IOSR Journal of Research & Method in Education (IOSR-JRME)
e-ISSN: 2320–7388,p-ISSN: 2320–737X Volume 6, Issue 6 Ver. III (Nov. - Dec. 2016), PP 07-20
www.iosrjournals.org

A MIRT Model for Tests with Multiple Subtests

Alvaro Mauricio Montenegro[a], Luisa Fernanda Parra[a], Emilio Berdugo[b]
[a]*Universidad Nacional de Colombia*
[b]*Universidad Santo Tomás*

Abstract: *The model proposed in this paper has been thought to be used in large-scale assessment tests designed explicitly to measure more than one latent trait. Those tests are usually split into subtests, in which each subtest is designed to measure mainly a unique unidimensional latent trait. Admission tests of some universities are typical examples of that type of tests.*
Keywords: *Multidimensional Item Response Theory Model, Multiple Subtests, Large-Scale Assessment Tests, Bayesian inference*

I. Introduction

In admission tests of some universities, it is common to split the test into several subtests. Each subtest is designed to measure a main latent trait. For this type of tests, there are at least three estimation procedures that are used, which are described as follows. The first procedure consists in the estimation of the parameters of each subtest separately, using unidimensional item response theory (UIRT) models. A unidimensional latent trait is estimated for each examinee from each subtest. If a global scale is required, the unidimensional latent traits are standardized and an average is computed. This average is a global synthetic trait.

The second procedure is based on the use of a multidimensional item response theory (MIRT) model, (1; 2; 3). This procedure requires the previous specification of the dimension of the latent trait space. In this procedure, the covariance matrix of the latent trait vector may be estimated. To estimate the latent trait measured by each subtest, a reference direction is calculated for each subtest. The reference directions of the subtests are computed from the directions of the items belonging to each subtest. Then, a composite is computed for each subtest. Each component of the latent trait vector may not have a direct interpretation. If a unidimensional global measure is required, the reference composite of the test is computed, (4).

The third procedure is based on the use of a simple structure model, called multi-unidimensional model. In this case, each subtest is modeled as a unidimensional test and a latent trait vector is estimated. Each component of such vector is the unidimensional latent trait that is measured by a subtest. See, for example, (5; 6; 7).

In this paper we introduce the Multiple Subtest MIRT (MSMIRT) model. The model is proposed to be used in large scale tests which include multiple subtests. It is assumed that each subtest measures essentially a unidimensional latent trait and that the dimension of the latent trait space is at most the number of subtests. The paper is organized as follows. In section 2 we review the equivalence between the MIRT models and the Factor Analysis models; section 3 is a discusion about the concept of dimension; we distinguish between the dimension of a test and the dimension of the data; section 4 is an introduction to the concept of composite; this concept is required to obtain the score of each subtest; section 5 presents the proposed model, including the main aspects of the estimation procedure; section 6 is the application of the model to the admission test that was ran in Universidad Nacional de Colombia in the second semester of 2009; finally, section 7 is the discussion about the main results of the paper.

II. Equivalence of the MIRT model and the Factor Analysis Model

In this section we introduce the equivalence between the compensatory MIRT models and the Factor Analysis models. We show that the Bayesian techique of augmented variables common in estimation procedures of the parameters of MIRT models uses some ideas derived from this equivalence.

Let us assume that we have a test with K dichotomous items which is responded by N examinees. Let us consider a classical two parameter normal ogive MIRT model. The model is specified by the probability of success of examinee i to item j, given by

$$P(Y_{ij} = 1 | \boldsymbol{\alpha}_j, \boldsymbol{\theta}_i, \gamma_j) = \Phi(\boldsymbol{\alpha}^t_j \boldsymbol{\theta}_i - \gamma_j), \qquad (1)$$

where Y_{ij} is the random variables which represents the response of examinee i to item j, $\boldsymbol{\alpha}_j = (a_{j1}, \cdots, a_{jd})^t$ is a vector of slopes of item j, $\boldsymbol{\theta}_i = (\theta_{i1}, \cdots, \theta_{id})^t$ is the latent trait vector of examinee i, γ_j is an intercept parameter associated to the difficulty of item j, and d is the dimension of the latent trait vector, (1). It is assumed that the $\boldsymbol{\theta}$-vectors are a sample from a random vector $\boldsymbol{\Theta}$ that has normal distribution $N_d(\mathbf{0}, \boldsymbol{\Sigma})$, in which the diagonal elements of $\boldsymbol{\Sigma}$ are 1. The latent trait space is the Euclidean space $\mathbf{R}^d$. The value d is called the *dimension of the latent trait space*. We adopt this definition for the MSMIRT model that is introduced in this paper, because we propose a special kind of MIRT model. The response pattern of an examinee is an element of $\{0,1\}^K$. Vector $\boldsymbol{\theta}_i$ has a reduced dimension $d < K$. In general the latent trait vector of an examinee is a representation of his/her response pattern in an Euclidean space of reduced dimension.

The reduction of the dimension can be thought in terms of a factorial analysis. This is not new. When (8) proposed the first formal method to estimate the parameters in a MIRT model, they supposed that the responses of the examinees could be modeled by the introduction of continuous latent variables Z_{ij} that govern the response process.

(9) proposed the factor analysis of dichotomized variables based on the same latent variables used by (8). (10)stated that the factor analysis of dichotomized variables is equivalent to the item response theory. They used in their proof the same continuous latent variables Z_{ij}. Additionally, (11;12) used this approach to propose a framework for the multidimensional item response theory.

In the Bayesian field the variables Z_{ij} are called augmented variables, and they are introduced to build Gibbs samplers that are easy to implement. In the field of the item response theory, this type of variables was introduced by (13). Similar approaches to estimate the parameters of MIRT models have been used by (14), (15) , (6), and other authors. In this work, the variables Z_{ij} are used in the estimation procedure.

The use of latent continuous variables in the classical item response theory is different from the use in the Bayesian field. However, the continuous latent variables Z_{ij} are essentially the same. This characteristic of the latent variables is exploited in this work. For fixed values of $\boldsymbol{\alpha}_j, \boldsymbol{\theta}_i$, and γ_j. Let Z_{ij} be the random variable defined as

$$Z_{ij} = \boldsymbol{\alpha}^t_j \theta_i - \gamma_j + e_{ij},\ e_{ij} \textit{ has distribution } N(0,1). \qquad (2)$$

Thus, we have that

$$\begin{aligned} &P(Z_{ij} \geq 0 \mid \boldsymbol{\alpha}_j, \boldsymbol{\theta}_i, \gamma_j) = 1 - P(e_{ij} \leq -(\boldsymbol{\alpha}^t_j \boldsymbol{\theta}_i - \gamma_j)) \\ &= 1 - \Phi(-\boldsymbol{\alpha}^t_j \boldsymbol{\theta}_i + \gamma_j) = \Phi(\boldsymbol{\alpha}^t_j \boldsymbol{\theta}_i - \gamma_j) \\ &= P(Y_{ij} = 1 \mid \boldsymbol{\alpha}_j, \boldsymbol{\theta}_i, \gamma_j). \qquad (3) \end{aligned}$$

