



HAL
open science

Modelling Missingness with a Rasch-Type model

Lucio Bertoli-Barsotti, Antonio Punzo

► **To cite this version:**

Lucio Bertoli-Barsotti, Antonio Punzo. Modelling Missingness with a Rasch-Type model. *Annales de l'ISUP*, 2013, 57 (1-2), pp.29-44. <hal-03615282>

HAL Id: hal-03615282

<https://hal.science/hal-03615282v1>

Submitted on 21 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Modelling Missingness with a Rasch-Type model

Lucio Bertoli-Barsotti, L.* and Antonio Punzo

Università di Bergamo et Università di Catania

Abstract. In this paper we focus on a model-based approach to the treatment of missing data due to examinees' nonresponse, in the context of Item Response Theory (IRT). With model-based approach we mean that item nonresponses are to be included in the analysis – indeed we assume that nonresponses are caused by a specific latent trait, summarizing the response propensity of the examinee. Then, the idea is to postulate the existence of two latent traits: one for response propensity and the other for ability/proficiency. Different models have been proposed in the literature. In this paper, a new class of multidimensional IRT models, called Rasch-Rasch models, is introduced. The Rasch-Rasch model belongs to the wider class of Rasch models – and, as a member of the exponential family, it can be viewed as a generalized linear mixed model. Real and artificial datasets are used to illustrate the characteristics of this new model.

Keywords: Multidimensional Rasch Model; nonignorable missing data; nonresponse; exponential family; marginal maximum likelihood estimation.

Introduction

A very common problem in applications of Item Response Theory (IRT) is the presence of nonresponses. Omitted responses are sometimes treated, by practitioners, as not administered – e.g. when the amount of missing data is small – but this method is adequate only if data are *missing at random* (MAR). Missing data may be structural (planned missingness) when it is introduced deliberately from the researcher – e.g. in targeted testing. Sometimes missing data may be detected post hoc – this is the case, e.g., when missingness is due to subject choice. This paper focuses on this latter situation, i.e. when nonresponses originate from respondent's choice – typically because the item has been skipped. Recently a model-based approach to dealing with nonresponses has been proposed by several authors ([1], [2], [3], [4], [5], [6], [7], [8], [9]; see also [10], [11], [12], [13]). The main idea is to incorporate the nonresponse mechanism into the analysis, under the assumption of a latent response propensity, say θ_1 ; the measurement model for θ_1 is added to the measurement model for the latent proficiency variable, say θ_2 . In the model presented in this paper we assume for simplicity a unidimensional proficiency parameter. This results in a multidimensional IRT (MIRT) model, where every person is characterized by a 2-dimensional latent trait $\theta^T = (\theta_1, \theta_2)$. The first dimension may be considered, in a loose sense, related to the general person's tendency to respond (response propensity), while the second dimension may represent the usual person's attitude, or "ability". The aim of this contribution is to develop, following the latent variable approach to taking into account of the person's tendency to respond, a MIRT model which belongs to the Rasch family of statistical models.

In this study we shall refer to the case of dichotomously scored data. Let consider the following response "codes" – as characters that are to be regarded as valid data for the dataset:

- A, representing a nonresponse (missing value code);
- B, if the subject gives an incorrect response to the item;
- C, if the subject gives a correct response to the item.

The probability of giving a response will depend on person's position on both the first and the second dimension of the latent trait, respectively θ_1 and θ_2 . To combine these two dimensions in the analysis, it is convenient to assume a random-effects approach, i.e. to

assume a distribution for θ , with corresponding density function, say, $g(\theta; \omega)$, where ω indicates a vector of population parameters. Holman and Glas ([2]; hereafter we shall refer to these authors as HG) presented a wide class of models incorporating the missing-data mechanism as an effect explained by a specific person's latent trait. We first introduce some notation to present briefly their taxonomy. Let D_{vi} be the response indicator random variable, with values $d_{vi} = 1$ if the response of person v on item i was observed, and $d_{vi} = 0$ otherwise (i.e. in the case of nonresponse), and let X be the matrix with elements x_{vi} , $i = 1, \dots, n$, $v = 1, \dots, k$, where x_{vi} represents the score of person v on item i , with values 0 or 1 corresponding to either incorrect or correct answer (or other binary item responses, e.g. false-true, no-yes, disagree-agree) when the response has been observed and equal to some arbitrary constant otherwise. Finally, let $p(x_{vi} | d_{vi})$ denote the probability function for the score x_{vi} , and let $p(d_{vi})$ the probability function for the response indicator random variable. The probability function of the joint random variable (X_{vi}, D_{vi}) may be factorized as $p(x_{vi}) = p(x_{vi} | d_{vi}) p(d_{vi})$. Note that both $X_{vi} | D_{vi}$ and D_{vi} are Bernoulli variables that depend on a set of person and item parameters to be specified. In their taxonomy, HG distinguish the following cases, which they called respectively models G_2 , G_3 and G_4 :

- a) $p(x_{vi} | d_{vi})$ depends only on θ_2 and $p(d_{vi})$ depends only on θ_1 ,
- b) $p(x_{vi} | d_{vi})$ depends only on θ_2 , while $p(d_{vi})$ depends on both θ_1 and θ_2 ,
- c) $p(x_{vi} | d_{vi})$ depends on both θ_1 and θ_2 , while $p(d_{vi})$ depends only on θ_1 .

In passing, note that model G_2 is an example of pattern-mixture model (while the others are not). If the correlation between θ_1 and θ_2 is constrained to be zero, then interesting special cases are obtained. In particular, under the further assumption of stochastic independence between θ_1 and θ_2 , the model G_2 is called G_1 by HG. In other terms, if $g(\theta; \omega)$ is the density of a 2-dimensional normal, model G_1 can be viewed as a constrained version of model G_2 in which the correlation between θ_1 and θ_2 is set to zero. By construction, when model G_1 holds, the missing-data mechanism is ignorable for likelihood inference; indeed, the missing data are MAR and the distinctness condition on the parameters θ_1 and θ_2 is satisfied because θ_1 and θ_2 have independent distributions ([14], pp. 119-120). HG interpreted this fact by considering the magnitude of the correlation between θ_1 and θ_2 - in model G_2 - as an indication of the extent to which ignorability is violated. It is important to realize that in all these models item and person parameters generally have quite different meanings; in particular, the correlation between the components of the latent variable θ will differ across these models.

