Psicológica (2013), 34, 97-123.

Rasch Analysis for Binary Data with Nonignorable Nonresponses

Lucio Bertoli-Barsotti* [1] & Antonio Punzo[2]

[1]*University of Bergamo, Italy;* [2]*University of Catania, Italy*

This paper introduces a two-dimensional Item Response Theory (IRT) model to deal with nonignorable nonresponses in tests with dichotomous items. One dimension provides information about the omitting behavior, while the other dimension is related to the person's "ability". The idea of embedding an IRT model for missingness into the measurement model is not new but, differently from the existing literature, the model presented in this paper belongs to the Rasch family of models. As a member of the exponential family, the model offers several advantages, such as existence of non trivial sufficient statistics and possibility of specific objective parameter estimation; feasibility of conditional inference; goodness of fit analysis via conditional likelihood ratio tests. Maximum likelihood estimation is discussed, and the applicability of the proposed model is illustrated by using a real data set.

Introduction

In applications of Item Response Theory (IRT), researchers are often faced with the problem of missing data. This problem may be particularly critical when the ignorability principle does not hold (Rubin, 1976; Little & Rubin, 2002). Missing data are said to be missing at random (MAR; Little & Rubin, 2002, p. 12) when the process generating the pattern of missingness does not depend on the unobserved data, although it may depend the observed data. When the probability of missingness is independent of both unobserved and observed data, the missing data mechanism is called missing completely at random (MCAR; note that MCAR implies MAR). When data are MAR, the missing data mechanism may be considered "ignorable" for likelihood-based inferences if the parameter of interest and the parameter of the missingness process are distinct (Little & Rubin, 2002, pp. 117-120). Roughly speaking, Rubin's

* Correspondence concerning this article should be addressed to Lucio Bertoli-Barsotti, Dipartimento di Matematica, Statistica, Informatica e Applicazioni, University of Bergamo, Via dei Caniana 2, 24127 Bergamo (Italy). E-mail: lucio.bertoli-barsotti@unibg.it

ignorability principle states the conditions under which ignoring the process that causes the missingness does not generate any systematic errors in the parameter estimates.

In a broad sense, we may distinguish two main cases of missingness. The case in which the observations result from an incomplete design of the test decided by the researcher (planned missingness; e.g. random incomplete designs, multistage testing designs, targeted testing designs), and the case in which the missing may be considered as the effect of a choice of the respondent (unplanned missingness). When data are missing by design, usually (for possible exceptions see Eggen & Verhelst, 2011) randomization warrants they are at least MAR (Little & Schenker, 1995, p. 43; Schafer, 1997, p.20-22 and p.62; Mislevy & Wu, 1996). In this paper we mostly consider the latter case, and more specifically the case in which a) an item is presented to a person, b) that person has time to consider it (thus, we do not refer to “not reached” items), but c) decides, for whatever reason, to not respond (Mislevy & Wu, 1996). This choice may be due to inability to understand, or unwillingness to respond for embarrassment, anger, discomfort, or other reason that may, or may not, depend on the latent trait to be measured. As it is well known, in these cases missingness is not generally ignorable. A typical situation is represented by items which are skypped because of the low “proficiency” of the respondent. But other examples can be mentioned. As an example, consider the problem of measuring “ability” to perform activities of daily life. With this aim, Holman and Glas considered the ALDS item bank (see Holman, Lindeboom, Vermeulen, Glas, & de Haan, 2001, for details), founding that "patients with a higher proficiency level tended to boost their rating by failing to respond, while the patients of low proficiency were less inclined or motivated to impress the nurses" (Holman & Glas, 2005, p. 9). These intentionally omitted responses represent a case of nonignorable missingness, because nonresponse depends on the unobserved data; in other words, the number of items endorsed is correlated with the respondent’s proficiency level.

In this article, we focus on the situation in which a person latent trait ϑ is measured by a test composed of k dichotomous items. The parameter ϑ_v may be thought as the amount of “ability” (i.e. proficiency, but also agreement, motivation, belief, attitude, capacity, intention, and so on) of respondent v ($v = 1,\dots,n$). For ease of exposition, in what follows we refer to this latent trait as “ability”. Original response categories (e.g. yes/no, right/wrong, agree/disagree, correct/incorrect, and so on) are supposed to be recoded as “1” and “0”, defining in this way a binary response variable X_{vi}, with reference to the subject v and the item indexed by i ($i = 1,...,k$). Conditionally to the fact that the response was observed, we assume that X_{vi} follows a Bernoulli distribution depending on the latent trait ϑ_v -and other item parameters- through an IRT model. We shall refer to this model as measurement model. The observations are collected into a $n \times k$ dataset under the basic assumption of the conditional independence between responses given ϑ (local independence).

All the models considered in the present paper allow to take into account the presence, if any, of a nonignorable missing data mechanism – due to omitting behavior. An appropriate method to deal with this type of missingness is to incorporate the mechanism that caused the missingness into the measurement model. A possible approach consists in using a latent variable model that is a function of a two-dimensional latent trait (ξ, ϑ), where ξ is related to the response propensity level. This model-based approach allows missing values to be included into the analysis and, equally important, it allows information about attitude to be inferred from nonresponse. From now on, for convenience, we will write $\xi = \theta_1$ and $\vartheta = \theta_2$, where θ_1 and θ_2 are the components of a vector valued latent variable $\boldsymbol{\theta}$. This approach was introduced by Knott, Albanese and Galbraith (1990) and Albanese and Knott (1992). Successively, the same model (in this paper this model is defined in equation (3), in the next section) was also studied and applied, among others, by Knott and Tzamourani (1997), Bartholomew, de Menezes and Tzamourani (1997), O'Muircheartaigh and Moustaki (1999), O'Muircheartaigh and Moustaki (1996) – and, in a slightly more general form, by Moustaki and Knott (2000), Moustaki and O'Muircheartaigh (2000), Moustaki and O'Muircheartaigh (2002). More recently, a unified approach to a more general class of models has been proposed in a seminal paper by Holman and Glas (Holman & Glas, 2005; see also Glas & Pimentel, 2006, and Pimentel, 2005, Chapter 2). Comparative studies of alternative models within the family of Holman and Glas (2005) are also given by Rose, von Davier and Xu (2010). Now, all the models in that literature are instances of two-dimensional IRT models; more specifically, by construction, they do not belong to the exponential family of distributions. What is more, this precludes the use of a conditional approach to the estimation.

Following the above said model-based approach for the treatment of nonignorable nonresponses, we introduce a new two-dimensional IRT model - that belongs to the Rasch family of models - for the analysis of dichotomously scored items in the presence of nonignorable nonresponses, called Rasch-Rasch Model (RRM). As a member of the exponential family, the RRM offers several advantages: 1) Existence of non trivial sufficient statistics 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 Conditional Maximum Likelihood (CML) estimates; 5) Existence of a conditional likelihood ratio test for goodness of fit.

This article is organized as follows. The next section contains a brief review of the existing models. The third section contains the formulation of the proposed model. Then, minimal sufficient statistics for the model parameters are presented. In the subsequent section the problem of the estimation is investigated. In the sixth section a small simulation study is carried out to evaluate parameter recovery. A case study is finally considered in the seventh section.

A standard non-Rasch approach to the problem

Let D_{vi} be the random variable response indicator of X_{vi}, assuming value 1 if the response of person v on item i was observed, and 0 otherwise. If D_{vi} = 1, the response variable X_{vi} may assume the values 0 and 1. Otherwise $X_{vi} = c$, where c is an arbitrary constant.