Hence, the variable $Y_{ij} = 1$ if $Z_{ij} > 0$ and $Y_{ij} = 0$ if $Z_{ij} \leq 0$. So the values of Y_{ij} are determined by the values of Z_{ij}. In other words, the latent continuous variable Z_{ij} governs the response process of variable Y_{ij}. Let $\mathbf{A}_{K \times d}$ be the matrix whose rows are the slope vectors $\boldsymbol{\alpha}_j$. Let $\boldsymbol{\Theta}$ be a random vector distribute as $N_d(\mathbf{0}, \boldsymbol{\Sigma})$. The latent traits of the examinees are considered samples from vector $\boldsymbol{\Theta}$. Let $\mathbf{e}$ be a random vector distributed as $N_K(\mathbf{0}, \mathbf{I}_K)$, where $\mathbf{I}_K$ represents the identity matrix of size K. It is assumed that $\boldsymbol{\Theta}$ and $\mathbf{e}$ are independent. Let $\boldsymbol{\gamma}$ be the vector of intercepts in the MIRT model. Let $\mathbf{Z} = (Z_1, \cdots, Z_K)^t$ be the random vector defined as

$$\mathbf{Z} = \mathbf{A}\boldsymbol{\Theta} - \boldsymbol{\gamma} + \mathbf{e}. \qquad (4)$$

Then,

$$\boldsymbol{Z} \sim N(-\boldsymbol{\gamma}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^t + \boldsymbol{I}_K) \qquad (5)$$

and

$$[\boldsymbol{Z}|\Theta = \boldsymbol{\theta}] \sim N(\boldsymbol{A\theta} - \boldsymbol{\gamma}, \boldsymbol{I}_K). \quad (6)$$

Let $\mathbf{y} = (y_1, \cdots, y_K)^t$ be the random vector representing any response pattern. Then, for $j = 1, \cdots, K$ we have that

$$p\{[Z_j \mid \boldsymbol{\theta}] > 0\} = \int_0^\infty \frac{1}{\sqrt{2\pi}} \exp\{-1/2(z - (\boldsymbol{\alpha}_j^t \boldsymbol{\theta} - \gamma_j))^2\} dz$$
$$= \Phi(\boldsymbol{\alpha}_j^t \boldsymbol{\theta} - \gamma_j)$$
$$= P(y_j = 1 \mid \boldsymbol{\alpha}_j, \boldsymbol{\theta}, \gamma_j) \quad (7)$$

Consequently, the random response pattern $\mathbf{y}$ is governed by the random vector vector $\mathbf{Z}$, i.e., the particular values of a response pattern are determined by the values of the random vector $\mathbf{Z}$. Furthermore, equation (4) represents a factorial analysis model with the nice property that the perturbation term $\mathbf{e}$ has distribution $N_K(\mathbf{0}, \mathbf{I}_K)$. This perturbation term is the Bayesian residual in the Bayesian item response theory models, (16).

Equation (4) has some important consequences. The equation expresses a relation between the classical models of the item response theory, the factorial analysis of dichotomized variables, and the technique of augmented variables used in some Bayesian procedures of estimation in the item response theory, (13). The dimension of the latent trait space may be determined as the minimum number of factors required to have a good representation of the random vector $\mathbf{Z}$. Obviously, this is only a theoretical construction because the variables Z_{ij} cannot be measured directly. However, they can be predicted using a Gibbs sampler algorithm. In this work, a data augmentation Gibbs sampler (DAGS) algorithm was implemented to estimate the parameters of the MSMIRT model; in the algorithm, the variables Z_{ij} were used, (4). In the classical item response theory, one of the more used strategies to determine the dimension of the latent trait space is through the eigenvalue structure of the tetrachoric correlation matrix. The tetrachoric correlation between two binary variables is the Person correlation one would obtain if the two variables were measured continuously. See for example (17). Some procedures have been developed to estimate the tetrachoric correlations. A recent function to estimate the tetrachoric correlations can be found in package *polycor*, (18), for R (2016). However, the estimated matrix of the sample tetrachoric correlation obtained from the classical algorithms is often nonpositive definite, (19).

We propose the following strategy to detect the dimension of the latent trait space. First, identify the dimension of the latent trait space through a principal component analysis; second, confirm the dimension of the space, based on the eigenvalue structure of the tetrachoric correlation matrix. This matrix can be estimated from the variables Z_{ij}, which can be predicted inside of the DAGS algorithm. For details, see (4).

III. A discussion about the concept of dimension

In the item response theory, an important amount of research has been devoted to determine whether the assumption of unidimensionality is reasonable, see, for example(20; 21), (22), (23), (24), (25), (26), (27), (28), (29).

The assumption of unidimensionality is a strong simplification of the reality. Unidimensionality can only be approximated, (30). (31) argues that related problems of dimensionality and bias of items are approached in an arbitrary and oversimplified fashion. Humphreys pointed out that a dominant attribute (i.e., dominant dimension) results from an attribute overlapping many items and asserts that attributes common to relatively few items or even unique to individual items are unavoidable and indeed are not detrimental to the measurement of a dominant dimension. In the same way, (11) argues the existence of "minor components" in factor analytic modeling of test data, and the existence of multiple determinants, which are common to some items.

According to (3), dimensionality is a property of the sample of examinees' latent traits that take a test, and it is not a property of the test itself. A common definition of the dimensionality is: the minimum dimension of the ability space required to obtain conditional independence. The dimensions required to have conditional independence in a test can change from a population to another. Reckase states that the number of dimensions needed to model accurately the relationships in the item response matrix dependent on two aspects of the data collection process: the number of dimensions on which the people taking the test differ and the number of dimensions on which test items are sensitive to differences. For example, in extreme cases, it is possible to imagine a group of individuals, which have been carefully selected to be identical on all dimensions except one. In this hypothetical case, the item response matrix that results from administering the test to them can represent differences on only one dimension. On the other hand, if the set of test items used are only sensitive to differences along one of the dimensions of variability of the examinee population, the resulting data will be essentially unidimensional.

(30), introduced the concept of essential unidimensionality. The main idea of Stout is that even though the ability space is multidimensional, the set of items used in a test may be sensitive mainly to differences along one of the dimensions, and the statistical tests to asses the unidimensionalidad can reject that assumption. He proposed to replace the usual assumption of unidimensionality by a weaker and arguably more appropriate statistically testable assumption of essential unidimensionality. Essential unidimensionality implies the existence of a "unique" unidimensional latent ability. To test essential unidimensionality, (29) developed the DIMTEST procedure.

