T E C H N I C A L
R E P O R T

0249

A RELATION BETWEEN A BETWEEN-ITEM MULTIDIMENSIONAL IRT MODEL AND THE MIXTURE RASCH

RIJMEN, F. and P. DE BOECK

IAP STATISTICS NETWORK

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

A Relation between a Between-item Multidimensional IRT Model and the Mixture Rasch Model

Frank Rijmen and Paul De Boeck

University of Leuven

Frank Rijmen, Department of Psychology; Paul De Boeck, Department of Psychology. Frank Rijmen was supported by the Fund for Scientific Research Flanders (FWO). This research is also funded by the GOA/2000/02 granted from the KU Leuven.

We would like to thank Kristof Vansteelandt for providing the data of the study on verbal aggression.

Correspondence should be addressed to Frank Rijmen, Department of Psychology, Tiensestraat 102, B-3000 Leuven, Belgium (e-mail: frank.Rijmen@psy.kuleuven.ac.be; tel.: (32) 16-325985; fax.: (32) 16-325916).

A Relation between a Between-item Multidimensional IRT Model and the Mixture Rasch Model

Two generalizations of the Rasch model are compared: the between-item multidimensional model (Adams, Wilson, & Wang, 1997), and the mixture Rasch model (Mislevy & Verhelst, 1990; Rost, 1990). It is shown that the between-item multidimensional model is formally equivalent with a continuous mixture of Rasch models for which, within each class of the mixture, the item parameters are equal to the item parameters of the multidimensional model up to a shift parameter that is specific for the dimension an item belongs to in the multidimensional model. In a simulation study, the relation between both types of models also holds when the number of classes of the mixture is as small as two. The relation is illustrated with a study on verbal aggression.

Key words: Rasch model, multidimensional IRT models, mixture Rasch model, Saltus model.

The Rasch model (Rasch, 1960) is a simple and widespread IRT model developed for binary test data. The model takes the dependency between the responses of the same participant into account by assuming a latent person variable. The latter represents the 'ability', 'scale', or 'trait' the test is measuring. Under the Rasch model, the probability that a participant n scores one on an item i ($Y_{ni} = 1$) is given as

$$\Pr\left(Y_{ni} = 1 \middle| \theta_n, \beta_i\right) = \frac{\exp(\theta_n + \beta_i)}{1 + \exp(\theta_n + \beta_i)} \quad . \tag{1}$$

β_i is an item parameter representing the position of item i on the latent scale, and θ_n represents the position of participant n on the same scale. The θ_n can be considered either as fixed parameters, or alternatively, as random variables with a common distribution defined over the population of participants, leading to respectively a functional and structural model formulation (de Leeuw & Verhelst, 1986). In this paper, we take a structural model perspective. The distribution of the latent variables is left unspecified in the derivation of our main result, however. Consequently, this result is free from distributional assumptions about the latent variables. In a separate section, we consider the special case of a two-dimensional model with a bivariate normal distribution for the latent variables.

For the structural Rasch model, the marginal probability of a response pattern $\boldsymbol{y}$ is

$$\Pr\left(\boldsymbol{y} \middle| \boldsymbol{\beta}\right) = \int_{\theta} \left\{ \prod_{i=1}^{I} \frac{\exp[y_i(\theta + \beta_i)]}{1 + \exp(\theta + \beta_i)} \right\} dF(\theta), \tag{2}$$

where I is the number of items, and $\boldsymbol{\beta}$ the vector of item parameters. Here and further on, the person subscript n is suppressed from notation because we assume a common distribution function F for all θ_n, $n = 1,\ldots, N$. A constraint is needed to identify the model because we can always add a constant to $\boldsymbol{\beta}$, and subtract the same constant from θ without altering the likelihood of the model. We constrain $E(\theta)$ to 0, but alternatively, one can can also estimate $E(\theta)$ from the data, and put some constraint on the item parameters.

According to the Rasch model, all dependencies between the responses of a participant are accounted for by the unidimensional latent variable θ. This is often referred to as the assumption of local stochastic independency: conditional on the item and person parameters, all the responses are assumed to be independent Bernoulli observations, with the probability of ‘success” given in Equation 1. The characterization of the latent person space in terms of a single unidimensional latent variable means that all items are located on the same scale, and that this scale is the same for all persons.