According to the model-based approach for the treatment of nonignorable nonresponses, the underlying process leading to the values of X_{vi} can be explained in a hierarchic way, by two different steps. At step 1, when subject v encounters item i, she/he can choose to answer or not. This step is described by the random variable D_{vi}. At step 2, if subject v has chosen to respond (i.e. D_{vi} = 1), then she/he can respond in the category coded as “0” (i.e. $X_{vi} = 0$) or in the category coded as “1” (i.e. $X_{vi} = 1$) depending on her/his proficiency level $\vartheta = \theta_2$. This step is governed by the random variable $X_{vi} \mid D_{vi}$, taking values 0 and 1. Summarizing, there are three possible response patterns for any item, say

$$A = (D = 0, X = c),\ B = (D = 1, X = 0) \text{ and } C = (D = 1, X = 1),$$

where the subscripts v and i have been omitted, for brevity. We will refer to these three patterns as “options” to distinguish them from the original dichotomous response categories coded as “0” and “1”. Then one can postulate the following model

$$P(A) = P(D=0),\ P(B) = P(X=0 \mid D=1)P(D=1),$$
$$P(C) = P(X=1 \mid D=1)P(D=1).$$

According to the described process, the probability of any response pattern $(D_{vi} = d_{vi}, X_{vi} = x_{vi})$ can be given explicitly as

$$P(D_{vi} = d_{vi}, X_{vi} = x_{vi}) = P(D_{vi} = d_{vi})\left[P(X_{vi} = x_{vi} \mid D_{vi} = 1)\right]^{d_{vi}}. \qquad (1)$$

Let us denote by π_{viA}, π_{viB} and π_{viC}, respectively, the probabilities of the events: $(D_{vi} = 0, X_{vi} = c)$, $(D_{vi} = 1, X_{vi} = 0)$ and $(D_{vi} = 1, X_{vi} = 1)$. A possible IRT model for each of the terms on the right-hand side of equation (1) is represented by the two-parameter logistic (2PL; Birnbaum, 1968) model, which leads to

$$\ln\frac{\pi_{viB} + \pi_{viC}}{\pi_{viA}} = a_{1i}\,\theta_{1v} - \delta_{1i}, \qquad \ln\frac{\pi_{viC}}{\pi_{viB}} = a_{2i}\,\theta_{2v} - \delta_{2i} \qquad (2)$$

where a_{1i} and a_{2i} are item discrimination parameters and the deltas are item parameters.

In order to obtain parameter estimates, one could consider Marginal Maximum Likelihood (MML) in which person parameters are integrated out by marginalization, assuming a two-dimensional normal distribution, with density $g(\boldsymbol{\theta}) = g(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$, for the latent variable $\boldsymbol{\theta}$, with mean vector $\boldsymbol{\mu} = \mathbf{0}$ (corresponding to an identifiability constraint) and unknown covariance matrix $\boldsymbol{\Sigma}$. This model is called G_2 in the taxonomy of Holman and Glas (2005); they also perform a simulation study by considering G_2 in its simplest form, given by $a_{1i} = a_{2i} = 1$. Interestingly, in this way, although each logit equation in (2) represents a Rasch Model (RM; Rasch, 1960) for dichotomous items, the whole model (1) does not belong to the Rasch family of models.

In order to provide possibly more meaningful parameters, the model given in (2) can also be reparameterized by writing $\boldsymbol{a}'_{1i}\,\boldsymbol{\theta}_v$ instead of $a_{1i}\,\theta_{1v}$ and $\boldsymbol{a}'_{2i}\,\boldsymbol{\theta}_v$ instead of $a_{2i}\,\theta_{2v}$, where $\boldsymbol{a}_{1i} = (a_{11i}, a_{12i})'$ and $\boldsymbol{a}_{2i} = (a_{21i}, a_{22i})'$ are vectors of item discrimination parameters. This means that the variables D_{vi} and/or $X_{vi} \mid D_{vi}$ may depend (possibly) on both the components θ_1 and θ_2 of the latent trait $\boldsymbol{\theta}$. In that way, we obtain the following more general version of the previous model (2)

$$\ln\frac{\pi_{viB}+\pi_{viC}}{\pi_{viA}} = a_{11i}\,\theta_{1v} + a_{12i}\,\theta_{2v} - \delta_{1i}\,,\ \ln\frac{\pi_{viC}}{\pi_{viB}} = a_{21i}\,\theta_{1v} + a_{22i}\,\theta_{2v} - \delta_{2i}$$

where it may be noted that both these logit equations are now instances of a multidimensional 2PL (M2PL) model for dichotomous items (Rackase, 2009, p.86). According to these logit equations, the response probability function can be expressed explicitly as follows

$$\pi_{viA} = \left[1+\exp\left(\boldsymbol{a}'_{1i}\boldsymbol{\theta}_v - \delta_{1i}\right)\right]^{-1}$$

$$\pi_{viB} = \exp\left(\boldsymbol{a}'_{1i}\boldsymbol{\theta}_v - \delta_{1i}\right)\left\{\left[1+\exp\left(\boldsymbol{a}'_{1i}\boldsymbol{\theta}_v - \delta_{1i}\right)\right]\left[1+\exp\left(\boldsymbol{a}'_{2i}\boldsymbol{\theta}_v - \delta_{2i}\right)\right]\right\}^{-1} \quad (3)$$

$$\pi_{viC} = \exp\left(\boldsymbol{a}'_{1i}\boldsymbol{\theta}_v + \boldsymbol{a}'_{2i}\boldsymbol{\theta}_v - \delta_{1i} - \delta_{2i}\right)\left\{\left[1+\exp\left(\boldsymbol{a}'_{1i}\boldsymbol{\theta}_v - \delta_{1i}\right)\right]\left[1+\exp\left(\boldsymbol{a}'_{2i}\boldsymbol{\theta}_v - \delta_{2i}\right)\right]\right\}^{-1}.$$

To identify this model, we need to impose some constraints on the discrimination parameters $\boldsymbol{a}_{1i}$ and $\boldsymbol{a}_{2i}$. In particular, by taking $a_{21i}=0$, we obtain a model studied by Knott and Tzamourani (1997) (see also O'Muircheartaigh & Moustaki 1999, Knott et al., 1990, Albanese & Knott, 1992), that is:

$$\ln\frac{\pi_{viB}+\pi_{viC}}{\pi_{viA}} = a_{11i}\,\theta_{1v} + a_{12i}\,\theta_{2v} - \delta_{1i}\,,\ \ln\frac{\pi_{viC}}{\pi_{viB}} = a_{22i}\,\theta_{2v} - \delta_{2i}\,. \quad (4)$$

Model (4) has also been studied by Rose et al. (2010; referred to as "Model 7"). Holman and Glas (2005) denoted this model as G_3, while the case $a_{12i}=0$ is referred to as model G_4 and model G_2 is obtained by taking $a_{21i}=a_{12i}=0$. 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 $\boldsymbol{\theta}$ will differ across these models. It should be noted that for technical reasons, in the general case of models G_3 and G_4, the covariance matrix $\boldsymbol{\Sigma}$ should be constrained to be an identity matrix to allow the model to be identified.

A Rasch approach to the problem

Formulation of the model

A well-known drawback of the MML approach is that, if the distributional assumptions about $\boldsymbol{\theta}$ are wrong, the method is not consistent (Pfanzagl, 1994). As an alternative to the MML procedure one could also consider the Conditional ML (CML) method for item parameter estimation.