The concept of essential unidimensional can be generalized to essential dimensionality. Under this perspective, the items of a test can be grouped in clusters in such a way that the items in each cluster are sensitive mainly to differences along one direction in the latent trait space. In this case, the essential dimensions that are measured by the cluster of items are not necessarily orthogonal. Procustes methodology permits to build non orthogonal rotations onto the ability space, see, for example, (32). Such non orthogonal latent traits become orthogonal, through linear transformations that do not change the probability patterns, but changing the correlation of the latent traits.

A more recent discussion about the concept of dimensionality in the item response theory is due to (33) and (34). (33) proved a result that he called the *submodel theorem*. The theorem states that, a multidimensional model which has a positive continuous item response function, is equivalent to some of its unidimensional submodels in the sense that, the multidimensional model and each one of those unidimensional submodels predict the same probability patterns. (34) reported experimental results in the same way, based on the use of nonparametric multidimensional scaling to synthesize a multidimensional model from several approximate one-dimensional models.

IV. Composites of latent traits

When the experts design a test with multiple subtests, their objective is the estimation the latent traits that are measured by each subtest. Hence, the design of the test explicitly leads to define a first concept of dimension. It is natural to define the *dimension of a test* as the number of subtests, or as the number of latent traits that the entire test attempts to measure. We adopt this definition. We call this latent traits as *main latent traits*. By definition the main latent traits have a direct interpretation derived from the design of the test.

On the other hand, it is not realistic to assume that the dimension of the test coincides with the dimension of the data. The binary responses to the items can be considered as partial signs of the latent traits of the examinees. If a test has K items, the response pattern of any examinee is a vector in the space $\{0,1\}^K$. In a MIRT model we assume that the latent traits are points in a Euclidean space of dimension d, such that $d < K$. The components of a vector of latent traits will be called *basic latent traits*. The basic latent traits may not have a direct interpretation.

Given a latent trait vector $\boldsymbol{\theta}$ and a unitary vector $\boldsymbol{\beta}$, the scalar product $\boldsymbol{\beta}^t\boldsymbol{\theta}$ is called a composition. Let α_j be the Euclidean norm of the vector $\boldsymbol{\alpha}_j$, and let $\boldsymbol{\beta}_j = \boldsymbol{\alpha}_j/\alpha_j$. According to previous section, the latent variable Z_j can be written as

$$Z_j = \alpha_j \beta_j^t \theta - \gamma_j + e_j, \quad e_j \text{ has distribution } N(0,1) \qquad (8)$$

where $\boldsymbol{\beta}_j$ is a unit vector called the direction of item j. (4) showed that the vector $\boldsymbol{\beta}_j$ is the direction along which the item j discriminates better. This means that the item j discriminates better between the values of the synthetic latent trait given by the composite $\boldsymbol{\beta}_j^t\boldsymbol{\theta}$.

According to (30), a test is *essentially unidimensional* if all of its items are sensitive, mainly to differences along one direction in the latent trait space. Clearly, if a test is essentially unidimensional, the direction vectors point in roughly the same direction. (35) define the reference direction of a cluster of items as the principal direction of the direction of the items in the cluster. They showed that the reference direction of a subtest defines the composite that is estimated if a UIRT model is used to fit the data of the subtest.

Along the reference direction, the subtest discriminates better on average. Furthermore, if the test is essentially unidimensional, a good unidimensional approximation of the multidimensional model can be obtained by replacing all the item directions with the reference direction of the test and making some changes in the item parameters. Consequently, if a test is essentially unidimensional there is a UIRT model that fit well the data, although the tests of unidimensionality can fail. In this work, we assume that the test is split into m subtests. Furthermore, we assume that each subtest is essentially unidimensional, so each subtest is designed to measure a main latent trait. The main latent traits will be composties of basic latent traits.

V. The Multiple Subtest MIRT model

In this section, we introduce the nomenclature and the assumptions of the Multiple Subtests MIRT (MSMIRT) model:

1. The test is split into m subtests. It is assumed that each subtest is essentially unidimensional. Hence, each subtest attempts to measure only one main latent trait. Each subtest has K_v, items, $v = 1,\ldots,m$, so the entire test has $K = K_1 + K_2 + \cdots K_m$ items.
2. It is assumed that the basic latent traits of the examinees are a random sample drawn from a multivariate normal distribution. $N_d(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is a correlation matrix, and $d \leq m$.
3. The main latent traits of the examinees being measured by each subtest are composites of the basic latent trait vectors.
4. The link function is the standard normal ogive, denoted $\Phi(\cdot)$
5. Guessing parameters are not included.

The j th item of subtest v will be called item vj. The MSMIRT model is specified by the probability of success of examinee i to item vj given by

$$P(Y_{vij} = 1 | \alpha_{vj}, \gamma_{vj}, \boldsymbol{\beta}_v, \boldsymbol{\theta}_\mathbf{i}) = \Phi(\alpha_{vj}\boldsymbol{\beta}_v^t\boldsymbol{\theta}_\mathbf{i} - \gamma_{vj}), \quad (9)$$

where α_{vj} and γ_{vj} will be called respectively the slope (the discrimination) parameter and the intercept parameter of item vj. $\boldsymbol{\beta}_v = (\beta_{v1}, \cdots, \beta_{vd})^t$ is a unit vector in the latent trait space that will be called the direction of subtest v and $\boldsymbol{\theta}_i = (\theta_{i1}, \cdots, \theta_{id})^t$ represents the vector of basic latent traits of examinee i.

The classical difficulty parameter is given by $b_{vj} = \gamma_{vj}/\alpha_{vj}$. For ease, the item parameters of item vj will be denoted $\boldsymbol{\zeta}_{vj}$, that is, $\boldsymbol{\zeta}_{vj} = (\alpha_{vj}, \gamma_{vj})^t$. The main latent trait measured by subtest v is given by the composite $\boldsymbol{\beta}_v^t\boldsymbol{\theta}$, $v = 1, \cdots, m$. The expression given by

$$\eta_{vj} = \alpha_{vj}\boldsymbol{\beta}_v^t\boldsymbol{\theta} - \gamma_{vj}, \quad (10)$$

will be called the *linear latent predictor* of the item vj.

For the distribution of the latent trait vectors, other symmetrical distributions are possible like the multivariate t-student distribution. Recently, some authors have proposed asymmetric distributions as the multivariate skew normal and the multivariate skew t-student distributions, (36; 37), (38). In this work, we only consider the multivariate normal distribution.

When the dimension of the latent trait coincides with the dimension of the test, the test will be called a *simple structure test*. In this case, the probability of success of examinee i to item j reduces to

$$P(Y_{vij} = 1 | \boldsymbol{\zeta}_{vj}, \boldsymbol{\theta}_i) = \Phi(\alpha_{vj}\theta_{iv} - \gamma_{vj}), \quad (11)$$

because in that case all the subtest directions can be identified with the vectors of the canonical base of de Euclidean space $\mathbf{R}^d$, as we will see in the next sections.

In the classical literature of MIRT models, the parameter α_{vj} is called the multidimensional discrimination (MDISC) parameter, and the parameter b_{vj} is called the multidimensional difficulty parameter (MDIFF) parameter, (1; 3).