In this paper, we focus on two generalizations of the Rasch model that relax the assumption of the Rasch model of a single underlying unidimensional trait.

The first extension is an extension towards multidimensionality. We relax the assumption that all items are located on the same scale, and instead assume that a test consists of K subscales or subgroups of items that each can be modeled according to a Rasch model. Let $\boldsymbol{r}_i$ be a K-dimensional vector that contains exactly one nonzero element, equal to one, indicating the scale the item belongs to. The model is then defined as

$$\Pr(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{R})=\int_{\boldsymbol{\theta}}\left\{\prod_{i=1}^{I}\frac{\exp[y_i(\boldsymbol{r}_i'\boldsymbol{\theta}+\beta_i)]}{1+\exp(\boldsymbol{r}_i'\boldsymbol{\theta}+\beta_i)}\right\}dF(\boldsymbol{\theta}) \quad , \tag{3}$$

where $\boldsymbol{\theta}$ is now a K-dimensional vector representing the positions on K continuous latent traits, and $\boldsymbol{R}$ is a $I \times K$ matrix with as i^{th} row $\boldsymbol{r}_i$'. Analogous to the Rasch model, we use the restriction $E(\boldsymbol{\theta}) = \mathbf{0}$ to identify the model. Since each item measures only one dimension, the model is called a between-item multidimensional model (Adams, Wilson, & Wang, 1997). The advantages of using the multidimensional model instead of analyzing the K scales separately with a Rasch model is that the test structure is explicitly taken into account, that estimates for the correlation between the latent dimensions are provided, and that more accurate parameter estimates are obtained by relying on the relationship between the dimensions (Adams et al., 1997). The model presented in Equation 3 is a specific instance of the multidimensional random coefficients multinomial logit model (Adams et al., 1997).

In the second extension of the Rasch model, the mixture Rasch model (Mislevy & Verhelst, 1990; Rost, 1990), we relax the assumption of the Rasch model that the scale a test is measuring is the same for all persons. Instead, it is assumed that the total group of persons consists of T qualitatively different subgroups. Since group-membership is unobserved, the subgroups correspond to latent classes. Within each subgroup, the unidimensional Rasch model is assumed to hold. That is, within a latent class, one latent continuous variable explains all dependencies between responses of the same participant, and items are located on the same scale. However, because the item parameters are class-specific, the particular scale a test is measuring differs across classes. According to the mixture Rasch model, the marginal probability of a response pattern $\boldsymbol{y}$ is:

$$\Pr\left(\boldsymbol{y}\middle|\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_T\right)=\sum_t^T \pi_t \int_\theta \left\{\prod_{i=1}^I \frac{\exp\left[y_i\left(\theta+\beta_{it}\right)\right]}{1+\exp\left(\theta+\beta_{it}\right)}\right\} dF_t\left(\theta\right), \tag{4}$$

where π_t is the probability of belonging to latent class t, $\boldsymbol{\beta}_t$ is the vector of item parameters for latent class t, and F_t the distribution function of θ for latent class t. Analogous to the Rasch model and the multidimensional model, we set $E(\theta) = 0$ for each class to identify the model.

In this paper, we explore a formal relation between both extensions of the Rasch model. Specifically, we are interested in how a mixture Rasch model solution will look like if the true model is a multidimensional model.

In investigating the relation between multidimensional IRT model and the mixture Rasch model of Rost (1990), we follow Reise and Gomel (1995), who compared the mixture Rasch model and the full information item factor model (Bock, Gibbons, & Muraki, 1988) with respect to their conceptual underpinnings and empirical fit. The full information item factor model differs from the multidimensional model presented in Equation 3 in that also discrimination parameters are included (that is, the elements of the $\boldsymbol{r}_i$ vectors are estimated and not specified in advance in the particular way we did), and that a probit instead of a logit link is used. Reise and Gomel (1995) found that the item factor model was somewhat more parsimonious than the mixture Rasch model, but also pointed out several research contexts where the mixture Rasch model is more appropriate. Our approach differs from Reise and Gomel (1995) in that we focus on the formal relation between a between-item multidimensional model and a mixture Rasch model, rather than comparing the empirical fit to a particular dataset as they did. Nevertheless, we also discuss a real data example at the end of the paper.