Knott and Tzamourani ([6]) introduced a formulation similar to that of HG for their model G_3 with $\text{logit } P(D=1) = \gamma_1 \theta_1 + \gamma_2 \theta_2 - \delta_1$ and $\text{logit } P(X=1 | D=1) = \gamma_3 \theta_2 - \delta_2$, where the γ s are discrimination parameters, and δ_1 and δ_2 are item parameters; moreover they also assume that θ_1 and θ_2 have independent standard normal distributions.

In its most simple form, by modelling $p(x_{vi} | d_{vi})$ and $p(d_{vi})$ with two simple Rasch models, model G_2 yields: $P_A \propto 1 + \exp(\theta_2 - \delta_2)$, $P_B \propto \exp(\theta_1 - \delta_1)$, $P_C \propto \exp(\theta_1 + \theta_2 - \delta_1 - \delta_2)$. Then G_2 is a *continuation-ratio logits* model:

$$(i) \log((P_B + P_C)/P_A) = \theta_1 - \delta_1, \quad (ii) \log(P_C/P_B) = \theta_2 - \delta_2. \quad (1)$$

A possible limit of these models, $G_1 - G_4$, is that they not belong to the Rasch family of statistical models (and then to the exponential family of distributions).

Modelling missingness with a Rasch-type model

As an alternative to the HG's approach, we propose a new class of models belonging to the exponential family of distributions. The main idea is to adopt, instead of the *continuation-ratio*, the following *adjacent-categories logits* approach:

$$(i) \log(P_B/P_A) = \mathbf{q}_1^T \boldsymbol{\theta} - \delta_1, \quad (ii) \log(P_C/P_B) = \mathbf{q}_2^T \boldsymbol{\theta} - \delta_2, \quad (2)$$

where $\mathbf{q}_1^T = (q_{11}, q_{12})$ and $\mathbf{q}_2^T = (q_{21}, q_{22})$ are linearly independent vectors of fixed constants, and $\boldsymbol{\theta}^T = (\theta_1, \theta_2)$. Since both the steps A-B and B-C are modeled with Rasch-type model, we shall call this special class of model *Rasch-Rasch Model* (RRM). Note that both the models (2)(i) and (2)(ii) are special cases of Reckase's ([15], [16]) Linear Logistic Multidimensional Model for dichotomous responses, - where the discrimination parameters are considered as constants to be fixed in advance. More properly, these two models may also be viewed as instances of the dichotomous MULTIRA model ([17]) (however this is not the case for the *whole* RRM). The most important difference between model (1) and model (2) is that the latter belongs to the exponential family of distributions. Then, in particular model (2) admits nontrivial sufficient statistics.

To better handle this model, a simple numerical recoding is needed. By means of the introduction of a selection vector $\mathbf{y}_{vi} = (y_{vi0}, y_{vi1}, y_{vi2})$ -with values (1,0,0), (0,1,0) and (0,0,1), respectively for A, B and C- the RRM can also be viewed, algebraically, as an instance of some well-known Rasch general forms of family models. Indeed, by defining the "score" of person v on item i as $S_{vi} = \sum_h h Y_{vih}$, the RRM can also be viewed as a *Generalized Rasch*

Model ([18]): in particular, the RRM is a *Multidimensional Random Coefficients Multinomial Logit Model* (MRCMLM, [19]) - the most general model structure of multidimensional Rasch models. For the MRCMLM, the probability of a response in category h of item i for person v is given by

$$P(Y_{vih} = 1 | \boldsymbol{\theta}_v, \boldsymbol{\xi}) = P(S_{vi} = h | \boldsymbol{\theta}_v, \boldsymbol{\xi}) = \frac{\exp(\mathbf{b}_{ih}^T \boldsymbol{\theta}_v + \mathbf{a}_{ih}^T \boldsymbol{\xi})}{\sum_{m_j} \exp(\mathbf{b}_{ij}^T \boldsymbol{\theta}_v + \mathbf{a}_{ij}^T \boldsymbol{\xi})} \quad (3)$$

where:

D is the number of hypothesized dimensions for the latent trait vector;

$\boldsymbol{\theta}_v^T = (\theta_{v1}, \theta_{v2}, \dots, \theta_{vD})$ is the latent trait vector for person v ;

$\boldsymbol{\xi}$ is the vector of item parameters;

$\mathbf{b}_{ih}^T = (b_{ih1}, b_{ih2}, \dots, b_{ihD})$ is a scoring weights vector of known constants,

b_{ihd} represents the score given to category h on dimension d of item i (by definition, the score for a response in the 0 category is 0 for both dimensions);

\mathbf{a}_{ih}^T is a design vector of known constants given to category h of item i ;
 $m_i + 1$ is the number of response categories in item i .

In a more special way -due to the dependence of logit expressions in the equations (2) on both the parameters θ_1 and θ_2 - the RRM can be seen as an instance of a MRCMLM, with $D=2$ and $m_i + 1 = 3$, with a *within-item* multidimensionality (see Figure 1), since different response "levels" (A, B and C) may entail different *linear combinations* of the latent traits θ_1 and θ_2 (and not simply different latent traits) where -by construction- $\mathbf{b}_{i1}^T = \mathbf{q}_1^T$, $\mathbf{b}_{i2}^T = \mathbf{q}_1^T + \mathbf{q}_2^T$.

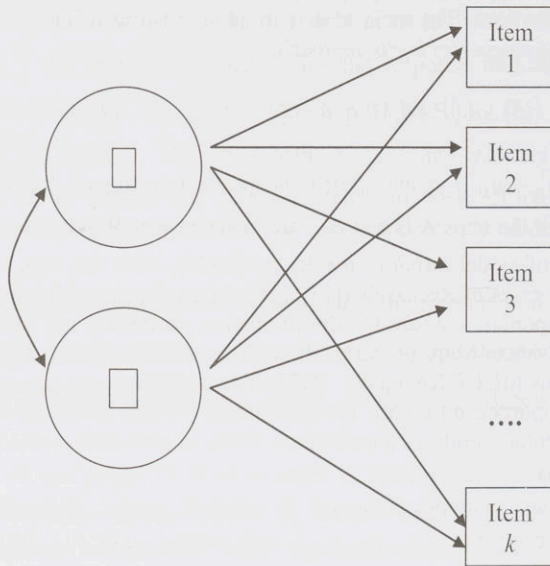


Figure 1. Conceptual path diagram of the within-item multidimensionality of a RRM
 Indeed, from equations (2) one can derive, equivalently, the probabilities

$$P_A = \left[1 + \exp\{\mathbf{q}_1^T \boldsymbol{\theta} - \delta_1\} + \exp\{(\mathbf{q}_1^T + \mathbf{q}_2^T) \boldsymbol{\theta} - (\delta_1 + \delta_2)\} \right]^{-1},$$

$$P_B = \exp\{\mathbf{q}_1^T \boldsymbol{\theta} - \delta_1\} \cdot P_A,$$

$$P_C = \exp\{(\mathbf{q}_1^T + \mathbf{q}_2^T) \boldsymbol{\theta} - (\delta_1 + \delta_2)\} \cdot P_A.$$