The logistic and the univariate standard normal cdf's are the more extended links. However, recently, asymmetric links have been proposed as the univariate skew normal distribution and the univariate skew t -student distribution, (37), (38). The normal ogive link function was selected in this work for several reasons that include the implementation of the data augmentation Gibbs sampler (DAGS), to estimate the parameters of the model.

5.1 Identifiability of the MSMIRT model

The MSMIRT model is not identifiable. To obtain an identifiable model, we note first that the vectors $\boldsymbol{\beta}_v$ are the reference directions of the subtests, so they are unit vectors. However, this constraint is not sufficient to have an identifiable model. In this section, we propose two parameterizations, one of which permits a nice interpretation of the parameters, including the basic latent traits. The parameterizations are based on the relationship between the main and the basic latent traits. Let $\tilde{\boldsymbol{\Theta}}$ be a $m\times 1$ random vector that represents the main latent traits of the test. Then,

$$\tilde{\boldsymbol{\Theta}}_{m\times 1} = \mathbf{B}_{m\times d}\Theta_{d\times 1} \qquad (12)$$

where $\mathbf{B}$ is the matrix whose rows are the vectors $\boldsymbol{\beta}_v$. The covariance matrix of vector $\tilde{\boldsymbol{\Theta}}$ is given by

$$\mathrm{Cov}(\tilde{\boldsymbol{\Theta}}) = \left(\mathbf{B}\,\boldsymbol{\Sigma}\,\mathbf{B}^{\mathbf{t}}\right) \qquad (13)$$

Equation (12) represents the relationship between the basic latent traits and the main latent traits. From this equation, two parameterizations are considered. Without loss of generality, suppose that matrix $\boldsymbol{\Sigma}$ is positive definite and that the first d rows of matrix $\mathbf{B}$ are linearly independent. Let $\boldsymbol{\Sigma}^{1/2}$ be the square root of matrix $\boldsymbol{\Sigma}$. To state the first parameterization, we rewrite $\tilde{\boldsymbol{\Theta}}$ as:

$$\tilde{\boldsymbol{\Theta}} = \left(\mathbf{B}\boldsymbol{\Sigma}^{1/2}\right)\left(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Theta}\right). \qquad (14)$$

Equation (14) implies that we can assume that $\boldsymbol{\Theta}$ has distribution $N_d(\mathbf{0},\mathbf{I}_d)$. In this case, the basic latent traits are not correlated. In this parameterization, the matrix $\boldsymbol{\Sigma}$ is the identity and consequently it is necessary to estimate the $m\times d$ components of the $\boldsymbol{\beta}_v$ directions.

The second parameterization is inferred as follows. Let $\tilde{\mathbf{B}}$ be the submatrix of $\mathbf{B}$ that contains its first d rows. Then, $\tilde{\boldsymbol{\Theta}}$ can also be written as

$$\tilde{\boldsymbol{\Theta}} = \left(\mathbf{B}\tilde{\mathbf{B}}^{-1}\right)\left(\tilde{\mathbf{B}}\boldsymbol{\Theta}\right). \qquad (15)$$

Equation (15) implies that the first d reference directions are aligned with the coordinate axes. If the variance of the basic latent traits is fixed in 1, as usual, there are only $(m-d)\times d + d\times(d-1)/2$ parameters to estimate. Those parameters correspond to the components of the reference directions that are not aligned with the coordinate axes and the non-diagonal elements of the correlation matrix.

The second parameterization has some advantages. Firstly, there are fewer parameters to be estimated; secondly, the basic latent traits are directly the reference composite of the first d subtests; thirdly, the coordinate axes are set in advance, so, identifiability problems caused by orthogonal transformations of the latent trait space are impossible. Furthermore, it is important to note that in the first parameterization the item directions have a better projection along the corresponding reference composite. Consequently, the values of the item parameters are closer to values of item parameters of the MIRT model, from which the MSMIRT model can be derived.
In the implementation of the DAGS algorithm to estimate the parameters of the MSMIRT model, the second parameterization was used. In the next section, we show how to change from one parameterization to the other.

5.2 Interchangeability between parameterizations

In this section, it is shown how to change from the first parameterization to the second and vice versa. First, assume that the parameters of the first parameterization are available. This means that it was assumed that the latent trait vector has distribution $N_d(\mathbf{0},\mathbf{I}_d)$, and that the reference directions $\boldsymbol{\beta}_v$, $v=1,\cdots,m$ were estimated. To obtain the parameters of the second parameterization, the following transformations are required:

1. Align the first d reference vectors with the coordinate axes, using the equation (15). The reference vectors in the second parameterization are given by

$$\mathbf{B}_v^* = \frac{\widetilde{\mathbf{B}}^{-1}\mathbf{B}_v}{\| \widetilde{\mathbf{B}}^{-1}\mathbf{B}_v \|}, \qquad v = 1, \cdots, m. \qquad (16)$$

where $\|\cdot\|$ denotes the norm of a vector. This implies that in the second parameterization the first d reference vectors are the vectors of the canonical base of $\mathbf{R}^d$.

2. The new covariance matrix is given by $\widetilde{\mathbf{B}}\widetilde{\mathbf{B}}^t$.
3. The new slope parameters are given by

$$\alpha_{vj}^* = \alpha_{vj} \| \widetilde{\mathbf{B}}^{-1}\mathbf{B}_v \|, \qquad v = 1, \cdots, m. \qquad (17)$$

4. The intercept parameters do not change.
5. The new latent trait vectors are given by $\boldsymbol{\theta}^* = \widetilde{\mathbf{B}}\boldsymbol{\theta}$.

Now, suppose that the second parameterization is given. In this case, it is assumed that the latent trait vector has distribution $N_d(\mathbf{0}, \boldsymbol{\Sigma})$ and that the first d reference directions were set to the canonical vectors of $\mathbf{R}^d$. To obtain the parameters of the first parameterization, the following transformations are required:

1. From equation (14), the first d reference vectors are the rows of matrix $\boldsymbol{\Sigma}^{1/2}$. In general the new reference vectors are given by

$$B_v^* = \frac{\boldsymbol{\Sigma}^{1/2}\mathbf{B}_v}{\| \boldsymbol{\Sigma}^{1/2}\mathbf{B}_v \|}, \qquad v = 1, \cdots, m. \qquad (18)$$

2. The new covariance matrix is $\mathbf{I}_d$.
3. The new slope parameters are given by

$$\alpha_{vj}^* = \alpha_{vj} \| \boldsymbol{\Sigma}^{1/2}\mathbf{B}_v \|, \qquad v = 1, \cdots, m. \qquad (19)$$

4. The intercept parameters do not change.
5. The new latent trait vectors are given by $\boldsymbol{\theta}^* = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\theta}$.