The Between-Item Multidimensional Model as a Continuous Mixture of Rasch Models

We start by rewriting the between-item multidimensional model of Equation 3 as

$$\Pr\left(\boldsymbol{y}\middle|\boldsymbol{\beta},\boldsymbol{R}\right)=\int_{\xi}\left\{\prod_{i=1}^{I}\frac{\exp\left[y_i\left(\boldsymbol{r}_i'\boldsymbol{S}\xi+\beta_i\right)\right]}{1+\exp\left(\boldsymbol{r}_i'\boldsymbol{S}\xi+\beta_i\right)}\right\}dQ\left(\xi\right), \tag{5}$$

where $\boldsymbol{\xi}=\boldsymbol{T\theta}$, with $\boldsymbol{T}$ a nonsingular transformation matrix; $\boldsymbol{S}=\boldsymbol{T}^{-1}$; and Q the distribution function of $\boldsymbol{\xi}$. Conditioning the k^{th} (linearly transformed) latent variable ξ_k on the vector $\boldsymbol{\xi}_{(k)}$ of K - 1 other latent variables, we obtain

$$\Pr\left(\boldsymbol{y}\middle|\boldsymbol{\beta},\boldsymbol{R}\right)=\int_{\xi_{(k)}}\int_{\xi_k}\left\{\prod_{i=1}^{I}\frac{\exp\left[y_i\left(\boldsymbol{r}_i'\boldsymbol{S}\xi+\beta_i\right)\right]}{1+\exp\left(\boldsymbol{r}_i'\boldsymbol{S}\xi+\beta_i\right)}\right\}dQ_k\left(\xi_k\middle|\xi_{(k)}\right)dQ_{(k)}\left(\xi_{(k)}\right). \tag{6}$$

If we can choose $\boldsymbol{T}$ such that the inner integral corresponds with a Rasch model as formulated in Equation 2, the between-item multidimensional model of Equation 3 can be rewritten as a continuous mixture of Rasch models, where each set of values for $\boldsymbol{\xi}_{(k)}$ corresponds to a class. That there is always such a reformulation is shown in the following.

Conditioning on $\boldsymbol{\xi}_{(k)} = \boldsymbol{a}_{(k)}$, the probability of a response pattern $\boldsymbol{y}$ is:

$$\Pr\left(\boldsymbol{y} \mid \xi_{(k)} = \boldsymbol{a}_{(k)}, \boldsymbol{\beta}, \boldsymbol{R}\right) = \int_{\xi_k} \prod_{i=1}^{I} \frac{\exp\left[y_i \left(\sum_{l=1}^{K} r_{il}\left(s_{lk}\xi_k + \boldsymbol{s}_l^{(k)}\boldsymbol{a}_{(k)}\right) + \beta_i\right)\right]}{1 + \exp\left(\sum_{l=1}^{K} r_{il}\left(s_{lk}\xi_k + \boldsymbol{s}_l^{(k)}\boldsymbol{a}_{(k)}\right) + \beta_i\right)} dQ_k\left(\xi_k \mid \xi_{(k)} = \boldsymbol{a}_{(k)}\right), \quad (7)$$

where $\boldsymbol{s}_l^{(k)}$ is row l, l = 1,…, K, of $\boldsymbol{S}$, without element s_{lk}. Since each item belongs to precisely one of the original scales in the between-item multidimensional model we are considering, Equation 7 can be simplified to

$$\Pr\left(\boldsymbol{y} \mid \xi_{(k)} = \boldsymbol{a}_{(k)}, \boldsymbol{\beta}, \boldsymbol{R}\right) = \int_{\xi_k} \prod_{i=1}^{I} \frac{\exp\left[y_i\left(s_{l_i k}\xi_k + \boldsymbol{s}_{l_i}^{(k)}\boldsymbol{a}_{(k)} + \beta_i\right)\right]}{1 + \exp\left(s_{l_i k}\xi_k + \boldsymbol{s}_{l_i}^{(k)}\boldsymbol{a}_{(k)} + \beta_i\right)} dQ_k\left(\xi_k \mid \xi_{(k)} = \boldsymbol{a}_{(k)}\right), \quad (8)$$