The advantage is that CML does not require distributional assumptions about $\boldsymbol{\theta}$. However, this method essentially requires that the model belongs to the exponential family. Now, to obtain a model that belongs to this family, in this paper we suggest to consider the following alternative formulation of model (2)

$$\ln \frac{\pi_{viB}}{\pi_{viA}} = \theta_{1v} - \delta_{1i}, \qquad \ln \frac{\pi_{viC}}{\pi_{viB}} = \theta_{2v} - \delta_{2i}. \tag{5}$$

Since each logit equation in (5) assumes the standard form of an RM, and the whole two-dimensional model model belongs to the Rasch family of models, (5) is called Rasch-Rasch Model (RRM). In its simplest form (but see below for generalizations), this model can be written succinctly as

$$P\left(D_{vi} = d_{vi}, X_{vi} = x_{vi}\right) = \frac{\exp\left[d_{vi}\left(\theta_{1v} - \delta_{1i}\right) + d_{vi}\, x_{vi}\left(\theta_{2v} - \delta_{2i}\right)\right]}{1 + \exp\left(\theta_{1v} - \delta_{1i}\right) + \exp\left(\theta_{1v} + \theta_{2v} - \delta_{1i} - \delta_{2i}\right)}.$$

From a comparison between the models given in (5) and in (2), it is straightforward to note that -while person and item parameters θ_2 and δ_2 maintain a similar interpretation for both these models, as “ability” and “difficulty” a difference may be found between the distinction *B* versus *A*, for the RRM, instead of (*B or C*) versus *A*, for the model given in (2). As a consequence, person and item parameters θ_1 and δ_1 have different meaning in the two models; more specifically, for example, in model (5) the item parameter δ_1 indicates the relative “difficulty” of choosing option *B* rather than *A*. This “slight” modification in the parameterization has the great advantage to pose the RRM in the exponential family preserving, although with a different interpretation of the parameters, the idea of embedding a model for missingness in the measurement model. Obviously, sometimes other response options may appear to be interesting and informative (e.g. *C* versus *A*). In these cases, alternative parameterizations of the RRM may also be adopted.

Alternative parameterizations and relationship with other families of IRT models

There are several types of alternative RRM parameterizations to choose from. Indeed, in analogy with the previous section, the model given in (5) can also be reparameterized by writing

$$\boldsymbol{a}_1' \boldsymbol{\theta}_v \text{ instead of } \theta_{1v} \text{ and } \boldsymbol{a}_2' \boldsymbol{\theta}_v \text{ instead of } \theta_{2v}, \tag{6}$$

where $\boldsymbol{a}_1 = (a_{11}, a_{12})'$ and $\boldsymbol{a}_2 = (a_{21}, a_{22})'$ are vectors of *known* and fixed coefficients of the trait parameters (rather than unknown discrimination parameters, as in the models of the previous section). According to the above reparameterization, the response probability function can be expressed explicitly as follows

$$\begin{aligned}
\pi_{viA} &= \left[1+\exp\left(\boldsymbol{a}_1'\boldsymbol{\theta}_v - \delta_{1i}\right)+\exp\left(\boldsymbol{a}_1'\boldsymbol{\theta}_v + \boldsymbol{a}_2'\boldsymbol{\theta}_v - \delta_{1i} - \delta_{2i}\right)\right]^{-1} \\
\pi_{viB} &= \exp\left(\boldsymbol{a}_1'\boldsymbol{\theta}_v - \delta_{1i}\right)\left[1+\exp\left(\boldsymbol{a}_1'\boldsymbol{\theta}_v - \delta_{1i}\right)+\exp\left(\boldsymbol{a}_1'\boldsymbol{\theta}_v + \boldsymbol{a}_2'\boldsymbol{\theta}_v - \delta_{1i} - \delta_{2i}\right)\right]^{-1} \quad (7) \\
\pi_{viC} &= \exp\left(\boldsymbol{a}_1'\boldsymbol{\theta}_v + \boldsymbol{a}_2'\boldsymbol{\theta}_v - \delta_{1i} - \delta_{2i}\right)\left[1+\exp\left(\boldsymbol{a}_1'\boldsymbol{\theta}_v - \delta_{1i}\right)+\exp\left(\boldsymbol{a}_1'\boldsymbol{\theta}_v + \boldsymbol{a}_2'\boldsymbol{\theta}_v - \right.\right. \\
&\left.\left.\delta_{1i} - \delta_{2i}\right)\right]^{-1},
\end{aligned}$$

where it is understood that the vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ have to be linearly independent to identify the model parameters. Note that since these constants are supposed to be known, the model remains within the Rasch family of models; in particular, the choice $\boldsymbol{a}_1 = (1,0)'$ and $\boldsymbol{a}_2 = (0,1)'$ yields the RRM in its simplest form, given in (5). Although all these models are identical up to reparameterization (6), the meaning of the latent variables θ_1 and θ_2 may change.

By comparing the probability functions in (3) and (7), it is apparent that the models are similar, but inherently different. In particular, it is worth noting that the RRM cannot be seen in any way as a special case of the class of models described in the second section. However, the RRM may be related to other known families of IRT models. More specifically, if *A* is recoded as "0", *B* is recoded as "1", and *C* is recoded as "2", the RRM (5) can also be viewed as an instance of the Multidimensional Polytomous Latent Trait (MPLT) model (Kelderman & Rijkes, 1994) – but it should be noted that the three response options *A*, *B* and *C* are not ordered, in the RRM framework. By the way, it is interesting to note that if we assume the unidimensionality of the latent trait, that is $\theta_1 = \theta_2$, the same recoding generates the well-known Partial Credit Model (PCM; Masters, 1982).

Since there is no total order among the options *A*, *B* and *C*, the RRM can also be considered as a special case of a Multidimensional Nominal Response Model (MNRM; Bolt & Johnson, 2009) where some of the parameters are constrained. Indeed, choosing option *B* as a reference point for every item, the MNRM can also be written as

$$\pi_{viA} \propto exp\left(\boldsymbol{b}_{1i}'\boldsymbol{\theta}_v + c_{1i}\right), \quad \pi_{viC} \propto exp\left(\boldsymbol{b}_{2i}'\boldsymbol{\theta}_v + c_{2i}\right),$$

where $\boldsymbol{b}_{1i}$ and $\boldsymbol{b}_{2i}$ are vectors of discrimination parameters, and c_{1i} and c_{2i} are intercept parameters (Thissen, Cai & Bock, 2010). Since the RRM (5) yields in particular the probabilities

$$\pi_{viA} = \frac{\exp\left[-\left(\theta_{1v} - \delta_{1i}\right)\right]}{1 + \exp\left[-\left(\theta_{1v} - \delta_{1i}\right)\right] + \exp\left(\theta_{2v} - \delta_{2i}\right)},$$

$$\pi_{viC} = \frac{\exp\left(\theta_{2v} - \delta_{2i}\right)}{1 + \exp\left[-\left(\theta_{1v} - \delta_{1i}\right)\right] + \exp\left(\theta_{2v} - \delta_{2i}\right)},$$

the model is equivalent to the MNRM by imposing the following constraints

$$\boldsymbol{b}_{1i} = \left(-1,0\right)' \text{ and } \boldsymbol{b}_{2i} = \left(0,1\right)'$$

for every item i. With the same parameterization, the RRM can also be viewed as an instance of a Multicategorical Multidimensional Rasch Model (Rasch, 1961; Andersen, 1973). Finally, by construction, the RRM may be seen (if the MML approach to the estimation is adopted) as an instance of a Multidimensional Random Coefficients Multinomial Logit Model (MRCMLM, Adams et al., 1997), which is the most general structure of a multidimensional Rasch model. It may also be noted that, since in our case each of the items relates to more than one dimension (θ_1 and θ_2), the model will be considered as a within-item multidimensional IRT model. To be precise about the specific kind of within-item multidimensionality, we may observe that in the RRM, for every item, the response process for each single option requires different latent traits.