5.3 Estimation of the Parameters

Let $p_{vij} = P(Y_{vij} = 1 | \boldsymbol{\theta}_i, \boldsymbol{\beta}_v, \boldsymbol{\zeta}_{vj})$. Let $\boldsymbol{\theta}_{N\times d}$ be the matrix of the latent traits of examinees in the sample. Let $\boldsymbol{\zeta}$ be the vector of all item parameters of the test. Let $\boldsymbol{\beta}_{m\times d}$ be the matrix of the m reference directions of the subtests. Then, under the assumption of local independence, the likelihood function is given by:

$$f(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\zeta}) = \prod_{i=1}^{N}\prod_{v=1}^{m}\prod_{j=1}^{k_v} p_{vij}^{y_{vij}} (1 - p_{vij})^{1-y_{vij}}. \qquad (20)$$

where y_{vij} is the observed response of examinee i to item vj, and $\mathbf{y} = [y_{ij}]_{N\times K}$.

A data augmentation Gibbs sampler (DAGS) algorithm was developed to estimate jointly the item and the latent trait parameters. Following the strategy proposed by (13), we introduced the augmented variables Z_{vij} with distribution $N(\eta_{vij}, 1)$, where $\eta_{vij} = \alpha_{vj}\boldsymbol{\beta}_v^t\boldsymbol{\theta}_\mathbf{i} - \gamma_{vj}$. It is easy to show that if we define $Y_{vij} = 1$ if $Z_{vij} > 0$ and $Y_{vij} = 0$ if $Z_{vij} \le 0$, then $P(Y_{vij} = 1 | \boldsymbol{\theta}_i, \boldsymbol{\beta}_v, \zeta_{vj}) = \Phi(\eta_{vij})$.

Prior distributions for the parameters were defined as follows. For the item parameters α_{vj} and γ_{vj}, the classical priors proposed in the literature were used. That is, we assume that $p(\alpha_{vj})$ has distribution $N(0,1)I(\alpha_{vj} > 0)$ and $p(\gamma_{vj})$ proportional to 1. Let us suppose that $\mathbf{\Theta}$ is distributed as $N_d(\mathbf{0},\mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is a correlation matrix, with ones on the diagonal and correlation σ_{st} between θ_s and θ_t, $s \neq t$. To model $\mathbf{\Sigma}$ we introduce an unconstraint covariance matrix $\mathbf{R} = [\rho_{st}]$ such that the covariance matrix $\mathbf{\Sigma}$ can be obtained from $\mathbf{R}$ using

$$\sigma_{st} = \frac{\rho_{st}}{\sqrt{\rho_{ss}\rho_{tt}}}, \; s \neq t. \quad (21)$$

A noninformative prior that can be assumed for $\mathbf{R}$ is the Jeffreys' prior, which is proportional to $|I(\theta)|^{\frac{1}{2}}$, where $I(\omega)$ is the expected Fisher information matrix of ω, (39). In this work we used the Jeffreys' prior, which in this case is proportional to $|\mathbf{R}|^{-(d+1)/2}$. Modeling the vectors $\boldsymbol{\beta}_v$ is new in the item response theory. Let $\boldsymbol{\beta}_v = (\beta_{v1},\cdots,\beta_{vd})^t$. Two prior distributions are proposed. First the improper non-informative $\prod_{k=1}^{d} I(\beta_{vk} \geq 0)$ and second, the informative truncate multivariate normal distribution $N_d(\mathbf{b},\mathbf{T})\prod_{k=1}^{d} I(\beta_{vk} \geq 0)$, where T is a diagonal matrix. We propose the hyperparameters $\mathbf{b} = \frac{1}{\sqrt{d}}\mathbf{1}_d$, where 1_d is the $d-$dimensional vector with ones in all its components, and $\mathbf{T} = diag(\frac{1}{d},\cdots,\frac{1}{d})$. In the simulations and in the real case the results were very similar with each one of the priors for the $\boldsymbol{\beta}_{\mathbf{v}}$. The joint posterior distribution of $(\boldsymbol{\theta},\boldsymbol{\beta},\boldsymbol{\zeta},Z,\mathbf{\Sigma})$ is proportional to

$$f(\mathbf{y}\,|\,\mathbf{Z})\,p(\mathbf{Z}\,|\,\boldsymbol{\theta},\boldsymbol{\beta},\boldsymbol{\zeta})\,p(\boldsymbol{\zeta})\,p(\boldsymbol{\beta})\,p(\boldsymbol{\theta}\,|\,\mathbf{\Sigma})\,p(\mathbf{R}). \quad (22)$$

The full conditional distributions are derived in (4). The code of the DAGS algorithm can be found in the home page of the SICS Research Group: *http://ciencias.bogota.unal.edu.co/departamentos/estadistica/investigacion/grupos-de-investigacion-departamento-estadistica/.*

5.4 Simulation

To test the DAGS algortihm a test of size $K = 100$ was generated. The 100 items were divided in four clusters(subtests), each one with 25 items. That is, $K_1 = K_2 = K_3 = K_4 = 25$. It was assumed that the latent trait space had dimension $d = 3$. A sample of 5000 people was generated. The algorithm recovered the item parameters and the latent traits very good. For details see (4).

VI. Application to a real case

The data are from the admission test at the Universidad Nacional de Colombia, applied in the second semester of 2009. The sample size was N=5096. The test was taken by more than 35,000 people. There were seven types of tests, but the only difference between them was the order of the questions. The data correspond to the complete sample of one type. The test size was $K = 113$ with 5 subtests. The subtests were: textual analysis (Textual) with $K_1 = 15$ items, mathematics (Math) with $K_2 = 26$ items, natural sciences (Science) with $K_3 = 29$ items, social sciences (Social) with $K_4 = 29$, and image analysis (Image) with $K_5 = 14$ items.

6.1 Missing data

In the test there were 1845 missing responses that correspond to 0.32% of the responses. The data were first fitted using the Bayesian imputation procedure. Then we used the usual procedure that is to replace the non-responses with 0. There were small differences in the estimations. Simulaton procedures showed that the parameters are better recovered when the imputation procedure is used than when the non-responses are replaced with 0. However, in this real case, there was an extreme case in which the examinee had 112 missing responses of the 113. From the Bayesian point of view, this is not a problem, and the DAGS algorithm worked well. However, we must be careful with the extreme cases, because in that cases the estimation of the latent traits of an examinee based solely on one, two, or very few responses is not consistent. Results completely different are obtained, depending if the only response is 1 or 0. Since our main goal is to illustrate the MSMIRT model, we finally decided to follow the usual procedure. So, we replaced the non-responses with 0.

6.2 Preliminary analysis of the data

To specify the MSMIRT model, the second parameterization is used. Then, it is necessary to state the dimension of the latent trait space, the number of clusters (subtests), and the main directions that will be aligned with the coordinate axes. In this case, and in similar situations, the clusters are predefined. If the number of clusters coincides with the dimension of the latent trait space, the model is of approximately simple structure, and the main directions are not necessary. Descriptive analysis suggests that the dimension of the latent trait space is 3. Additionally the results suggest to align the reference directions of the subtests Textual (axis 1), Math (axis 2), and Image (axis 3) with the coordinate axes. This configuration was adopted.