There are many different ways to decompose the person parameters and to define the multidimensional structure of $\boldsymbol{\theta}$. The choice of a specific model within the RRM class should be guided by the goal of maximizing the interpretability of the parameters. It is important to note that, as a consequence of the expression of P_A , the probability of a missing value is allowed structurally to depend on *both* the traits θ_1 and θ_2 . Now, since the specific interpretation of the latent variables θ_1 and θ_2 in RRM depends upon the choice of the scoring vectors \mathbf{q}_1 and \mathbf{q}_2 , in order to properly interpret θ_2 as the sole dimension of

“ability”, one should take $q_{21} = 0$. With this constrained parameterization, an RRM may be considered as a Rasch-type counterpart of HG’s model G_3 .

Among the many possible choices, we will consider the two following ones:

- 1) $\mathbf{q}_1^T = (1, 0)$, $\mathbf{q}_2^T = (0, 1)$ (that corresponds to: $\mathbf{b}_{i1}^T = (1, 0)$ and $\mathbf{b}_{i2}^T = (1, 1)$);
- 2) $\mathbf{q}_1^T = (1, -1)$, $\mathbf{q}_2^T = (0, 2)$ (that corresponds to: $\mathbf{b}_{i1}^T = (1, -1)$ and $\mathbf{b}_{i2}^T = (1, 1)$).

Let us call these RRMs B_1 and B_2 , respectively. For completeness, we shall also mention a third interesting version, say B_3 , of the RRM, that is obtained by posing $\mathbf{q}_1^T = (1, 0)$ and $\mathbf{q}_2^T = (-1, 1)$ (hence $\mathbf{b}_{i1}^T = (1, 0)$ and $\mathbf{b}_{i2}^T = (0, 1)$). Model B_3 is known as the Multicategorical Multidimensional Rasch Model ([20], [21]). For this model, each response *category* ($s_{vi} = 1, 2$) corresponds to (exactly) one of the latent dimensions. However, this model does not meet the above mentioned requirement of $q_{21} = 0$, then it will not be considered further in the analysis.

Note that model B_1 -that arises from a simple Rasch Model for both the logit expressions (i) and (ii) in equation (2)- can also be viewed as a member of the family of *Multidimensional Partial Credit Models* ([22]). For this model, each *step* (i.e. both the conditional probabilities $P(S_{vi} = 1 | S_{vi} = 0 \text{ or } S_{vi} = 1)$ and $P(S_{vi} = 2 | S_{vi} = 1 \text{ or } S_{vi} = 2)$) depends on a different latent trait.

Finally, let denote by R_0 the simple Rasch Model by treating all omitted responses as not administered (then, ignoring the nonresponse). The probability functions for models B_1 and B_2 are summarized in Table 1, as well as those of model R_0 , G_2 and G_3 . For each of the models considered in Table 1, the distribution of $p(x_{vi} | D_{vi} = 1, \theta_{v2}, \delta_{i2})$ is (equivalent to) a simple Rasch Model.

The advantage of these models over others (for example, model B_3) is the interpretability of θ_2 as the usual latent “ability” score. Note that model G_2 differs from the others (without considering the special situation of model R_0) in that the probability function of the response indicator variable $p(D_{vi} | \theta_{v1}, \delta_{i1})$ (see equation (1)(i)) and the probability function of the outcome variable $p(x_{vi} | D_{vi} = d_{vi}, \theta_{v2}, \delta_{i2})$, are parameterized by distinct set of parameters.

Since the meaning of the parameters in the different model is not the same, in general, we will emphasize this by writing θ as θ^* , where $*$ = B_1, B_2, R_0, G_2 and G_3 , and analogously for the other parameters of the model.

Table 1. Alternative models for dichotomous data in presence of nonresponses.

B_1 and B_2 are RRM; G_2 and G_3 are simple instances of HG-type models; R_0 represents the simple Rasch Model under the assumption of ignorable nonresponses

Family of models		Resp. codes	probability
RM	R_0	A	<i>not present (ignorable missing)</i>
		B	$\left[1 + \exp(\theta_2^{R_0} - \delta_2^{R_0})\right]^{-1} (= c_0^{-1})$
		C	$c_0^{-1} \exp(\theta_2^{R_0} - \delta_2^{R_0})$
RRMs	B_1	A	$\left[1 + \exp(\theta_1^{B_1} - \delta_1^{B_1}) + \exp(\theta_1^{B_1} + \theta_2^{B_1} - \delta_1^{B_1} - \delta_2^{B_1})\right]^{-1} (= c_1^{-1})$
		B	$c_1^{-1} \exp(\theta_1^{B_1} - \delta_1^{B_1})$
		C	$c_1^{-1} \exp(\theta_1^{B_1} + \theta_2^{B_1} - \delta_1^{B_1} - \delta_2^{B_1})$
	B_2	A	$\left[1 + \exp(\theta_1^{B_2} - \theta_2^{B_2} - \delta_1^{B_2}) + \exp(\theta_1^{B_2} + \theta_2^{B_2} - \delta_1^{B_2} - \delta_2^{B_2})\right]^{-1} (= c_2^{-1})$
		B	$c_2^{-1} \exp(\theta_1^{B_2} - \theta_2^{B_2} - \delta_1^{B_2})$
		C	$c_2^{-1} \exp(\theta_1^{B_2} + \theta_2^{B_2} - \delta_1^{B_2} - \delta_2^{B_2})$
H&G models	G_2	A	$\left[1 + \exp(\theta_1^{G_2} - \delta_1^{G_2})\right]^{-1}$
		B	$\exp(\theta_1^{G_2} - \delta_1^{G_2}) \left[(1 + \exp(\theta_1^{G_2} - \delta_1^{G_2})) (1 + \exp(\theta_2^{G_2} - \delta_2^{G_2})) \right]^{-1}$
		C	$\exp(\theta_1^{G_2} + \theta_2^{G_2} - \delta_1^{G_2} - \delta_2^{G_2}) \left[(1 + \exp(\theta_1^{G_2} - \delta_1^{G_2})) (1 + \exp(\theta_2^{G_2} - \delta_2^{G_2})) \right]^{-1}$
	G_3	A	$\left[1 + \exp(\theta_1^{G_3} + \theta_2^{G_3} - \delta_1^{G_3})\right]^{-1}$
		B	$\exp(\theta_1^{G_3} + \theta_2^{G_3} - \delta_1^{G_3}) \left[(1 + \exp(\theta_1^{G_3} + \theta_2^{G_3} - \delta_1^{G_3})) (1 + \exp(\theta_2^{G_3} - \delta_2^{G_3})) \right]^{-1}$
		C	$\exp(\theta_1^{G_3} + 2\theta_2^{G_3} - \delta_1^{G_3} - \delta_2^{G_3}) \left[(1 + \exp(\theta_1^{G_3} + \theta_2^{G_3} - \delta_1^{G_3})) (1 + \exp(\theta_2^{G_3} - \delta_2^{G_3})) \right]^{-1}$