Minimal sufficient statistics for the RRM parameters

Given a $n \times k$ dataset, under the usual assumption of local independence, the log-likelihood function for the RRM in its most general form, with reparameterization (6), is

$$l = \sum_{v=1}^{n}\left[\left(a_{11}d_{v\cdot} + a_{21}x_{v\cdot}\right)\theta_{1v} + \left(a_{12}d_{v\cdot} + a_{22}x_{v\cdot}\right)\theta_{2v}\right] - \sum_{i=1}^{k}\left(d_{\cdot i}\delta_{1i} + x_{\cdot i}\delta_{2i}\right) + C \qquad (8)$$

where

$$C = -\sum_{v=1}^{n}\sum_{i=1}^{k} \ln\left[1 + \exp\left(\boldsymbol{a}_1'\boldsymbol{\theta}_v - \delta_{1i}\right) + \exp\left(\boldsymbol{a}_1'\boldsymbol{\theta}_v + \boldsymbol{a}_2'\boldsymbol{\theta}_v - \delta_{1i} - \delta_{2i}\right)\right]$$

and where

$$d_{v\cdot} = \sum_{i=1}^{k} d_{vi} \text{ , } d_{\cdot i} = \sum_{v=1}^{n} d_{vi} \text{ , } x_{v\cdot} = \sum_{i=1}^{k} d_{vi} x_{vi} \text{ , } x_{\cdot i} = \sum_{v=1}^{n} d_{vi} x_{vi} \text{ .}$$

In particular, if $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are linearly independent, $a_{11}d_{v\cdot} + a_{21}x_{v\cdot}$, $a_{12}d_{v\cdot} + a_{22}x_{v\cdot}$ -or, equivalently, $(d_{v\cdot}, x_{v\cdot})$– and $(d_{\cdot i}, x_{\cdot i})$ result to be the joint minimal sufficient statistics for the parameters $(\theta_{1v}, \theta_{2v})$ and $(\delta_{1i}, \delta_{2i})$, respectively. Note that, by definition, the totals $d_{v\cdot}$ and $d_{\cdot i}$ represent the number of given answers by row and by column, respectively. Similarly, the totals $x_{v\cdot}$ and $x_{\cdot i}$ represent the number of responses in the category coded as "1" by row and by column, respectively.

Model estimation

Conditional maximum likelihood estimation approach

As said before, the MML method constitutes the standard estimation technique for the models presented in this paper. Nevertheless, for the RRM, we can further consider the CML method. In the CML approach, which is the natural approach in the exponential family framework (Lehmann, 1983), person parameters θ_{1v} and θ_{2v}, considered as nuisance parameters, are eliminated by conditioning on their sufficient statistics $d_{v\cdot}$ and $x_{v\cdot}$. In order to estimate the item parameters of the RRM (5), one can maximize the conditional log-likelihood function

$$l_C\left(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2 / d_{v\cdot}, x_{v\cdot}\right) = -\sum_{i=1}^{k} d_{\cdot i}\delta_{1i} - \sum_{i=1}^{k} x_{\cdot i}\delta_{2i} - \sum_{v=1}^{n} \gamma\left(d_{v\cdot}, x_{v\cdot} / \boldsymbol{\delta}_1, \boldsymbol{\delta}_2\right)$$

where

$$\gamma\left(d_{v\cdot}, x_{v\cdot} / \boldsymbol{\delta}_1, \boldsymbol{\delta}_2\right) = \sum_{\boldsymbol{d}^*,\, \boldsymbol{x}^* / d_{v\cdot}, x_{v\cdot}} \exp\left[-\sum_{i=1}^{k} d_i^*\left(\delta_{1i} + x_i^*\delta_{2i}\right)\right]$$

with the summation $\sum_{\boldsymbol{d}^*,\, \boldsymbol{x}^* / d_{v\cdot}, x_{v\cdot}}$ running across all answer patterns $\boldsymbol{x}^* = \left(x_1^*, x_2^*, \ldots, x_k^*\right)'$, coupled with the response indicator vector

$\boldsymbol{d}^* = \left(d_1^*, d_2^*, \ldots, d_k^*\right)'$, such that $\sum_i d_i^* = d_{v\cdot}$ and $\sum_i x_i^* d_i^* = x_{v\cdot}$. Then, $\gamma\left(r,t \mid \boldsymbol{\delta}_1, \boldsymbol{\delta}_2\right)$ denotes a special case of "elementary symmetric function" of order (r, t) (where, by construction, $r \geq t$) of the parameters $\delta_{11}, \delta_{12}, \ldots, \delta_{1k}$ and $\delta_{21}, \delta_{22}, \ldots, \delta_{2k}$. For example, consider a test with $k = 4$ items; for the case of $d_{v\cdot} = 3$ and $x_{v\cdot} = 2$ we find the elementary symmetric function

$$\gamma(3,2) = \eta_1\eta_2\tau_3 + \eta_1\eta_3\tau_2 + \eta_2\eta_3\tau_1 + \eta_1\eta_2\tau_4 + \eta_1\eta_4\tau_2 + \eta_2\eta_4\tau_1 + \\ + \eta_1\eta_3\tau_4 + \eta_1\eta_4\tau_3 + \eta_3\eta_4\tau_1 + \eta_2\eta_3\tau_4 + \eta_2\eta_4\tau_3 + \eta_3\eta_4\tau_2$$

where $\tau_i = \exp\left(-\delta_{1i}\right)$ and $\eta_i = \exp\left(-\delta_{1i} - \delta_{2i}\right)$.

CML item parameter estimates are obtained by maximizing l_C under two identifiability constraints, say $\sum_{i=1}^{k} \delta_{1i} = 0$ and $\sum_{i=1}^{k} \delta_{2i} = 0$. In a second step, these CML estimates can be substituted, as if they were the "true" item parameters, into the log-likelihood function (8) to give a profiled version of l that can be maximized with respect to $\boldsymbol{\theta}$. In this way, both item and person parameter estimates are obtained.

Marginal maximum likelihood estimation approach

Obviously, the MML approach is also possible for estimation of the parameters of an RRM. We remember that this approach is inherently based on what Holland (1990) called the *random sampling rationale* – i.e. the assumption that the subjects are sampled at random from a population with a specified distribution (e.g. a normal distribution). This method makes explicit use of the density function $g(\boldsymbol{\theta})$ of the latent variable, in order to obtain a likelihood function where the person parameters are integrated out. In other terms, in the random sampling perspective, the parameter $\boldsymbol{\theta}$ is simply a variable of integration and cannot be estimated. In our case, as already assumed in the second section, we will use a normal density $g\left(\boldsymbol{\theta} \mid \mathbf{0}, \boldsymbol{\Sigma}\right)$, leading to the following marginal log-likelihood function

$$l_M\left(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \boldsymbol{\Sigma}\right) = \sum_{v=1}^{n} \ln \int_{\Re^2} \prod_{i=1}^{k} P\left(d_{vi}, x_{vi} \mid \boldsymbol{\theta}, \delta_{1i}, \delta_{2i}\right) g\left(\boldsymbol{\theta} \mid \mathbf{0}, \boldsymbol{\Sigma}\right) d\boldsymbol{\theta}$$

which can be maximized to obtain estimates of item parameters and of the unknown covariance matrix $\boldsymbol{\Sigma}$. Possibly, estimates of the individual's position in the two-dimensional latent space are obtainable with a similar procedure as described for the CML approach, or by using the expected *a posteriori* (EAP) values given the response patterns. In the perspective of adopting the MML approach, it is useful to recognize that the RRM is an instance of a MRCMLM – for a convenient choice of both the design and the scoring matrices. Operationally, the computer program ConQuest (Wu, Adams, Wilson & Haldane, 2007), allowing for MML estimation of the parameters for a MRCMLM, can be used for the RRM too.