6.3 Fitting the data

Now, we review the parameters recovered by the DAGS algorithm. In the algorithm, the second parameterization of the model was used. In this case, we used a burning period of 5,000 iterations. After burning, we ran 10,000 iterations with a thin period of 1. That is, we obtained 10,000 iterations to compute the Bayesian estimations. To estimate the variance of the estimations, we used 100 batches of length 100. In all cases, the Bayesian estimator was the sample mean, because the mean and the median were very similar in all cases.

6.3.1 Estimation of the reference direction of the subtests

Table 1 contains the components of the subtest directions estimated by the DAGS algorithm. As mentioned before, the dimension of the latent trait space is 3. Let $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ be the ordered canonical base of $\mathbf{R}^3$. The reference directions of Math, Textual, and Image subtests were aligned with coordinate axes 1, 2 and 3 respectively, such that $\boldsymbol{\beta}_1 = \mathbf{e}_1$, $\boldsymbol{\beta}_2 = \mathbf{e}_2$, and $\boldsymbol{\beta}_5 = \mathbf{e}_3$. The reference directions $\boldsymbol{\beta}_3$ and $\boldsymbol{\beta}_4$ corresponding to the Science and Social subtests were estimated. The complete subtest directions are given in table 2. These results imply that Science is basically a composition of the Math and, Textual latent traits with a little component of Image and that Social is basically equivalent to Textual, with a little component of Image.

parameter	mean	sd	mcmc error	2.5%	0.50%	97.5%
$\widehat{\beta}_{31}$	0.70970	0.02716	0.00027	0.58928	0.71067	0.81753
$\widehat{\beta}_{41}$	0.98962	0.00275	0.00003	0.97460	0.99077	0.99800
$\widehat{\beta}_{32}$	0.69120	0.02870	0.00029	0.56307	0.69490	0.80182
$\widehat{\beta}_{42}$	0.07053	0.02104	0.00021	0.00329	0.06325	0.18015
$\widehat{\beta}_{33}$	0.09999	0.01379	0.00014	0.02550	0.09982	0.17695
$\widehat{\beta}_{43}$	0.10966	0.01505	0.00015	0.03974	0.10921	0.18203

Table 1: Estimated parameters of the subtest directions. Data from Admission Test of U.N.C., 2009

direction	component 1	component 2	component 3
β_1	1.000	0.000	0.000
β_2	0.000	1.000	0.000
$\widehat{\beta}_3$	0.710	0.691	0.100
$\widehat{\beta}_4$	0.990	0.071	0.110
β_5	0.000	0.000	1.000

Table 2: Estimated subtest directions. Data from Admission Test in U.N.C., 2009

6.3.2 Estimation of the covariance matrix

The covariance matrix that was estimated by the DAGS algorithm is given by

$$\widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} 1.0000 & 0.7273 & 0.5180 \\ 0.7273 & 1.0000 & 0.5544 \\ 0.5180 & 0.5544 & 1.0000 \end{pmatrix} \quad (23)$$

param.	mean	sd	mcmc err.	2.5%	0.50%	97.5%
$\widehat{\sigma}_{12}$	0.72726	0.00931	0.00009	0.68635	0.72836	0.76483
$\widehat{\sigma}_{13}$	0.51797	0.00909	0.00009	0.47631	0.51807	0.55885
$\widehat{\sigma}_{23}$	0.55439	0.00606	0.00006	0.51674	0.55479	0.58941

Table 3: Estimated parameters of the covariance matrix. Data from Admission Test of U.N.C., 2009

Table 3 shows the statistical information of the components of the covariance matrix estimated by the DAGS algorithm. From the covariance matrix estimated by the DAGS algorithm, we conclude that the main latent traits are highly correlated and therefore, the reference composite of the test is a good unidimensional synthesis of the latent trait vector.

The subtest directions in the uncorrelated space (parameterization 1) are obtained from $\widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\boldsymbol{\beta}}_{v}$ after normalizing these vectors. The subtest directions in the uncorrelated space are shown in table 4.

direction	component 1	component 2	component 3
$\widehat{\boldsymbol{\beta}}_1$	0.899	0.373	0.230
$\widehat{\boldsymbol{\beta}}_2$	0.373	0.892	0.256
$\widehat{\boldsymbol{\beta}}_3$	0.675	0.666	0.319
$\widehat{\boldsymbol{\beta}}_4$	0.852	0.417	0.316
$\widehat{\boldsymbol{\beta}}_5$	0.230	0.256	0.939

Table 4: Estimated subtest directions in the uncorrelated latent trait space. Data from Admission Testin U.N.C., 2009

Let $\mathbf{B}$ be the matrix whose rows are the subtest directions shown in table 4. The reference direction of the entire test in the uncorrelated space was computed as the first eigenvalue of $\mathbf{B}^t\mathbf{B}$. That direction was given by

$$\widehat{\boldsymbol{\beta}} = (0.693, 0.584, 0.423)^t. \quad (24)$$

6.3.3 Estimation of the item parameters

Table 5 shows the estimations of some of the slope parameters and table 6 shows the estimations of the corresponding intercept parameters. In the tables, the items have their original identifier. The items of each subtest were the following: Textual 1-15, Math 16-41, Science 42-70, Social 71-99 and, Image 100-113. The slope parameters were small in general, including the first parameterization. Apparently, some of the items could be omitted from the test. However, this issue will not be discussed.

item	mean	sd	mcmc error	2.5%	0.50%	97.5%
1	0.30793	0.00409	0.00004	0.26520	0.30769	0.35262
10	0.30494	0.00533	0.00005	0.25737	0.30447	0.35655
17	0.31501	0.00509	0.00005	0.27257	0.31512	0.35851
30	0.17121	0.00409	0.00004	0.12989	0.17133	0.21221
42	0.51616	0.00627	0.00006	0.47294	0.51591	0.56125
55	0.09379	0.00335	0.00003	0.06244	0.09392	0.12491
83	0.17787	0.00406	0.00004	0.14011	0.17754	0.21655
101	0.58574	0.00738	0.00007	0.53233	0.58575	0.63923
113	0.77583	0.01069	0.00011	0.71284	0.77555	0.83960

Table 5: Some estimated slope parameters. Data from Admission Test of U.N.C., 2009

item	mean	sd	mcmc error	2.5%	0.50%	97.5%
1	-0.27720	0.00294	0.00003	-0.31345	-0.27722	-0.24099
10	-0.84630	0.00448	0.00004	-0.88828	-0.84626	-0.80408
17	0.37025	0.00308	0.00003	0.33342	0.37012	0.40871
30	0.39290	0.00292	0.00003	0.35728	0.39274	0.42878
42	-0.62438	0.00675	0.00007	-0.67098	-0.62454	-0.57799
55	0.86565	0.00382	0.00004	0.82651	0.86551	0.90583
83	-0.21238	0.00289	0.00003	-0.24793	-0.21231	-0.17605
101	0.06880	0.00441	0.00004	0.02916	0.06893	0.10884
113	-0.18519	0.00607	0.00006	-0.22828	-0.18546	-0.14100