where the index l_i refers to the original scale l item i belongs to (for the items belonging to the first original scale $s_{l_i k} = s_{1k}$ and $\boldsymbol{s}_{l_i}^{(k)} = \boldsymbol{s}_1^{(k)}$, for the items belonging to the second original scale $s_{l_i k} = s_{2k}$ and $\boldsymbol{s}_{l_i}^{(k)} = \boldsymbol{s}_2^{(k)}$, and so on). Equation 8 corresponds to the marginal formulation of a Rasch model (see Equation 2) if the discrimination parameters for all items equal one, that is if s_{lk} equals one for all l, l = 1,…, K. Hence, in order for the inner integral of Equation 6 to correspond with the marginal formulation of a Rasch model, the k^{th} column of $\boldsymbol{S}$ should consist of ones only. The latter means that the k^{th} basis vector of the natural basis of the $\boldsymbol{\xi}$ space has coordinates (1, 1, …, 1) in the original $\boldsymbol{\theta}$

space. By consequence, ξ_k forms the superdiagonal in the $\boldsymbol{\theta}$ space, and $\boldsymbol{\xi}_{(k)}$ spans the (K – 1)-dimensional subspace orthogonal to ξ_k . Hence, each latent class corresponds to a straight line parallel to the superdiagonal, through the point with coordinates $\xi_k = 0$, $\boldsymbol{\xi}_{(k)} = \boldsymbol{a}_{(k)}$, lying in the subspace orthogonal to the superdiagonal. For a geometrical illustration for a two-dimensional model, see Figure 1. The figure represents the case where one conditions ξ_1 on $\boldsymbol{\xi}_{(1)} = \xi_2$. ξ_1 is the superdiagonal of the $\boldsymbol{\theta}$ space, that is, the straight line trough the origin with slope one. Each class is defined by a particular value a of ξ_2, which spans the subspace orthogonal to ξ_1, and hence is the straight line trough the origin with slope minus one.

It follows from Equation 8 that an item parameter *within a class* is the sum of (a) the item parameter β_i of the between-item multidimensional model (b) a term that is common for all the items belonging to the same original dimension l and that depends on the value of $\boldsymbol{a}_{(k)}$ one is conditioning upon, $\boldsymbol{s}_l^{(k)}\boldsymbol{a}_{(k)}$, and (c) a term that is the same for all items, $E\left(\xi_k \middle| \boldsymbol{\xi}_{(k)} = \boldsymbol{a}_{(k)}\right)$. The last term originates from our identification restriction that the mean of the latent variable within each class, $E\left(\xi_k \middle| \boldsymbol{\xi}_{(k)} = \boldsymbol{a}_{(k)}\right)$, is equal to zero. This is accomplished by subtracting $E\left(\xi_k \middle| \boldsymbol{\xi}_{(k)} = \boldsymbol{a}_{(k)}\right)$ from ξ_k and adding it to each item parameter. Hence, within each class, the item parameters are equal to the item parameters of the multidimensional model up to a shift parameter that is specific for the dimension an item belongs to in the multidimensional model. The shift parameter equals $\boldsymbol{s}_l^{(k)}\boldsymbol{a}_{(k)} + E\left(\xi_k \middle| \boldsymbol{\xi}_{(k)} = \boldsymbol{a}_{(k)}\right)$.

INSERT FIGURE 1 ABOUT HERE

It also follows from the equality that the item parameters for items belonging to the same dimension in the between-item multidimensional model are equal across classes in the continuous mixture of Rasch models up to a class-specific shift parameter: $\beta_{it} = \beta_{i1} + \eta_{kt}$; $t = 2, \ldots, T$. In the literature, a mixture Rasch model with this property is described as a Saltus model (Mislevy & Wilson, 1996; Wilson, 1989).