Simulation

A simulation study is conducted to assess parameter recovery and the effect of the latent correlation parameter on these estimates. The data are randomly generated from the RRM, in its form given in (5), and analyzed using both

a) the generating model;

b) the simple RM (with "difficulty" parameter δ^{RM}) ignoring the missing data mechanism.

The R software (R Development Core Team, 2008) was used to generate the data. For every replication, a random sample of size $n = 10000$ of latent trait values (θ_1, θ_2) is drawn from a two-dimensional normal distribution with means zero, variances 1 and correlation ρ, with ρ =0.0,0.2,0.6,0.8; two different sets, $\delta_{11}, \delta_{12}, ..., \delta_{1k}$ and $\delta_{21}, \delta_{22}, ..., \delta_{2k}$, of $k = 15$ item parameters values are randomly selected from a standard normal distribution. Then, for each person a response vector is generated. One hundred replications are made for each fixed value of ρ. For each dataset, we compute the MML item parameter estimates, say $\hat{\delta}_i^{RM}$ for the RM, and the estimates $\hat{\delta}_{1i}$, $\hat{\delta}_{2i}$ and $\hat{\rho}$ for the RRM (note that in this section we use a different notation to distinguish the "difficulty" parameter between the two models, i.e. $\hat{\delta}_i^{RM}$ for the RM and $\hat{\delta}_{2i}$ for the RRM), using the calibration program ConQuest. Parameter estimation is done with the Monte Carlo method, using 2000 nodes and a convergence criterion of 0.0001.

The accuracy of the parameter estimates is assessed with the criterion of the evaluation of the average bias (BIAS), and a sample version of the root mean squared error (RMSE; i.e. the root of the averaged squared deviation between a parameter and its estimate) as follows

$$\text{BIAS} = \frac{1}{r}\sum_{j=1}^{r}\frac{1}{k}\sum_{i=1}^{k}\left(\hat{\phi}_{ij} - \phi_{ij}\right)$$

$$\text{RMSE} = \sqrt{\frac{1}{r}\sum_{j=1}^{r}\frac{1}{k}\sum_{i=1}^{k}\left(\hat{\phi}_{ij} - \phi_{ij}\right)^2}$$

where r is the number of replications, $\hat{\phi}_{ij}$ is the estimated parameter of item i (i.e. δ_i^{RM} for the RM, and δ_{1i}, δ_{2i} for RRM) in replication j, and ϕ_{ij} represents the corresponding generating value of that parameter and replication. Table 1 gives the results for the item parameters, while Table 2 shows the results for the correlation parameter.

Table 1. Item parameter recovery for RRM and RM

			Generating value of ρ				
Model	Parameter		*0.0*	*0.2*	*0.4*	*0.6*	*0.8*
RRM	δ_1	BIAS	-0.0006	0.0018	0.0043	0.0028	0.0042
		RMSE	0.0336	0.0316	0.0328	0.0330	0.0327
	δ_2	BIAS	-0.0006	-0.0019	-0.0001	-0.0001	-0.0015
		RMSE	0.0340	0.0316	0.0326	0.0331	0.0333
RM	δ^{RM}	BIAS	-0.0610	-0.0890	-0.1153	-0.1442	-0.1771
		RMSE	0.0758	0.1021	0.1301	0.1601	0.1947

Note. The calibration of RM is done under the assumption of MAR.

The BIAS for parameters δ_1 and δ_2 is always very close to zero, while for δ^{RM} it grows, in its absolute value (the estimator seems to be biased downward), as ρ increases – as suggested by Figure 1. As a consequence, as can be seen in Table 1, this produces an inflation of the RMSE for the parameter δ^{RM}; as expected, with respect to the RM, the RRM produced the smallest RMSEs under all the considered values of ρ. These results indicate that item parameters can be estimated accurately under the proposed model, whereas the approach based on the simple RM ignoring missingness may yield biased estimates when missingness depends on the latent variable to be measured (i.e. θ_2).

Table 2. Correlation parameter recovery

	Generating value of ρ				
	0.0	*0.2*	*0.4*	*0.6*	*0.8*
BIAS	-0.0017	-0.0058	-0.0086	-0.0076	-0.0033
RMSE	0.0175	0.0184	0.0192	0.0154	0.0149

Application to data on racial prejudices

In this section we describe the application of the RRM to data analyzed by Knott and Tzamourani (1997) through model (4). As a special case of the models given in (3), this model is the most similar to the RRM (5) because:

- θ_2 represents the latent trait that governs the distinction *C* versus *B*. Then, for both the models, θ_2 may be interpreted as the "ability";
- The probability of a nonresponse is explicitly allowed to depend on both the latent trait dimensions θ_1 and θ_2, for both the models.

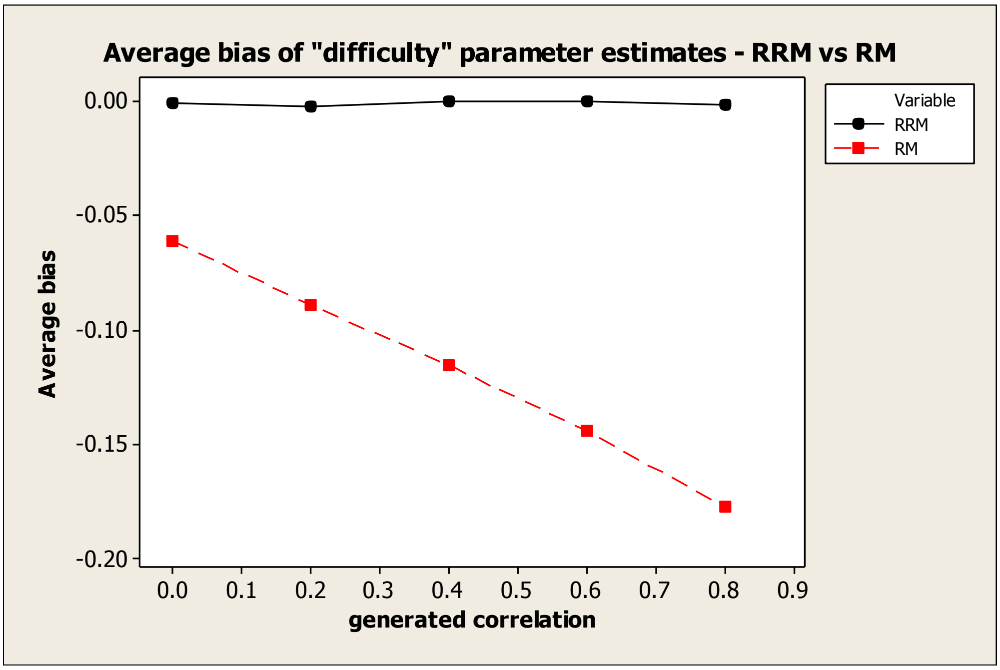

Figure 1. Bias of the "difficulty" parameter estimates, under RRM (δ_2) and RM (δ^{RM}): relationship with the correlation parameter ρ.