Table 6: Some estimated intercept parameters. Data from Admission Test of U.N.C., 2009

6.3.4 Goodness of fit

Some measures of goodness of fit of the model were computed inside the DAGS algorithm. For the complete details about goodness of fit, Bayesian latent residual and other Bayesian issues see (4). Let Z_{vij} be the underling latent continuous response of examinee i to the j th item of subtest v. This is the augmented variable used in the DAGS algorithm. For any fixed linear latent predictor η_{vij}, the latent variable Z_{vij} is given by

$$Z_{ij} = \eta_{vij} + \varepsilon_{vij}, \quad \varepsilon_{vij} \textit{ has distribution } N(0,1). \quad (25)$$

The Bayesian latent residual corresponding to the binary observation Y_{vij} is defined as

$$\varepsilon_{vij} = Z_{vij} - \eta_{vij}. \qquad (26)$$

These Bayesian latent residuals are the basis to define the statistics to asses the fitting of the model to the data, (16).

6.3.5 Outlier detection

According to (16), an observation is considered an outlier if the absolute value of the residual is greater than some pre-specified value q times the standard deviation. We used $q = 2$ and computed the posterior probability. The number of residuals with probability greater than 0.2 to be outliers was 2922 that correspond to (0.51%) of the total of observations. Additionally, the number of residuals such $|e_{vij}| > 1.5$ was 8069 that correspond to (1.4%) of the responses. Since this percent is less that 5% there is no reason to concern, (16).

6.3.6 Person Fit

A measure to evaluate the fit of a response pattern of a person i under the MSMIRT model based on the Bayesian latent residuals is given by

$$X^2_{p,i} = \sum_{v=1}^{m}\sum_{j=1}^{K_v}(Z_{vij} - \eta_{vij})^2 = \sum_{v=1}^{m}\sum_{j=1}^{K_v}\varepsilon^2_{vij}. \quad (27)$$

Each Bayesian latent residual has standard normal distribution, and under the assumption of conditional independence the statistic $X^2_{p,i}$ has a chi-square distribution with K degrees of freedom. That distribution can be used as a reference distribution to evaluate the extremeness of the sum of square residuals. The corresponding posterior p -value is defined as

$$p_0(X^2_{p,i}) = \int P(\chi^2_K > X^2_p(z_i))\, p(z_i \mid y_i) dz_i, \qquad (28)$$

where y_i represents the response pattern of a person i and z_i the corresponding latent response pattern. The posterior p -value is computed at each step of the DAGS algorithm, and the mean is the estimate of the posterior p -value. The p-values in real case data were between 0.13 and 0.80.

6.3.7 Item Fit

Similarly, an item fit statistic is defined as

$$X^2_{item,j} = \sum_{i=1}^{N}(Z_{vij} - \eta_{vij})^2 = \sum_{i=1}^{N}\varepsilon^2_{vij}, \qquad (29)$$

and the corresponding posterior p -value is defined as

$$p_0(X^2_{item,vj}) = \int P(\chi^2_N > X^2_{item}(z_j))\, p(z_j \mid y_j) dz_j, \qquad (30)$$

The p-values in real case data were between 0.48 and 0.52.

VII. Discussion and conclusions

In this paper we introduced the Multiple Subtests MIRT (MSMIRT) model. The model has been thought to be used in large-scale assessment tests designed explicitly to measure more than one latent trait. It was assumed that the tests are split into subtests and that each subtest is designed to measure mainly a unique unidimensional latent trait.

A discussion about the concept of dimension in the item response theory was the central issue in the paper. (3) points out that the dimension of the latent trait space and the dimension of a test are different. According to Reckase, the dimension of the latent trait space is an underlying property of the examinees, while the dimension of the test is a design property of the test. Obviously, the latent trait space depends on the design of the test. When a test is designed to measure some specific latent traits, the examinees require certain abilities to answer the test successfully. However, the dimension of the ability space does not coincide necessarily with the number of latent traits that the test attempts to measure.

The MSMIRT model is a multidimensional item response theory model in which the items have a cluster structure. The model is based on the assumption that the dimension of the latent trait space is smaller than the number of subtests (clusters) of the test. The MSMIRT model is equivalent to a factor analysis model of dichotomized variables,in which the factors are just the latent traits. The dimension of the test was defined as the number of clusters of the test, and the dimension of the latent trait space was defined as the number of factors of that model. Consequently, the dimension of the test is a design property while the dimension of the latent trait space is a characteristic of the response data.

In the MSMIRT model, there are two types of latent traits that are considered: the main latent traits and the basic latent traits. The main latent traits correspond to those abilities that the test attempts to measure. Thus, the main latent traits are defined by the design of the test, and can be interpreted directly form the underlying theory that leads the test design. On the other hand, the basic abilities are the components of the latent trait vector of the examinees. In general, these latent traits are not interpretable directly. In the MSMIRT model, the main latent traits are linear combinations of the basic latent traits.

We defined the concept of reference direction of a subtest as the direction along which the subtest discriminates better on average. The reference direction of the subtests are estimated directly in the MSMIRT model. This is an important characteristic of the MSMIRT model, because the main latent traits are just the reference composites of the subtests. This implies that basic and the main latent traits are estimated directly. Furthermore, the covariance matrix of the basic latent traits is also estimated.

Two equivalent parameterizations were proposed for the model. In the first, it is assumed that the basic latent traits are uncorrelated, so any linear combination of them has the same scale. In this parameterization, the basic latent traits do not have a direct interpretation, and all the main latent traits that are measured by the test are linear combinations of the basic latent traits.

To estimate the parameters of the model, the second parameterization of the model was adopted. Following this parameterization, some of the main latent traits that the test attempts to measure are identified with the coordinate axes of the latent trait space. The other latent traits can be described as combinations of the basic latent traits. This interpretation may be useful to the experts.

A data augmentation Gibbs sampler (DAGS) algorithm was implemented to fit the MSMIRT model. The simulation results showed that the parameters are recovered well by the DAGS algorithm.

To illustrate the use of the MSMIRT model, we utilized the response data of a test from Universidad Nacional de Colombia. The test had 5 subtest to measure respectively Textual, Math, Science, Social, and Image. Each subtest was unidimensional. However, all the items in the test are correlated and the dimension analysis reveals that the data have dimension 3. One can consider basically two types of classical models to fit the data. The first option is to consider a MIRT model. In this case, the dimension of the latent space is 3. However, in this case the cluster structure of the items and the fact that each cluster measures an unidimensional latent trait is omitted.

The second option is to adopt a simple structure model. In this case, it is assumed that the dimension of the latent trait space is 5. In both cases, the models are over parameterized. The MSMIRT model seemed to be a better option to fit the data. The statistical analysis of goodness of fit showed that the MSMIRT model fitted the data well.