Finally, since $\boldsymbol{T}$ (and $\boldsymbol{S}$) is not completely specified, it may seem that there is a whole family of continuous mixtures of Rasch models that correspond to a particular two-dimensional model, each characterized by its own class-specific shift parameters that are a function of $\boldsymbol{T}$. However, as was outlined above, for all admissible $\boldsymbol{T}$, the $\boldsymbol{\theta}$ space is partitioned in the same way: Each class corresponds to a straight line parallel with the superdiagonal, and, by integrating over $\boldsymbol{\xi}_{(k)}$ (see Equation 6), there is a class corresponding to each point in the subspace orthogonal to the superdiagonal. Hence, regardless of the specific $\boldsymbol{T}$ that is chosen, the parameters of the continuous mixture of Rasch models should be a function of the two-dimensional item parameters and of the distribution function of $\boldsymbol{\theta}$ only. In the following section, we will also show algebraically that the resulting continuous mixture of Rasch models is independent of $\boldsymbol{T}$ for the special case of a two-dimensional model with a bivariate normal distribution for the latent variables.

Special Case: A Two-dimensional Model with a Bivariate Normal Distribution for the Latent Variables

In the previous section, we showed that the between-item multidimensional model of Equation 3 can be reformulated as a continuous mixture of Rasch models without making any assumptions on the distribution of the latent variables. The general result is now applied to the case of a multivariate normal distribution. The latter offers the advantage that more precise expressions can be derived for the parameters of the continuous mixture of Rasch models. The expressions are derived for a two-dimensional model. For more than two dimensions, the derivations are analogous, although they become rather cumbersome.

Applying the general results from the previous section to the present case, the between-item two-dimensional model can be presented as a mixture Rasch model with an infinite number of classes, where each class is formed by a straight line with slope one, see also Figure 1. Specifically, $\boldsymbol{S}$ has the following form:

$$\boldsymbol{S} = \begin{bmatrix} 1 & s_{12} \\ 1 & s_{22} \end{bmatrix},$$

and hence $\boldsymbol{T} = \boldsymbol{S}^{-1} = \begin{bmatrix} \frac{s_{22}}{s_{22} - s_{12}} & \frac{-s_{12}}{s_{22} - s_{12}} \\ \frac{-1}{s_{22} - s_{12}} & \frac{1}{s_{22} - s_{12}} \end{bmatrix}$.

Each class is a line parallel with ξ_1, and thus of the form $\theta_1 - \theta_2 = b$.

Conditioning on $\theta_1 - \theta_2 = b$ corresponds to conditioning on $\xi_2 = \frac{-b}{s_{22} - s_{12}} = a$. Since $\boldsymbol{\theta}$ is bivariate normal, it follows that $\xi_1 | \theta_1 - \theta_2 = b$ is also normal, with the following mean and variance:

$$E\left(\xi_1 | \theta_1 - \theta_2 = b\right) = E\left(\xi_1\right) + \rho_{\xi_1\xi_2} \frac{\sigma_{\xi_1}}{\sigma_{\xi_2}}\left[a - E\left(\xi_2\right)\right] = -b\rho_{\xi_1\xi_2} \frac{\sigma_{\xi_1}}{\sigma_{\theta_1-\theta_2}},$$

$$\sigma^2_{\xi_1|\theta_1-\theta_2=b} = \sigma^2_{\xi_1}\left(1 - \rho^2_{\xi_1\xi_2}\right) = \frac{\sigma^2_{\theta_1}\sigma^2_{\theta_2} - \sigma^2_{\theta_1\theta_2}}{\sigma^2_{\theta_1-\theta_2}}, \quad (9)$$

where $\rho_{\xi_1\xi_2}$ is the correlation between ξ_1 and ξ_2, and $\sigma_{\theta_1\theta_2}$ the covariance between θ_1 and θ_2. From these expressions, it is clear that the variances are the same for all classes, and independent of the actual choice for any of the admissible $\boldsymbol{T}$'s.

For the item parameters within each class, applying the general result from the previous section, we can derive the following expressions for respectively the items belonging to the first and to the second dimension:

$$\beta_{itd_1} = \beta_i + s_{12}a - E\left(\xi_1 | \xi_2 = a\right)$$

$$= \beta_i - \frac{bs_{12}}{s_{22} - s_{12}} - b\rho_{\xi_1\xi_2}\frac{\sigma_{\xi_1}}{\sigma_{\theta_1-\theta_2}}$$

$$= \beta_i + b\frac{\sigma^2_{\theta_1} - \sigma_{\theta_1\theta_2}}{\sigma^2_{\theta_1-\theta_2}}.$$

$$\beta_{itd_2} = \beta_i + s_{22}a - E\left(\xi_1 \middle| \xi_2 = a\right)$$

$$= \beta_i - \frac{bs_{22}}{s_{22} - s_{12}} - b\rho_{\xi_1\xi_2}\frac{\sigma_{\xi_1}}{\sigma_{\theta_1-\theta_2}}$$

$$= \beta_i + b\frac{-\sigma_{\theta_2}^2 + \sigma_{\theta_1\theta_2}}{\sigma_{\theta_1-\theta_2}^2} \quad . \tag{10}$$

Again, the expressions are independent of the actual choice for any of the admissible ***T***'s.