Therefore, for comparative purposes, we decided to calibrate the same dataset using the RRM. The dataset considered by Knott and Tzamourani (1997) consists of the answers of n = 1408 "white" respondents to the following k = 4 items from the British Social Attitudes Survey 1991 (Brook, Prior & Taylor, 1992):

Item 1. Thinking of black people - that is, people whose families were originally from West Indies or Africa - who now live in Britain. Do you think there is a lot of prejudice against them in Britain nowadays, a little, or hardly any?

Item 2. Do you think most white people in Britain would mind or not mind if a suitably qualified person of Asian origin were appointed as their boss? If "would mind", a lot or a little?

Item 3. And you personally? Would you mind or not? If "would mind", a lot or a little?

Item 4. Do you think that most white people in Britain would mind or not mind if one of their close relatives were to marry a person of Asian origin? If "would mind", a lot or a little?

Table 3. Total scores and CML estimates for the item parameters of the RRM

	Item 1	Item 2	Item 3	Item 4
# of given answers	1374	1342	1377	1334
$\hat{\delta}_1$	-0.053	0.322	-1.074	0.805
	(0.229)	(0.189)	(0.257)	(0.184)
# of "1"	707	266	72	520
$\hat{\delta}_2$	-1.568	0.317	2.195	-0.944
	(0.079)	(0.094)	(0.158)	(0.078)

Note. Standard errors of the estimates are also provided in parentheses.

Knott and Tzamourani (1997) recoded the data as follows: "a lot" and "mind a lot" were coded as 1, "mind a little", "a little", "hardly any" and "not mind" were coded as 0, and "don't know" and "not answered" were coded as missing (c, with the notation introduced in this paper). In this way, θ_2 should measure an attitude towards "non-white" people, or racial prejudice, while θ_1 should measure, in a broad sense, the tendency to express an opinion. The primary goal of this case study is to apply the RRM to the items from the same dataset.

The CML item parameter estimates for the RRM, along with their standard errors, are given in Table 3. These estimates are obtained in Mathematica Version 7.0 (Wolfram Research, 2008; the corresponding code is detailed in the Appendix).

To evaluate the RRM adequacy, we could assume that the distribution of the observed frequencies, over the possible response patterns, follows a multinomial distribution. In this case traditional goodness-of-fit tests could be applied. Unfortunately, as underlined by Knott and Tzamourani (1997, p. 249) for these data, the number of possible response patterns ($3^4 = 81$) is substantial; thus, the expected frequencies will tend to be very small for certain patterns, such that the usual χ^2-approximations are not valid (see, e.g., Tjur, 1982). For Rasch models, Andersen (1973) suggested an alternative goodness-of-fit test, based on the CML approach to the estimation. For this goodness-of-fit test, the observed counts for each response pattern are not required. Instead, the test is based on a partition of the total sample of subjects into a convenient number $M > 1$ of disjoint subgroups of subjects, say *score groups*, with n_m subjects in each ($m = 1,...,M$), "homogenous" with respect to the joint minimal sufficient statistics $(d_{v\cdot}, x_{v\cdot})$. The Andersen test is based on the fact that, under the Rasch paradigm, we should expect the overall CML-estimates, say $\hat{\boldsymbol{\delta}}$, obtained using the total group of subjects, to be approximately equal to those, say $\hat{\boldsymbol{\delta}}^{(m)}$, obtained by maximizing the log-likelihood function $l_C^{(m)}$ corresponding to the score group m, for all $m = 1,\ldots,M$. Then, in order to evaluate the fit, one can compare how close the CML-estimates $\hat{\boldsymbol{\delta}}^{(m)}$ are to the overall CML-estimates. This can be done graphically by plotting the score group estimates against the overall estimates, and numerically by considering a conditional likelihood ratio test based on the test statistic $S = -2\, l_C\left(\hat{\boldsymbol{\delta}}\right) + 2\sum_m l_C^{(m)}\left(\hat{\boldsymbol{\delta}}^{(m)}\right)$. Under regularity conditions, S is asymptotically χ^2 distributed, with $\nu = q\left(M-1\right)$degrees of freedom when each $n_m \to \infty$, where q is the number of unconstrained item parameters (Andersen, 1973). No precise theory exists for forming these score groups. Nevertheless, in order to avoid ill-conditioned datasets (i.e. datasets for which CML estimates do not exist; see Fischer, 1981), a possible choice is given by $M = 2$ groups according to what detailed in Table 4. Figure 2 shows the score group CML-estimates plotted against the overall CML-estimates. The benchmark line is also superimposed in Figure 2 to facilitate comparisons with respect to the optimal situation; it leads us to conclude that the main structure is good. This graphical result is corroborated by the observed test statistic $s = 5.576$ which, on $v = 6$ degrees of freedom, results to be no statistically significant (p-value equal to 0.47). In conclusion, with a more parsimonious RRM, we have obtained an adequate fit to these data without recurring to the 20-parameter model of Knott and Tzamourani (1997).

We also illustrate the MML approach by considering the output provided by ConQuest. Table 5 shows the item parameter estimates and corresponding fit information. From Table 6 we see that the item parameter estimates are very similar under CML and MML approaches (by setting to zero the mean for both the $\boldsymbol{\delta}_1$ and the $\boldsymbol{\delta}_2$ parameters, to facilitate comparisons).

Table 4. Contingency table for number of given answers and number of answers in category coded as "1"

# of "1" / # of given answers	0	1	2	3	4	Total
0	4					4
1	8	6				14
2	13	9	3			25
3	44	39	11	3		97
4	379	485	249	129	26	1268
Total	448	539	263	132	26	1408

Note. Two score groups are individuated for the Andersen conditional likelihood ratio test: score group G1 (composed by $n_1 = 87$ subjects) in light gray and score group G2 (composed by $n_2 = 1321$ subjects) in dark gray.

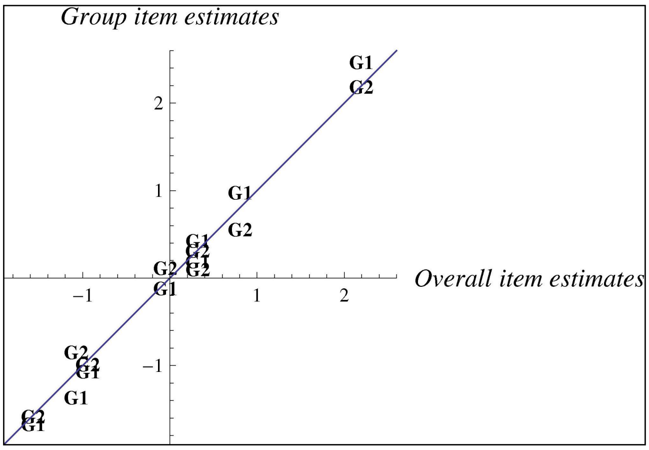

Figure 2. Andersen test. Within-score group CML-estimates against overall CML-estimates of δ_1 and δ_2 (G1 = score group 1; G2 = score group 2).

