; Figure 4d - for variance with value 2.1 in Set Six).

6. Illustration with Real Data

To illustrate a real application of the model, and estimation method, an exploration of the dimensionality of the Australian Population Two science data collected as a part of the Third International Mathematics and Science Study (TIMSS) (Lokan, Ford & Greenwood (1996)) was undertaken. The data consisted of a sample of 12 852 students who each responded to approximately 40 items. In the TIMSS design, a pool of 140 items (102 multiple choice, 23 short answer and 15 extended response) was distributed over eight, linked test booklets (Adams & Gonzalez 1996). Each sampled student was assigned a single booklet. The pool of 135 items was constructed so that it included items that assessed five content sub-domains. Table 6 shows the distribution of items across the booklets, both by sub-domain and item type.

The interest in analysing these data is two-fold. First, it is interesting to determine if the five sub-domains are distinct latent variables. Second, there is an interest in determining whether the different item types tap different latent variables.

To study these questions three models were tested with Australian TIMSS data. In the first analysis a uni-dimensional marginal model was tested. It gave 496405.993 for the deviance (-2*loglikelihood). In the second model, where items were allocated to three dimensions according to their type, the resulting deviance was 496232.145.

Table 6. Distribution of items across the booklets by sub-domains and item types (MC - Multiple Choice; SA - Short Answer; ER - Extended Respond).

Booklet	Earth Science			Life Science			Physics			Chemistry			Environment and Other		
	MC	SA	ER	MC	SA	ER	MC	SA	ER	MC	SA	ER	MC	SA	ER
1	7	-	-	9	-	-	10	-	-	2	-	-	5	1	-
2	5	-	3	8	1	3	9	-	-	6	1	-	3	-	-
3	6	-	-	11	-	-	11	1	-	2	-	-	2	1	-
4	6	-	-	11	-	3	8	-	3	3	-	-	2	-	-
5	5	-	-	6	-	1	7	2	-	5	-	-	3	-	-
6	5	-	-	9	-	-	7	1	2	5	1	3	6	2	-
7	7	-	4	6	2	-	10	1	-	4	-	-	1	-	-
8	5	2	-	3	2	1	9	4	-	3	1	-	2	-	-

For the third analysis, a five-dimensional model with items allocated to dimension by sub-domain, a deviance of 495698.130 was obtained. In the second model there are five parameters more than in the first one, so as the difference 496405.993 - 496232.145 = 173.848 is significantly larger than 15.1 which correspond to 0.01 quantile of the chi-square distribution with 5 degree of freedom, it means that the second model is significantly better than the first one. Similar comparison of the third model with the first one gives value 807.863 which is much higher than 23.2 which is correspond to chi-square distribution with 10 degree of freedom. This analysis shows that both multidimensional models fit significantly better than unidimensional model. Note, that as the second and third models are not hierarchical it is not possible to formally compare their fit. Perhaps it is worth noting, however, that the decrease in the deviance for each degree of freedom increase is greater for the *sub-domain* based model than it is for the *item-type* model At the same time Table 7 shows that the correlation between the *item-type*

dimensions are, on average, slightly lower than those for the sub-domain based dimensions.

Table 7 Estimated Covariance and Correlation Matrices for the Australian Science Data

	Scale				
	Earth Science	Life Science	Physics	Chemistry	Environment and Other
	Covariances				
Earth Science	0.718				
Life Science	0.716	0.872			
Physics	0.550	0.608	0.512		
Chemistry	0.816	0.916	0.694	1.158	
Environment and Other	0.830	0.944	0.706	1.072	1.262
	Correlations				
Earth Science	1.000				
Life Science	0.905	1.000			
Physics	0.908	0.910	1.000		
Chemistry	0.894	0.912	0.901	1.000	
Environment and Other	0.872	0.900	0.878	0.887	1.000

	Scale		
	Multiple Responce	Short Answer	Extended Response
	Covariances		
Multiple Response	0.695		
Short Answer	0.746	1.009	
Extended Response	0.730	0.895	1.013
	Correlations		
Multiple Response	1.000		
Short Answer	0.891	1.000	
Extended Response	0.870	0.885	1.000

7. Discussion

The application of multidimensional item response models has been limited by a number of factors, including but obviously not restricted to, the fact that existing methods can only be applied when the number of dimensions is limited to three or perhaps four. In this paper we have illustrated how the use of Monte Carlo integration methods allows the application of a Rasch based multidimensional to at least six dimensions and we have illustrated this through application to data collected as part of the Third International Mathematics and Science Study. Benefits which would accrue from the application of such a model include improved estimation of the correlations between latent quantities, improved reliability of estimation for each dimension and better approaches modelling complex performance data (Adams, Wilson and Wang, 1997).

The exploration of the Monte Carlo method with multidimensional item response models has only just begun and many important questions remained to be answered. Perhaps the most important one is, at what number of dimensions does the Monte Carlo method become a better choice than quadrature. When using quadrature we have found that as few as eight nodes are adequate for estimation in one dimension and that the desirable number of nodes increases exponentially with the dimensionality (Wang, 1994). When using Monte Carlo methods our preliminary work suggests that many hundreds of nodes may be necessary in one dimension, but the desirable number of nodes increases linearly with the dimensionality.