Interpretation of the parameters and derivation of the sufficient statistics

In its exponential form, the probability of a given response pattern $s_v = (s_{v1}, s_{v2}, \dots, s_{vk})$ depends upon the linear combinations

- 1) $\theta_1^{B_1} y_{v,1} + (\theta_1^{B_1} + \theta_2^{B_1}) y_{v,2}$ for model B_1 , and
- 2) $(\theta_1^{B_2} - \theta_2^{B_2}) y_{v,1} + (\theta_1^{B_2} + \theta_2^{B_2}) y_{v,2}$ for model B_2 ,

where $y_{v,h} = \sum_i y_{vih}$. Then, the jointly sufficient statistics for the person parameter are $(y_{v,1} + y_{v,2}, y_{v,2})$ and $(y_{v,1} + y_{v,2}, y_{v,2} - y_{v,1})$, respectively, for model B_1 and model B_2 . Analogously, it is easy to see that $(y_{i,1} + y_{i,2}, y_{i,2})$ are jointly sufficient for the item parameter

$(\delta_{i1}, \delta_{i2})$, where $y_{vih} = \sum_v y_{vih}$. These statistics are in fact equivalent, but the person parameters have slightly different meaning for model B_1 and model B_2 . More specifically, for model B_1 the greater the value of the person parameter $\theta_1^{B_1}$, the greater is the tendency of the person of choosing option B rather than option A, while for model B_2 the greater is the value of the person parameter $\theta_1^{B_2}$ the greater is the tendency of the person of choosing options B or C rather than option A. In this sense the parameter $\theta_1^{B_2}$ is more clearly related to the response propensity. It is worth noting that both the models yields the same estimates of the parameters, up to a (linear) reparameterization.

Estimation

As for all the other Rasch-type models, the parameters of the RRM can be estimated by means of unconditional, conditional and marginal methods. Despite statistical optimality of Conditional Maximum Likelihood (CML) estimators, the Marginal Maximum Likelihood (MML) tends to be in wider use nowadays, and in this study we shall consider this latter, also because the MML approach provides estimates of the correlation of the latent traits – and under our framework the latent correlation plays a crucial role for the interpretation of ignorability.

MML requires to postulate a multivariate density function $g(\boldsymbol{\theta}_v; \boldsymbol{\omega})$ for person's v latent trait $\boldsymbol{\theta}_v$, where $\boldsymbol{\omega}$ indicates the vector parameter characterizing the distribution. In MML estimation, the vector of item parameters $\boldsymbol{\delta}$ is simultaneously estimated with the parameter $\boldsymbol{\omega}$ of the latent trait distribution by maximizing the marginal likelihood function

$$L_M = \prod_v \int_{\boldsymbol{\theta}} p(s_v, \boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\omega}) d\boldsymbol{\theta},$$

where $p(s_v, \boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\omega}) = \prod_i P(S_{vi} = s_{vi} | \boldsymbol{\theta}, \boldsymbol{\delta}) g(\boldsymbol{\theta}, \boldsymbol{\omega})$ represents the simultaneous distribution of the response pattern s_v and the ability $\boldsymbol{\theta}$ for any person v in the sample. Now, if g is constrained to be a normal distribution with mean set equal to zero to identify the model, then $\boldsymbol{\omega} = (\boldsymbol{\Sigma})$, where the covariance matrix $\boldsymbol{\Sigma} = (\sigma_{ij})$ is an estimand. The software ConQuest ([18]) provides MML estimates of the parameters $\boldsymbol{\delta}$ and $\boldsymbol{\omega}$ of the Generalized Rasch model stated by equation (3) – by adopting Bock and Aitkin's ([23]) formulation of the Expectation-Maximization algorithm ([24])– under normality assumption for the density g . The estimation procedure has been described in [19]. The software ConQuest was used for the estimation of RRM parameters in this paper. For a test containing k dichotomous items, the software produce estimates for a total of $2k + 3$ parameters: $2k$ item parameters and 3 elements for the covariance matrix. It should be noted that B_1 , B_2 (and B_3) produce identical estimates of item parameters: then $\delta_j^{B_1} = \delta_j^{B_2} (= \delta_j^{B_3})$. This may be easily proved by considering the conditional approach to the estimation. Indeed, since the sufficient statistics $(y_{v+1} + y_{v+2}, y_{v+2})$ and $(y_{v+1} + y_{v+2}, y_{v+2} - y_{v+1})$ (and also (y_{v+1}, y_{v+2}) , for the case of model B_3) are equivalent, then they define the same conditional likelihood function. Besides, the estimates of the elements of the covariance matrix $\boldsymbol{\Sigma}$ will differ across these models, but the person parameter estimates of one model can be converted, ex-post, into the other model's parameterization by means of a linear transformation. In particular, one obtains

$$COV(\theta_1^{B_1}, \theta_2^{B_1}) = COV(\theta_1^{B_2} - \theta_2^{B_2}, 2\theta_2^{B_2}) = 2COV(\theta_1^{B_2}, \theta_2^{B_2}) - 2VAR(\theta_2^{B_2}).$$

The coefficient of correlation can be used to measure the extent to which the latent dimensions are related. As a limit case, when these load on the same trait the correlation equals one, in its absolute value; in particular, if $\theta_1^{B_1} = \theta_2^{B_1}$, B_1 leads to a Partial Credit Model (PCM): let us denote this model as B_0 . Hence, since the B_0 is hierarchically related to B_1 , the change in deviance value (the difference in deviance may be compared to a chi-squared distribution with 2 degrees of freedom) may be used to compare statistically the fit of these two models – as a statistical test of multidimensionality.