Table 5. MML item parameter estimates for the RRM

(a) δ_1 - summary informations

Item	$\hat{\delta}_1$	MNSQ	CI	T
1	-5.368 (0.20)	1.28	(0.68,1.32)	1.6
2	-4.984 (0.15)	0.99	(0.79,1.21)	0.0
3	-6.274 (0.21)	0.99	(0.67,1.33)	0.0
4	-4.486 (0.15)	0.97	(0.81,1.19)	-0.3

(b) δ_2 - summary informations

Item	$\hat{\delta}_2$	MNSQ	CI	T
1	-0.074 (0.06)	1.10	(0.95,1.05)	3.8
2	1.803 (0.08)	0.92	(0.92,1.08)	-1.9
3	3.562 (0.13)	0.98	(0.80,1.20)	-0.2
4	0.595 (0.06)	1.00	(0.95,1.05)	0.1

Note. Item parameter estimates, as given by the ConQuest output. In this calibration, the mean of the latent variables is set to zero.

Table 6. Comparison of CML-estimates and MML-estimates of item parameters for the RRM

(a) δ_1 estimates

Item	CML	MML
1	-0.053	-0.090
2	0.322	0.294
3	-1.074	-0.996
4	0.805	0.792

(b) δ_2 estimates

Item	CML	MML
1	-1.568	-1.545
2	0.317	0.332
3	2.195	2.091
4	-0.944	-0.876

Note. To facilitate comparisons, the mean of the item parameters on each dimension is constrained to be zero, for both the estimation methods.

As we can see from Table 5(a) and Table 5(b), ConQuest produces the mean squared (MNSQ) fit statistic for every estimated parameter, which is based on a standardized comparison between expected and observed scores. When the model fits the data, the MNSQ statistics have a unitary expected value. These statistics are transformed by ConQuest to approximate normal deviates, denoted by *T*. The software also provides a 95% confidence interval (CI) for the expected value of the MNSQ. If the MNSQ fit statistic lies outside the CI, then the corresponding *T* statistic will have an absolute value that roughly exceeds 2 (see Wu et al., 2007, p. 23). It is apparent that all the item fit statistics are good, with the exception of the parameter δ_{21}. To obtain a simple measure about the global fit of the RRM we have computed, for each item, the observed and expected frequencies for each response option: *A* (nonresponse); *B* (response in category coded as "0"); *C* (response in category coded as "1"). As we can see from Table 7, the fit seems to be good.

Table 7. Observed and expected frequencies. A comparison between model (3) (Knott & Tzamourani, 1997) and the RRM (5).

(a) Observed				(b) Expected – Model (3)				(c) Expected – RRM			
Item	*A*	*B*	*C*	Item	*A*	*B*	*C*	Item	*A*	*B*	*C*
1	34	667	707	1	33.95	667.03	707.02	1	33.56	665.84	708.59
2	66	1076	266	2	65.97	1080.56	261.47	2	64.59	1075.75	267.66
3	31	1305	72	3	30.62	1304.87	72.51	3	29.57	1306.55	71.88
4	74	814	520	4	74.84	813.46	519.71	4	73.08	813.11	521.80

Note. Response options: A = nonresponse; B = response in category coded as “0”; C = response in category coded as “1”.

Interestingly, there is agreement with several of the points made by Knott and Tzamourani (1997, p. 248). (Note that in their model the coefficients $\boldsymbol{a}_{1i}$ and $\boldsymbol{a}_{2i}$ play the role of “factor loadings”). In particular:

- they find that the discrimination parameter of item 1 (a_{221}) "is close to zero indicating that this item does not "load" as high on the underlying factor as the other items do. The first item does not seem to measure racial prejudice. This can be explained by the fact that the question asked for item 1 is too abstract." Indeed, we find evidence of the misfit of this item to the RRM also, which is reflected by an under-discriminating item (large positive value of T);
- they find that "item 2 is the most discriminating item (highest value of discrimination parameter)." Indeed, for item 2, we found the largest negative value of T;
- they find that items 2, 3, and 4 correspond to "low probability for a positive response for the median individual, i.e., the median individual has a small probability of saying that people would have something against a "non-white" person." Indeed, the δ_2 s are all positive for these items.

Conclusions

In this paper we introduce the RRM, a two-dimensional generalization of the simple RM for the treatment of binary data in presence of nonignorable nonresponses. The model is suited for intentional omissions – i.e. situations in which respondents may decide, for whatever reason, to skip the item. The model depends on a vector-valued person variable for two latent dimensions: θ_1 related to the response propensity, and θ_2 that may be interpreted (depending on the parameterization) as the usual "ability" parameter. In its simplest form, the RRM combines 1) an RM for the response variable and 2) an RM for the response indicator variable, but the model can also be reparameterized via linear combinations of both the latent traits θ_1 and θ_2 –with different interpretations of the parameters.

In particular, the model presented can cope with situations where the issues under study are sensitive, whether a respondent is embarrassed or not willing to reveal his opinion, depending upon circumstances (e.g., within the context of polling data, see Rubin, Stern, & Vehovar, 1995, and also Smith, Skinner, & Clarke, 1999). As a member of the exponential family, the RRM allows the use of conditional inference, but the MML approach can also be used to estimate the model parameters. More specifically, under the MML approach the RRM can be seen as an instance of a MRCMLM, with a within-item multidimensionality, because each item in the test is designed to measure both the dimensions θ_1 and θ_2; then, the computer program ConQuest can be directly adopted for fitting the model to data as well. The results of a simulation indicate that item parameters can be estimated accurately under the proposed model.

REFERENCES

Adams, R., Wilson, M., & Wang, W. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*, 1-23.

Albanese, M., & Knott, M. (1992). *TWOMISS: A computer program for fitting a one-or two-factor logit-probit latent variable model to binary data when observations may be missing*. Technical report, London School of Economics and Political Science, London.

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

Bartholomew, D.J., de Menezes, L.M., and Tzamourani, P. (1997). Latent trait class of models applied to survey data. In J. Rost & R. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences* (pp. 219-232). Munster, Germany: Waxmann.

Birnbaum, A. (1968). Some latent trait models. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores*, (pp. 397-479). Reading, MA: Addison-Wesley.

Bolt, D. M., & Johnson, T. R. (2009). Applications of a MIRT model to self-report measures: Addressing score bias and DIF due to individual differences in response style. *Applied Psychological Measurement*, 33, 335-352.

Brook, L., Prior, G., & Taylor, B. (1992). *British social attitudes 1991 survey*. Technical report, Social & Community Planning Research.

Eggen, T., & Verhelst, N. (2011). Item calibration in incomplete testing designs. *Psicológica, 32*, 107-132.

Fischer, G. H. (1981). On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. *Psychometrika, 46*, 59-77.

Glas, C. A. W., & Pimentel, J. L. (2006). *Modeling nonignorable missing data process in item calibration*. Technical report, LSAC.

Holland, P. W. (1990). On the sampling theory foundation of item response theory models. *Psychometrika*, *55*, 577-601.

Holman, R., Lindeboom, R., Vermeulen, R., Glas, C.A.W., and de Haan, R.J. (2001). The Amsterdam Linear Disability Score (ALDS) project. The calibration of an item bank to measure functional status using item response theory, *Quality of Life Newsletter* 27, 4–5.

Holman, R., & Glas, C. A. W. (2005). Modeling non-ignorable missing-data mechanism with item response theory models. *British Journal of Mathematical and Statistical Psychology, 58*, 1-17.

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

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* (pp. 244-252). Munster, Germany: Waxmann.