After fitting the data, we can explore an interesting characteristic of the MSMIRT model, which can be useful for the experts. When the first parameterization is used, some of the main latent traits can be identified with the basic latent traits. Consequently, the other main latent traits can be interpreted in terms of that latent traits. For example, in the current case, the main latent trait Social is a composite of Textual (85%), Math (6%), and Image (9%).

The MSMIRT model introduced in this paper seems to be more natural to fit data from tests designed to measures several specific latent traits, in which a cluster structure is available. The MSMIRT model is in general more parsimonious than the existing models. In the simulations, the responses were generated using MIRT models. However, the cluster structure of the tests and the fact that each subtest measures essentially a main latent trait were incorporated in the MIRT models. The results of simulations showed that the MSMIRT model fitted well the data in this case, so the classical MIRT model can be replaced by a MSMIRT model in these situations.

Acknowledgements

This work was supported by Colciencias, grant contract number 0039-2013, and Universidad Nacional de Colombia. This is a work of the SICS Research Group of Universidad Nacional de Colombia.

References

[1]. M. Reckase. The difficulty of test items that measure more than one ability. *Applied Psycological Measurement*, 9(9):401–412, 1985.
[2]. M. Reckase. Multidimensional item response theory. *Handbook of Statistics*, 26:607–642, 2007.
[3]. M. Reckase. *Multidimensional Item Response Theory*. Statistics for Social and Behavior Sciences. Springer, 2009.
[4]. A. Montenegro. *Multidimensional Item Response Theory Models where the Ability has a Linear Latent Structure*. Phd dissertation, Universidad Nacional de Colombia, 2010.
[5]. Y. Sheng. Comparing multiunidimensional and unidimensional item response theory models. *Educational and Psychological Measurement*, 67(6):899–919, 2007.
[6]. Y. Sheng. *Bayesian IRT Models with Genaral and specific Traits*. VDM Verlag Dr. Müller, 2008.
[7]. Y. Sheng. A matlab package fo markov chain monte carlo with a multi-unidimensional irt model. *Journal of Statistical Software*, 28(10):1–20, 2008.
[8]. R. D. Bock and M. Lieberman. Fitting a response model for n dichotomously scored items. *Psychometrika*, 35:179–197, 1970.
[9]. A. Christofferson. Factor analysis of dichotomized variables. *Psychometrika*, 40:5–32, 1975.
[10]. Y. Takane and J. Leeuw. On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, 52(3):393–408, 1987.
[11]. R. P. McDonald. The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, 34:100–117, 1981.
[12]. R. P. McDonald. A basis for multidimensional item response theory. *Applied Psychological Measurement*, 24:99–114, 2000.
[13]. J. Albert. Bayesian estimation of normal item response curves using gibbs sampling. *Journal of Educational Statistics*, 17(3):251–269, 1992.
[14]. A. Béguin and C. A. Glass. Mcmc estimation and some model-fit analysis of multidimensional irt models. *Psychometrika*, 66(4):541–562, 2001.
[15]. H. Lee. *Markov Chain Monte Carlo Methods for Estimating Multidimensional Ability in Item Response Analysis*. Ph.d. thesis, University of Missouri, Columbia, MO, 1995.
[16]. J. P. Fox. *Bayesian Item Response Modeling*. Springer, 2010.
[17]. F. Drasgow and CK Parsons. Applications of unidimensional item response theory to multidimensional data. *Applied Psychological Measurement*, 7:189–199, 1983.
[18]. John Fox. *polycor: Polychoric and Polyserial Correlations*, 2010. R package version 0.7-8.
[19]. R. D Bock, R. Gibbons, and E. Muraki. Full-information item factor analysis. *Applied Psychological Measurement*, 12:261–280, 1988.
[20]. T. Ackerman. Unidimensional irt calibration of compensatory and noncompensatory multidimensional items. *Applied Psychological Measurement*, 13:113–127, 1989.
[21]. T. Ackerman. A didactic explanation of items bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement*, 29(1):67–91, 1992.
[22]. M. Reckase and T. Ackerman. Building a unidimensional test using multidimensional items. *Journal of Educational Measurement*, 25(3):193–203, 1988.
[23]. M. Reckase, J. Carlson, and T. Ackerman. The interpretation of the unidimensional irt parameters when estimate from multidimensional data. In *Paper presented at the annual meeting of Psychometrics Society, Toronto*, 1986.
[24]. T. Ansley and R. Forsyth. An examination of the charateristics of unidimensional irt parameter estimates derived from two dimensional data. *Applied Psychological Measurement*, 9:27–48, 1985.
[25]. W. Way, T. Ansley, and R. Forsyth. The comparative effects of compensatory and noncompensatory two-dimensional data items on unidimensional irt estimates. *Applied Psychological Measurement*, 12:239–252, 1988.
[26]. R. Nandakumar. Traditional dimensionality versus essential dimensionality. *Journal of Educational Measurement*, 28:99–117, 1991.
[27]. W. Stout. A nonparametric approach for assesing latent trait dimensionality. *Psychometrika*, 52:589–617, 1987.
[28]. ME. Gesaroli and AF. De Champlain. Using an approximate chi-square statistic to test the number of dimensions underlying the responses to a set of items. *Journal of Education Measurement*, 33:157–179, 1996.
[29]. W. Stout, B. Douglas, B. Junker, and L. Roussos. Dimtest. Computer software, The William Stout Institute for Measurement, Champaing, IL, 1999.

[30]. W. Stout. A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, 55:293–325, 1990.

[31]. L. Humphreys. A theoretical and empirical study of psychometric assessment of psychological test dimensionality and bias. Onr research proposal, Washington, DC: Office of Naval Research., 1984.

[32]. JC. Gower and GB. Dijksterhuis. *Procustes Problems*. Oxford University Press, Oxford, England, 2004.

[33]. M. Levine. Dimension in latent variable models. *Journal of Mathematical Psychology*, 47:450–466, 2003.

[34]. B. Carroll, J. Williams and M. Levine. Multidimensional modeling with unidimensional approximations. *Journal of Mathematical Psychology*, 51:207–228, 2007.

[35]. A. Montenegro and E. Cepeda. Synthesizing the ability in multidimensional item response theory models. *Revista Colombiana de Estadística*, 33:127–147, 2010.

[36]. J. Bazán. *Uma Familia de Modelos de Resposta ao Ítem Normal Assimétrica*. Phd dissertation, Universidade de São Pablo, 2005.

[37]. J. Bazán. A skew item response model. *Bayesian Analysis*, 1(4):861–892, 2006.

[38]. G. Da Silva. *Modelos multidimensionais da TRI com distribuições assimétricas para os traços latentes*. Phd dissertation, Universidade de São Pablo, 2008.

[39]. D. Gamerman and H. Lopes. *Markov Chain Monte Carlo*. Chapman and Hall/CRC, 2th edition, 2006.