Statistical simulations

Two simulation studies are carried out to assess parameter recovery properties of this new model – when the computer program ConQuest is used to calibrate its parameters. The R program was utilized to generate the data. The accuracy of the parameter estimates is assessed with the criterion - commonly used in parameter recovery studies - of the evaluation of the average bias (BIAS), the sample version of the root mean squared error (RMSE) -i.e. the root of the averaged squared deviation between a parameter and its estimate- and the sample standard deviation of the estimates (STD), respectively

$$BIAS(\phi^*) = \sum_{j=1}^r (\phi_j^* - \phi_j) / r,$$

$$RMSE(\phi^*) = \sqrt{MSE(\phi^*)}, \text{ where } MSE(\phi^*) = \sum_{j=1}^r (\phi_j^* - \phi_j)^2 / r,$$

$$STD(\phi^*) = \sqrt{\sum_{j=1}^r (\phi_j^* - \bar{\phi}^*)^2 / (r-1)},$$

where r is the number of replications, ϕ_j^* is the estimate of one of the considered parameters (i.e. δ_2 for model R_0 , and δ_1 , δ_2 , σ_{11} , σ_{22} and ρ for model B_2) on a particular replication j , ϕ_j represents the generating value of that parameter and replication (whenever it is changed) and $\bar{\phi}^*$ the mean estimate of ϕ^* across replications.

Simulation 1: Datasets generated from model G_2 . In the formulation of model G_2 , the strength of the stochastic dependency between the ability and the response propensity may be interpreted an index of the nonignorability of the missing data mechanism ([2]). Besides, as seen from above, model B_2 mimics model G_2 . Indeed, for both these models the person parameters θ_1 and θ_2 assume inherently different but similar interpretation. To better understand similarities and differences between these models a simulation study is performed to evaluate latent correlation parameter recovery – in a 12-item test, when the data are generated from model G_2 and analyzed using the RRM in its form B_2 . A relatively small data set is considered. This relatively small dataset is used to simulate similar conditions to that of the case study in the following section. For every replication, a random sample of size $n = 250$ of latent trait values (θ_1, θ_2) is drawn from a 2-dimensional normal distribution with means zero, variances 1 and correlation ρ^{G_2} equal to $-0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6$ and 0.8 , and two different sets, one for $\delta_1^{G_2}$ and one for $\delta_2^{G_2}$, of 12 item parameters

values are randomly selected from a standard normal distribution. Then, for each person a response vector is generated using model G_2 . One hundred replications are made for each fixed value of the correlation parameter. As illustrated in Table 2, the estimate of $|\rho^{B_2}|$ appears to be very well recovered, especially when the generating correlation parameter ρ^{G_2} is close to zero. (That is not the case of model B_1 , obviously; results not reported here). This result corroborates the non-obvious fact that $CORR(\theta_1, \theta_2)$ has not only similar interpretation for models B_2 and G_2 , but also same/similar numerical values in a neighborhood of the zero. As a consequence, the following procedure may be suggested: 1) if ρ^{G_2} , and then ρ^{B_2} , are near to zero there is no real need to use these models, because the response propensity of a person is not related to the proficiency of interest; 2) if ρ^{G_2} , and then ρ^{B_2} , are substantially different from zero both these models are equally suited to account for nonignorable missing data. Note that the condition $\rho^{G_2} \neq 0$ implies that the non-response process is *missing not at random* (MNAR).

Table 2. Estimation of correlation parameter ρ^{B_2} - parameter recovery.

Generating value	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8
BIAS	0.039	0.005	0.017	-0.011	-0.011	-0.015	-0.004	-0.006	-0.033
STD	0.078	0.101	0.132	0.142	0.134	0.124	0.126	0.108	0.078

Data generated from HG's model G_2 and analyzed using RRM in its form B_2 . Sample size $n=250$.

Simulation 2: Datasets generated from an RRM. This simulation study is conducted to assess parameter recovery and the effect of latent correlation parameter as an index of ignorability. The data are generated from an RRM in its form B_2 and are analyzed using both the generating model and model R_0 . For every replication, a random sample of size $n = 10000$ of latent trait values (θ_1, θ_2) is drawn from a 2-dimensional normal distribution with means zero, variances 1 and correlation ρ , with $\rho = 0, 0.2, 0.4, 0.6, 0.8$, and two different sets of 15 item parameters values, δ_1 and δ_2 , are randomly selected from a standard normal distribution. Then for each person a response vector is generated using an RRM in its form B_2 . One hundred replications are made for each fixed value of ρ . Parameter estimation is done with the Monte Carlo method, using 2000 nodes and a convergence criterion of 0.0001. For each dataset, we compute the estimates $\tilde{\delta}_2$ from model R_0 , and the estimates $\hat{\delta}_1$, $\hat{\delta}_2$ and $\hat{\Sigma}$ from the generating model B_2 , using MML as implemented with the calibration program ConQuest. Model B_2 differs from model B_1 in that it gives different estimates for the elements of Σ - but the estimates of the delta parameters coincide. Table 3 and Table 4 show the results from the simulations - where the RMSE of the delta parameters are summarized with a single value as $RMSE(\delta_u^*) = \sqrt{\sum_{i=1}^{15} MSE(\delta_{ui}^*)} / 15$, $u = 1, 2$, and analogously for the BIAS, as $BIAS(\delta_u^*) = \sum_{i=1}^{15} BIAS(\delta_{ui}^*) / 15$.

Table 3. Estimation the covariance matrix $\Sigma = (\sigma_{ij})$: parameter recovery.

parameter	Generating value of ρ					
	0.0	0.2	0.4	0.6	0.8	
ρ	BIAS	-0.001	-0.004	-0.004	-0.003	-0.004
	RMSE	0.016	0.016	0.013	0.011	0.009
σ_{11}	BIAS	0.003	0.007	0.005	0.002	0.001
	RMSE	0.022	0.023	0.018	0.018	0.019
σ_{22}	BIAS	-0.007	-0.006	-0.008	-0.008	-0.011
	RMSE	0.027	0.025	0.026	0.025	0.026

Data generated from RRM in its form B_2 and analyzed using the generating model. Sample size $n=10000$.

Table 4. Item parameter recovery.