Knott, M., Albanese, M., & Galbraith, J. (1990). Scoring attitudes to abortion. *The Statistician, 40*, 217-223.

Lehmann, E. (1983). *Theory of point estimation*. New York, NY: Wiley.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. New York, NY: Wiley.

Little, R. J. A., & Schenker, N. (1995). Missing Data. In: G. Arminger, C. C. Clogg & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 39-75). New York, NY: Plenum.

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

Mislevy, R. J., & Wu, P. K. (1996). *Missing responses and IRT ability estimation: omits, choice, time limits, adaptive testing*. ETS Research Report RR-96-30-ONR, Princeton, NJ: Educational Testing Service.

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 (Statistics in Society)*, *163*, 445-459.

Moustaki, I., & O'Muircheartaigh, C. (2000). A one dimension latent trait model to infer attitude from nonresponse for nominal data. *Statistica, 60*, 259-276.

Moustaki, I., & O'Muircheartaigh, C. (2002). Locating 'Don't Know', 'No Answer' and Middle Alternatives on an Attitude Scale: A latent Variable Approach. In G.A.

Marcoulides and I. Moustaki (Eds.), *Latent Variable and Latent Structure Models* (pp. 15-40). London: Lawrence Erlbaum Associates.

O'Muircheartaigh, C., & Moustaki, I. (1996). Item non-response in attitude scales: a latent variable approach. *Proceedings of the American Statistical Association, Section of Survey Research Methods*, (pp. 938-943).

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.

Pfanzagl, J. (1994) . On item parameter estimation in certain latent trait models. In G H. Fischer & D. Laming (Eds.), *Contributions to Mathematical Psychology, Psychometrics and Methodology* (pp. 249-263). New York, NY: Springer Verlag.

Pimentel, J. L. (2005). *Item Response Theory modeling with nonignorable missing data*. Ph.D. thesis, University of Twente, The Netherlands.

R Development Core Team (2008). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. The Danish Institute of Educational Research, Copenhagen.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In: Neyman, J. (Ed.), *Proceedings of the IV. Berkeley Symposium on Mathematical Statistics and Probability* (pp. 321-333), Berkeley, CA: Berkeley University Press.

Reckase, M.D. (2009). *Multidimensional Item Response Theory*. New York, NY: Springer.

Rose, N., von Davier, M., & Xu, X. (2010). *Modeling nonignorable missing data with item response theory (IRT)*. Technical report, ETS, Princeton, New Jersey.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*, 581-592.

Rubin, D., Stern, H., & Vehovar, V. (1995). Handling "don't know" survey responses: the case of the Slovenian plebiscite. *Journal of the American Statistical Association, 90*, 822-828.

Schafer, J.L . (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman and Hall/CRC.

Smith, P. W. F., Skinner, C. J., & Clarke, P. S. (1999). Allowing for non-ignorable non-response in the analysis of voting intention data. *Journal of the Royal Statistical Society: series C (Applied Statistics), 48*, 563-577.

Thissen, D., Cai, L., & Bock, R. D. (2010). The nominal categories item response model. In M.L. Nering & R. Ostini (Eds.), *Handbook of Polytomous Item Response Theory Models: Development and applications* (pp. 43-75). New York, NY: Taylor & Francis.

Tjur, T. (1982). A connection between Rasch's item analysis model and a multiplicative Poisson model. *Scandinavian Journal of Statistics, 9*, 23-30.

Wolfram Research, I. (2008). *Mathematica, Version 7.0*. Champaign, Illinois.

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). *ACER ConQuest: Generalised Item Response Modeling Software - Version 2.0,* ACER Press edition.

APPENDIX

Mathematica code used for computing CML estimates of item parameters in the RRM

```
(* Import data matrix and preliminary quantities *)

X   = Import["X.dat"];
Dim = Dimensions[X];
n   = Dim[[1]];  (* number of subjects *)
k   = Dim[[2]];  (* number of items *)
c   = 9;         (* missing value *)
delta = Array[Subscript[\[Delta], ##] &, {2, k}];

(* Response indicator *)

d = X;
For[v = 1, v <= n, v++,
       For[i = 1, i <= k, i++,
              d[[v, i]] = If[X[[v, i]] == c, 0, 1]
       ]
]

(* "1"-answer indicator *)

Xd = X;
For[v = 1, v <= n, v++,
       For[i = 1, i <= k, i++,
              Xd[[v, i]] = If[X[[v, i]] == 1, 1, 0]
       ]
]

(* Number of given responses *)

tot01 = ConstantArray[0, n];
For[v = 1, v <= n, v++,
       tot01[[v]] = Count[d[[v]], 1]
];

(* Number of "1" for subject *)

tot1 = ConstantArray[0, n];
For[v = 1, v <= n, v++,
       tot1[[v]] = Count[Xd[[v]], 1]
];

(* Single conditional likelihood numerator *)

num = ConstantArray[0, n];
For[v = 1, v <= n, v++,
       num[[v]] = Exp[-(delta[[1]].d[[v]]+delta[[2]].Xd[[v]])]
]
```

```
(* All possible 3^k tuples in the denominator *)

Tup = Tuples[{c, 0, 1}, k];
For[given = 0, given <= k, given++,
       For[pos = 0, pos <= given, pos++,
              Subscript[a, given, pos]={}
       ]
]
For[given = 0, given <= k, given++,
       For[pos = 0, pos <= given, pos++,
              For[i = 1, i <= 3^k, i++,
                     If[Count[Tup[[i]],c]==k-given &&
Count[Tup[[i]],1]==pos,

       Subscript[a,given,pos]=AppendTo[Subscript[a,given,pos],Tup[[i
]]],
                            ]
              ];
       Print[Subscript[a, given, pos]//MatrixForm]
       ]
]

(* Single denominator *)

den = ConstantArray[0, n];
For[v = 1, v <= n, v++,
       den[[v]] = Sum[
              Exp[-delta[[1]].Boole[(# != c) &
/@Subscript[a,tot01[[v]],tot1[[v]]][[u]]]]*
              Exp[-delta[[2]].Boole[(# == 1) &
/@Subscript[a,tot01[[v]],tot1[[v]]][[u]]]],
                     {u,1,Length[Subscript[a,tot01[[v]],tot1[[v]]]]}
       ]
]

(* Conditional likelihood function *)

LC = Product[num[[v]]/den[[v]],{v,1,n}];
res =
FindMaximum[Log[LC],Union[delta[[1]],delta[[2]]],WorkingPrecision -
>20]

(* Identifiability constraints *)

deltahat = Array[0, {2, k}];
deltahat[[1]]=res[[2,1;;k,2]]-Mean[res[[2,1;;k,2]]];
deltahat[[2]]=res[[2,k+1;;2*k,2]]-Mean[res[[2,k+1;;2*k,2]]];
deltahat//MatrixForm
```

Note. This code is written in Mathematica Version 7.0 (Wolfram Research, 2008). It has been used to obtain the CML-estimates of the item parameters δ_1 and δ_2 in the application to racial prejudice data and can be successfully adopted for other short tests (approximately up to k=10 items). The observed data are supposed to be collected in a ($n \times k$)-matrix saved (in the working directory of Mathematica) as a plain text file format, named `X.dat`, with tab-separated values coded as c, 0, 1. The code returns a ($2 \times k$)-matrix, named `deltahat`, having the CML-estimates of δ_1' and δ_2' in the first and second row, respectively.

(Manuscript received: 8 December 2011; accepted: 8 March 2012)