Finite Number of Classes

In the above, the equivalence between a between-item multidimensional model and a mixture Rasch model with a particular structure was derived for the case with an infinite number of classes. We can expect the derivation also to hold approximately for a finite but sufficiently large number of classes, approximating the outer integral of Equation 6 by a weighted sum over a discrete grid with nodes φ_t , $t = 1,\ldots,T$, defined on $\boldsymbol{\xi}_{(k)}$.

The smaller the number of classes, the poorer the approximation of the outer integral in Equation 6 is expected to be. In the next section, we describe a simulation study for assessing to what amount the relation between a multidimensional model and a mixture Rasch model described above holds when the number of classes is only two. In the simulation, the data were generated under a two-dimensional model with a bivariate normal distribution for the latent variables. The more precise expressions for the parameters of a continuous mixture of Rasch models that is equivalent with such a model (given in Equation 9 and 10) can be used as a benchmark to assess an approximation with only two classes.

Simulation Study.

Data were generated according to a two-dimensional model with a bivariate normal distribution for the latent variables. The first 10 items belonged to the first dimension, the next 10 items to the second dimension. For both sets of items, the item parameters were –2, -1.5, -1, -.5, -.10, .10, .5, 1, 1.5, 2. Two factors were manipulated: (1) the correlation between the two latent variables, which could be 0, .25, .50, or .75, and (2) whether the variances of the two latent variables were equal or not: $\sigma^2_{\theta_1} = \sigma^2_{\theta_2} = 1$, or $\sigma^2_{\theta_1} = 1$ and $\sigma^2_{\theta_2} = 2$. Hence, there were eight conditions in total. For each condition, 100 datasets where generated, each with 1000 persons.

Assuming that the mixture Rasch model with two classes is a suitable approximation of a continuous mixture of Rasch models, the variances and item parameters were calculated with Equations 9 and 10. These expected parameter estimates are given in Table 1 and 2, respectively

INSERT TABLE 1 ABOUT HERE

INSERT TABLE 2 ABOUT HERE

The datasets were analysed with a mixture Rasch model with two classes. An EM algorithm was programmed by the first author to estimate the model, based on the description of the algorithm by Mislevy and Verhelst (1990) for a model that consists of a mixture of linear logistic test models (of which the mixture Rasch model is a special case, since a Rasch model is a linear logistic test model with an identity matrix as design matrix). A normal distribution was supposed for the latent variable within each class. The variances were allowed to differ across classes.

To avoid local maximum solutions, each dataset was analysed 10 times, with different random starting values. The solution with the highest loglikelihood was retained. As a matter of fact, local maximum solutions did occur very rarely: only for 10 datasets (all in the condition with equal variances for the latent variables and a correlation of .75), one or more parameter estimates differed more than .001 across the 10 runs of the EM-algorithm (we checked this after relabeling the classes as described next).

Due to the fact that the class labels are arbitrary, one can always obtain a second solution that is also a (global) maximizer of the likelihood by switching the class labels. So, by having random starting values for all parameters, the EM-algorithm will randomly converge to either one of these equivalent solutions. This problem can be solved by relabeling the classes such that a particular class always corresponds to a particular kind of solution. For example, one can always take the smallest class to be Class 1. A necessary condition for such a procedure is that the classes are well separated with respect to the parameters upon which the labeling is based. Otherwise, averaging the estimated parameters across datasets, one artificially creates differences between classes that are not present in the data. For example, suppose that datasets are generated according to a model in which both of two classes have an equal class weight of .5. If we then, for each analysis, take the first class to be the smallest one, we will conclude that, on average, the first class is the smallest one. In this study, we expected that the two classes showed a different pattern of item parameters. To check whether the two equivalent solutions were well separated in this respect, we visually inspected the plot of the item number versus the difference between the 100 estimated mixture Rasch model item parameters and the generating item parameters of the two-dimensional model, separately for each class and for each of the eight conditions. In Figure 2, such a plot is displayed for the condition with unequal variances for the latent variables and a correlation of .25. The points originating from the same dataset are connected with a line, so that the results of the different dataset are distinguishable. The two bands in the plot represent the two kinds of solutions, and we can conclude that the two kinds of solutions are well separated indeed with respect to the item parameters. Also for the other conditions, the two kinds of solutions showed up in the form of two bands.