Model	Parameter	Generating value of ρ					
		0.0	0.2	0.4	0.6	0.8	
B_2	δ_1	BIAS	0.004	0.006	0.003	0.004	0.002
		RMSE	0.037	0.036	0.036	0.033	0.032
	δ_2	BIAS	0.001	-0.003	-0.003	0.001	0.001
		RMSE	0.041	0.037	0.039	0.039	0.037
R_0	δ_2	BIAS	0.000	-0.019	-0.037	-0.051	-0.071
		RMSE	0.053	0.051	0.061	0.074	0.093

Data generated from RRM in its form B_2 . Sample size $n=10000$.

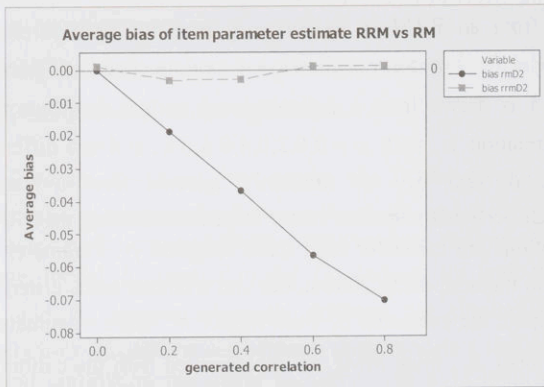


Figure 2. Relationship between the BIAS (100×15 parameters) of the estimators $\hat{\delta}_2$ and $\tilde{\delta}_2$ of item parameter δ_2 (item “difficulty”) and the generating correlation parameter ρ

The BIAS of estimators $\hat{\delta}_1$ and $\hat{\delta}_2$ is always very close to zero, while that of $\tilde{\delta}_2$ clearly grows, in its absolute value, as ρ increases (see Figure 2). As a consequence, this produces, for $\tilde{\delta}_2$, an inflation of its MSE (see Figure 3). As expected, with respect to model R_0 , the model B_2 (generating model) produced smallest MSEs under all the considered values of ρ (see Figure 3), with the only exception of the case of $\rho=0$, where the estimated error is similar for both the models B_2 and R_0 . This confirms that, when the correlation between δ_1 and δ_2 is near to zero, there is no real need to use the RRM.

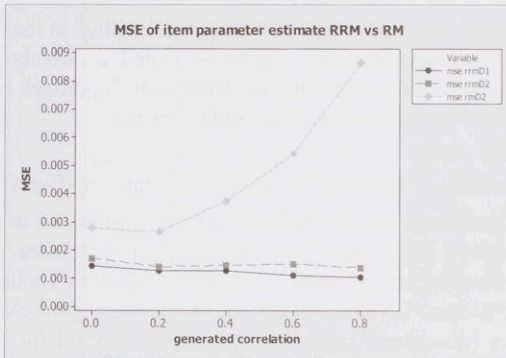


Figure 3. Relationship between the MSE (100×15 parameters) of the estimators $\hat{\delta}_1$, $\hat{\delta}_2$ and $\tilde{\delta}_2$ of item parameters δ_1 and δ_2 and the generating correlation parameter ρ

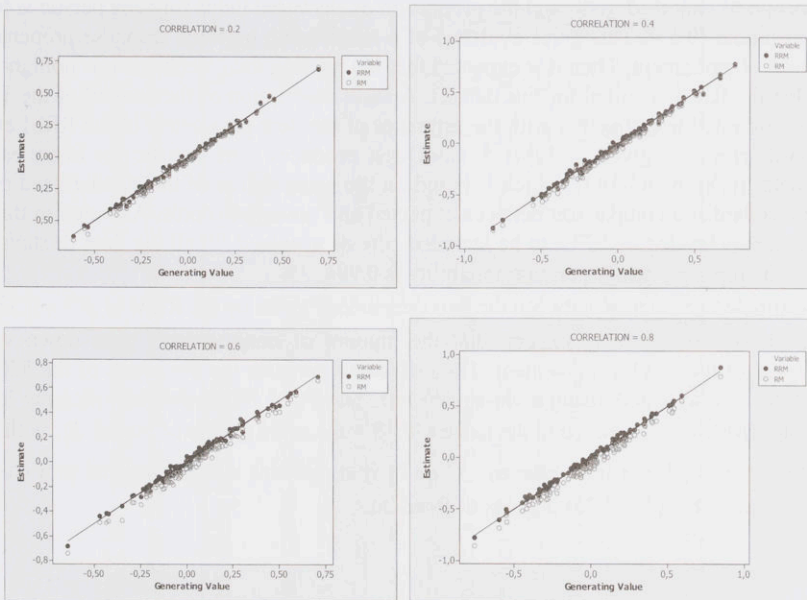


Figure 4. Relationship between estimated and generating values (all the values are averaged across 15 items, for each replication). The solid line is the bisectric.

Interesting enough, $\tilde{\delta}_2$ is clearly biased downward: see also Figure 4, where it is shown the relationship between true and estimated item parameters, for both the models B_2 and R_0 , for generating values of ρ equal to 0.2, 0.4, 0.6 and 0.8. For each replication j , $j = 1, \dots, 100$, the estimates of delta parameters – averaged across the 15 estimates of item parameters as $\sum_{i=1}^{15} \phi_i^* / 15$ – were compared with the true delta values – averaged across the 15 generating values as $\sum_{i=1}^{15} \phi_i / 15$. The fact that model R_0 yielded more negative bias as ρ increases may be easily explained by realizing that the tendency of an item to be skipped is partially explained, by model B_2 , by its difficulty, while that is not the case for model R_0 that ignores the missing data mechanism and tends to overestimate the “ability” of the respondent. These results indicate that item parameters can be estimated accurately under the proposed model (with the use of the software ConQuest), whereas the approach based on the simple Rasch Model ignoring missing data may yield biased estimates.

Illustration with real data

We illustrate the applicability of the RRM B_2 using a real data set. The Voluntary Counseling and Testing Efficacy Study is a randomized clinical trial conducted to test the efficacy of voluntary HIV-1 counseling and testing in reducing sexual risk behaviour ([25], [26]). A selection of 12 dichotomously scored items – concerning opinions about condom use – from this dataset has been considered for this application. Due to the negative wording of six out of the 12 items, the scores (disagree=0, agree=1) have been reversed for these questions (see Table 5, column 3). A total of 258 male subjects are considered in this study: all and only those with at least one nonresponse across the 12 items. The total amount of missing data is about 20%. As seen from above, as a general rule the strength of the stochastic dependency between the ability (here the level of agreement) and the response propensity is an index of the nonignorability of the missing data mechanism. In this case, the correlation between the proportion of endorsed items and the proportion of answered items for each person is found to be substantial (0.53). This gives evidence of a relationship between response propensity and the level of agreement. Then it is expected that the missing data mechanism is nonignorable – and that the RRM is suited for this dataset. A short description of the content of the 12 items of the calibrated test, together with the estimates of the item parameters of the RRM and their standard errors are given in Table 5. ConQuest produces a fit statistic for every estimated parameter ([18], p.143-144), which is based on the computation of the standardized residual, i.e. a standardized comparison between expected and observed scores. As noted in this Table (see columns headed by “ T ” – to be intended here as *weighted fit*) all the item fit statistics are good. The reported item separation reliability is 0.994.