A second important question is, what is the maximum number of dimensions that can be estimated with the Monte Carlo method? We can see no substantial difficulty in applying the approach in many-many dimensions and suspect that the limiting factor in applying the model in high dimensions will be the suitability of the data for providing stable estimates of the covariances and not the estimation complexity.

A third area for future development is the consideration of the use of the Monte Carlo method in conjunction with a Newton-Raphson, rather than an EM algorithm. In context of item response models the EM algorithm is analytically simpler than a Newton-Raphson method but it increases the number of iterations substantially. A more efficient approach may be to use Monte Carlo integration in conjunction with a Newton-Raphson method.

REFERENCES

Ackerman, T. A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement*, 29, 67-91.

Adams, R. J., & Gonzalez, E. J. (1996). The TIMSS Test Design. In M. O. Martin and D. L. Kelly, (Eds.), *Third International Mathematics and Science Study. Technical Report*. Vol. I: Design and Development. Boston: Chestnutt Hill, Massachusets.

Adams, R. J., & Wilson, M. R. (1997). Formulating the Rasch model as a mixed coefficients multinomial logit. In G. Engelhard and M. Wilson, (Eds.), *Objective measurement: Theory into Practice*. Vol III. Norwood, NJ: Ablex.

Adams, R.J., Wilson, M.R. and Wang, W.C. (1997) The multidimensional random coefficients multinomial logit. *Applied Psychological Measurement*,

Adams, R. J., Wilson, M. R. and Wu M. (1997). Multilevel Item Response Models: An Approach to Errors in Variables Regression. *Journal of Educational and Behavioral Statistics*, 22. 47-76.

Adams, R.J. and Wu, M.L. (2002) *Programme for International Student Achievement: Technical Report*. Organisation for Economic Cooperation and Development. Paris.

Andersen, E. B. (1985). Estimating latent correlations between repeated testings. *Psychometrika*, 50, 3-16.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.

Batley, R..M., & Boss, M. W. (1993). The effects on parameter estimation of correlated dimensions and a distribution — restricted trait in a multi-dimensional item response model. *Applied Psychological Measurement*, 17, 131-141.

Beaton, A.E., (1987). *Implementing the new design: The NAEP 1983-84 Technical Report*. Report No. 15-TR-20. Educational Testing Service.

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

Camilli, G. (1992). A conceptual analysis of differential item functioning in terms of a multidimensional item response model. *Applied Psychological Measurement*, 16, 129-147.

Glas, C. A. W. (1992). A Rasch model with a multivariate distribution of ability. In, M. Wilson (Editor), *Objective Measurement: Theory into Practice*. Volume. 1 (pp. 236-258). Norwood, NJ: Ablex Publishing Corporation.

Embretson, S. E. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika*, 56, 495-515.

Fischer, G. H. (1973). The linear logistic model as an instrument in educational research. *Acta Psychologica*, 37, 359-374.

Kalos, M. H. & Whitlock, P. A. (1986). *Monte Carlo Methods*. Volume I: Basics: New York: John Wiley & Sons.

Kelderman, H., & Rijkes, C. P. M. (1994). Loglinear multidimensional IRT models for polytomously scored items. *Psychometrika*, 59, 149-176.

Linacre, J. M. (1989). *Many-faceted Rasch measurement*. Chicago: MESA Press.

Lokan, J., Ford, P. & Greenwood, L. (1996). *Maths & Science on the Line. Australian Junior Secondary Students' Performance in the Third International Mathematics and Science Study*. Melbourne: ACER.

Luecht, R. M., & Miller, R. (1992). Unidimensional calibrations and interpretations of composite traits for multidimensional tests. *Applied Psychological Measurement*, 16, 3, 279-293.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.

Mislevy, R.J. (1985). Estimation of latent group effects. *Journal of the American Statistical Association*, 80, 993-997.

Ralston, A., & Rabinowitz, P. (1978). *A first course in numerical analysis* (2 nd ed.). Tokyo: McGraw Hill.

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

Reckase, M. D., & McKinley, R. L. (1991). The discriminating power of items that measure more than one dimension. *Applied Psychological Measurement*, 15, 361-373.

Tanner, M.A. (1993). *Tools for statistical inference*, Second Edition. Springer-Verlag. New York.

Wang, W.C. (1994). Implementation and application of the multidimensional random coefficients multinomial logit. *Unpublished doctoral dissertation*. University of California, Berkeley.

Wang, W., Wilson, M. R., and Adams, R. J. (under preparation). Implications and applications of the multidimensional random coefficients multinomial logit Model.

Whitely, S. E. (1980). Multicomponent latent trait models for ability tests. *Psychometrika*, 45, 479-494.

Wilson, M. R. (1992). The partial order model: An extension of the partial credit model. *Applied Psychological Measurement*, 16, 309-325.

Zwick, R. (1992) (editor) *Journal of Educational Statistics*, 17. Special Issue: National Assessment of Educational Progress.