The estimated correlation between the two dimensions given by the RRM is $\hat{\rho}^{B_2} = 0.338$. This result shows, as one may expect, that the amount of nonresponses goes down with the increment of the level of agreement. The difference between the deviances of the RRM and the PCM is also tested, using a chi-square test. The PCM fit the data significantly less well than the RRM: the difference in deviance (38.48 – the same as that of model B_1 , with respect to which the PCM is a limit case as $\rho^{B_1} \rightarrow 1$) is significant when compared to a chi square distribution with 2 ($2=27-25$) degrees of freedom.

Table 5. Item parameter estimates for the *Attitudes towards condoms* test. The number of respondents in the categories A, B and C are given. The item parameters estimates are given along with their standard errors in parentheses.

Item wording (abbreviated)	reverse scoring	A	B	C	δ_i	MNSQ	IC	T	δ_j	MNSQ	IC	T
I1 using condoms good protection from aids		1	25	232	-3.42 (1.00)	0.98	(0.00, 2.91)	0.30	-2.45 (0.21)	0.98	(0.69, 1.31)	-0.10
I2 sex not as good when you use a condom	yes	66	117	75	-0.66 (0.15)	1.10	(0.85, 1.15)	1.40	0.54 (0.14)	1.06	(0.87, 1.13)	0.90
I3 embarrassing to buy condoms	yes	6	77	175	-2.76 (0.42)	0.92	(0.27, 1.73)	-0.10	-0.91 (0.14)	0.98	(0.89, 1.11)	-0.30
I4 using condoms good pregnancy prevention		10	31	217	-1.30 (0.33)	0.98	(0.45, 1.55)	0.00	-2.14 (0.18)	0.99	(0.78, 1.22)	-0.10
I5 embarrassing put on condom	yes	20	63	175	-1.31 (0.24)	0.94	(0.64, 1.36)	-0.30	-1.12 (0.14)	0.97	(0.89, 1.11)	-0.60
I6 frnds think use condoms inclng w/spouse		45	115	98	-1.06 (0.17)	0.94	(0.80, 1.20)	-0.60	0.22 (0.14)	0.99	(0.90, 1.10)	-0.10
I7 condoms often break or slip	yes	103	81	74	0.23 (0.13)	1.09	(0.91, 1.09)	1.80	0.17 (0.15)	1.00	(0.87, 1.13)	0.00
I8 if sex partner wants condom I suspect	yes	17	129	112	-2.22 (0.26)	1.00	(0.60, 1.40)	0.10	0.18 (0.13)	1.03	(0.92, 1.08)	0.60
I9 friends use condoms w/new partner		118	47	93	0.95 (0.13)	1.04	(0.91, 1.09)	0.90	-0.65 (0.14)	1.00	(0.90, 1.10)	-0.10
I10 friends think that condoms uncomfortable	yes	79	131	48	-0.58 (0.14)	0.90	(0.88, 1.12)	-1.60	1.15 (0.17)	1.05	(0.81, 1.19)	0.60
I11 friends think always use condom new person		69	48	141	0.31 (0.15)	0.94	(0.86, 1.14)	-0.80	-1.13 (0.13)	0.92	(0.91, 1.09)	-1.90
I12 most people your age using condoms now		77	64	117	0.14 (0.14)	1.00	(0.87, 1.13)	-0.10	-0.60 (0.13)	1.01	(0.91, 1.09)	0.20

Conclusions

It is well-known that the mechanism causing nonresponse may lead to a violation of the ignorability principle ([14]). As a method for handling missing data, when nonresponses originate from a respondent's choice, the missing data information may be taken into account by considering a general person's tendency to respond as a latent trait. In this work, proficiency and latent response propensity are both incorporated in a new multidimensional class of IRT model, called RRM, belonging to the Rasch family of statistical models. The RRM is a two-dimensional generalization of the simple RM for the treatment of binary data in presence of nonignorable nonresponses –for example in the case of intentionally omitted items, i.e. when a person choose to not respond. More specifically, the RRM incorporates the response propensity of the person, as a latent variable. In its form called B_2 , the RRM mimics model G_2 described by HG ([2], [5], [9]). The simple interpretability of their parameters is an obvious quality of models B_2 and G_2 –respectively, within the class of the RRM, and within the HG family of models. In particular: 1) models B_2 and G_2 share the same number of parameters; 2) for both these models the 2-dimensional latent trait, (θ_1, θ_2) , assume inherently different but similar interpretation, i.e., θ_1 is related to the response propensity and θ_2 is related to the ability. Moreover: 3) a simulation study gives empirical evidence that in a neighborhood of the zero, the correlation assumes same/similar values for both. As a consequence of all the above, the following rule may be drawn. For small levels of correlation between ability and response propensity there is no need to apply an RRM, simply because the RM that ignores the missing data, R_0 , works well. Otherwise, i.e. for moderate to high levels of correlation (MNAR case), both the models G_2 and B_2 are equally suited to account for nonignorable missing data (while it is proved that R_0 leads to biased estimates).

However, strictly speaking, these findings should be considered with caution as a general rule. The reader should be aware that the MNAR assumption is not verifiable, since non-random processes allows the missingness to depend on the unobserved values. As shown in [27], an omnibus assessment of MAR versus MNAR is not feasible. This is due to the fact that every MNAR model can be replaced with an MAR version (counterpart) of this model which produce the same fit to the observed data as the original MNAR model. Nevertheless, this “counterpart” does not lead to a member of a conventional parametric family. Then, as far as concerns the comparison between the models G_2 and B_2 , our above considerations remain valid.

Ultimately, with regard to definition, interpretation and estimation, these two models are substantially equivalent, and the only relevant difference between model G_2 and model B_2 is that the latter belongs to the exponential family of distribution, while the former not.

Being a member of the exponential family of distributions, the RRM can be viewed as a generalized linear mixed model ([28], [29], [30], [31]). It can also be viewed as an instance of a MRCMLM, with a structure of within-item multidimensionality. As a member of the exponential family, the RRM offers several specific advantages, e.g.: 1) Existence of non trivial sufficient statistics for the parameters, and possibility of specific objective parameter estimation; 2) Feasibility of conditional inference; 3) Known conditions for identifiability of the model parameters; 4) Known necessary and sufficient conditions for existence and the uniqueness of the CML estimates; 5) Larger class of diagnostic goodness of fit tests (e.g. conditional likelihood ratio test); 6) Availability of specific software for Rasch Analysis; 7) Availability of simplified (natural exponential family) methods for bias reduction ([32]).

While a fixed effect approach to estimation it is also possible (e.g. via CML estimation), in this work a random effect approach has been adopted. Simulation analyses are used to evaluate parameter recovery for the RRM. Results of the simulations indicate that the parameters of the RRM can be recovered adequately, using ConQuest software. Moreover, the simulation study confirms that treating omitted responses as ignorable –which is the default option for this software- may lead to heavily distorted item parameters. Finally, a real data set was analyzed to illustrate the interpretation of the parameters of the model.

References

- [1] Moustaki, I., & O'Muircheartaigh, C. (2000). A one dimension latent trait model to infer attitude from nonresponse for nominal data. *Statistica*, 259-276.
- [2] Holman, R., & Glas, A.W. (2005). Modelling non-ignorable missing-data mechanism with item response theory models. *British Journal of Mathematical and Statistical Psychology*, 1-17.
- [3] Korobko, O. B., Glas, C. A. W., Bosker, R. J., & Luyten, J. W. (2008). Comparing the difficulty of examination subjects with item response theory. *Journal of Educational Measurement*, 45, 139-157.
- [4] Glas, C. A. W., & Korobko, O. B. (2007). A marginal maximum likelihood procedure for an IRT model with single-peaked response functions. OMD Progress Report No. 07-01, University of Twente.
- [5] Glas, C. A. W., & Pimentel, J.L. (2006). Modelling non-ignorable missing-data process in item calibration. LSAC Research Report Series, CTR 04-07.
- [6] Knott, M., & Tzamourani, P. (1997). Fitting a latent trait model for missing observations to racial prejudice data. In J. Rost & R. Langeheine (Eds) *Applications of latent trait and latent class models in the social sciences*, Waxmann, 224-252.
- [7] O'Muircheartaigh, C., & Moustaki, I. (1999). Symmetric pattern models: a latent variable approach to item non-response in attitude scales. *Journal of the Royal Statistical Society, Series A*, 162, 177-194.
- [8] Moustaki, I., & Knott, M. (2000). Weighting for item non-response in attitude scales by using latent variable models with covariates. *Journal of the Royal Statistical Society, Series A*, 163, 445-459.
- [9] Rose, N., von Davier, M., & Xu, X. (2010). Modeling non-ignorable missing data with IRT. ETS Research Report No. RR-10-10, Princeton.
- [10] Bernaards, C. A., & Sijtsma, K. (1999). Factor analysis of multidimensional polytomous item response data suffering from ignorable item nonresponse. *Multivariate Behavioral Research*, 34, 277-313.
- [11] Bernaards, C. A., & Sijtsma, K. (2000). Influence of imputation and EM methods on factor analysis when item nonresponse in questionnaire data is nonignorable. *Multivariate Behavioral Research*, 35, 321-364.
- [12] Conaway, M. R. (1992). The analysis of repeated categorical measurements subject to nonignorable nonresponse. *Journal of the American Statistical Association*, 87, 817-824.
- [13] Park, T., & Brown, M. B. (1994). Models for categorical data with nonignorable nonresponse. *Journal of the American Statistical Association*, 89, 44-52.
- [14] Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. 2nd Edition, Wiley.
- [15] Reckase, M.D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement*, 9, 401-412.
- [16] Reckase, M.D. (1997). A linear logistic multidimensional model for dichotomous item response data. In W.J. van der Linden & R.K. Hambleton (Eds) *Handbook of item response theory*, Springer-Verlag.

- [17] Rost, J., & Carstensen, C. H. (2002). Multidimensional Rasch measurement via item component models and faceted designs. *Applied Psychological Measurement*, 26, 42-56.
- [18] Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S., (2007). ACER ConQuest 2.0: General item response modelling software [computer program manual], ACER Press.
- [19] Adams, R.J., Wilson, M.R., & Wang, W.C. (1997). The multidimensional random coefficients multinomial logit. *Applied Psychological Measurement*, 21, 1-24.
- [20] Rasch, G. (1961). On general laws and the meaning of measurement in psychology, *Proceedings of the IV. Berkeley Symposium on mathematical statistics and probability*, Vol. IV, 321-333.
- [21] Andersen, E. B. (1973). Conditional inference for multiple choice questionnaires. *British Journal of Mathematical and Statistical Psychology*, 26, 31-44.
- [22] Kelderman, H., & Rijkes, C. P. M. (1994). Loglinear multidimensional IRT models for polytomously scored items. *Psychometrika*, 59, 149-176.
- [23] Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of the EM algorithm. *Psychometrika*, 46, 443-459.
- [24] Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39, 1-38.
- [25] The Voluntary HIV-1 Counseling and Testing Efficacy Study Group (2000). Efficacy of voluntary HIV-1 counselling and testing in individual and couples in Kenya, Tanzania, and Trinidad: a randomised trial. *Lancet*, 256, 103-112.
- [26] The Voluntary HIV-1 Counseling and Testing Efficacy Study Group (2000) The Voluntary HIV-1 Counseling and Testing Efficacy Study: Design and Methods. *AIDS and Behavior*, 4, 5-14.
- [27] Molenberghs, G, Beunckens, C, & Sotito, C. (2008). Every missingness not at random model has a missingness at random counterpart with equal fit. *Journal of the Royal Statistical Society B*, 70, 371-388.
- [28] De Boeck, P., & Wilson, M. R. (Eds) (2004), *Explanatory item response models: A generalized linear and nonlinear approach*. New York: Springer-Verlag.
- [29] McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.), Chapman & Hall.
- [30] McCulloch, C. E., & Searle, S. R. (2001). *Generalized, linear, and mixed models*, Wiley.
- [31] Rijmen, F., Tuerlinckx, F., De Boeck, P., & Kuppens, P. (2003). A nonlinear mixed model framework for item response theory. *Psychological Methods*, 8, 185-205.
- [32] Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80, 27-38.