INSERT FIGURE 2 ABOUT HERE

Therefore, for each of the analyses, the classes were relabeled such that the sum of the mean of the item parameters of the first 10 items (belonging to the first dimension of the two-dimensional model) in the first class and the mean of the item parameters of the last 10 items (belonging to the second dimension) in the second class was larger than the sum of the mean of the item parameters of the first 10 items in the second class and the mean of the item parameters of the last 10 items in the first class.

Figure 3 shows, for each of the eight conditions of the simulation study, the box plots of the estimated variances, the class weights, and the difference between the estimated mixture Rasch model item parameters and the generating item parameters of the two-dimensional model. With regard to the variance parameters, the box plots matched the expected estimates given in Table 1.

INSERT FIGURE 3 ABOUT HERE

With respect to the item parameters, we expect that the item parameters equal the item parameters of the between-item two-dimensional model up to a shift parameter that is specific for the dimension the item belongs to in the model. Consequently, we expect that, within a class, the differences between the item parameters estimated in the mixture Rasch model and the item parameters for the between-item two-dimensional model are constant over items belonging to the same dimension in the between-item two-dimensional model. Specifically, we expect the following differences for the items belonging to respectively the first and the second dimension in the between-item two-dimensional model (using Equation 10):

$$\beta_{itd_1} - \beta_i = b\frac{\sigma^2_{\theta_1} - \sigma_{\theta_1\theta_2}}{\sigma^2_{\theta_1-\theta_2}}$$

$$\beta_{itd_2} - \beta_i = b\frac{-\sigma^2_{\theta_2} + \sigma_{\theta_1\theta_2}}{\sigma^2_{\theta_1-\theta_2}} \qquad (11)$$

The box plots of the difference between the estimated mixture Rasch model item parameters and the generating item parameters of the two-dimensional model fits in with these expectations. First, within each class, the box plots were located at about the same height for the items belonging to the same dimension. Second, for the conditions with equal variances for θ_1 and θ_2, the box plots for the items belonging to respectively the first and the second dimension are located at the same distance but at the opposite side of the *X*-axis. Third, for the conditions with unequal variances for θ_1 and θ_2, the box plots of the items belonging to the second dimension are located at a larger distance from the *X*-axis than the box plots of the items belonging to the first dimension. These three results are in line with our expectations with respect to the item parameters given in Table 2.

In addition, the box plots are located closer to the *X*-axis with increasing correlation between θ_1 and θ_2, especially when the variances of θ_1 and θ_2 are equal. It indicates that the values *b* of θ_1 - θ_2 (see Table 2) the EM-algorithm "conditions upon" become smaller with increasing correlation between θ_1 and θ_2. That is, the distances between the classes (the straight lines with slope one, see Figure 1) and the origin of the $\boldsymbol{\theta}$ space become smaller with an increasing correlation between θ_1 and θ_2.

The *b* values for the different classes can be estimated with a regression without intercept, with as dependent variable the difference between the mean estimated values of the class-specific item parameters and the generating item parameters of the two-dimensional model, and as predictor the values of Table 2 (see Equation 11). The regression weight corresponds to b. The classes can then be located as lines in the original two-dimensional $\boldsymbol{\theta}$ space, following the equation:

$$\text{Class } t\text{: } \theta_1 - \theta_2 = b_t\,,\ t = 1, 2. \qquad (12)$$

The locations of the two classes in the two-dimensional space of the generating two-dimensional model is presented in Figure 4, in which also the .05 contour line is plotted for the respective bivariate normal distribution of $\boldsymbol{\theta}$. The values for b become smaller indeed with increasing correlation between θ_1 and θ_2, both when the variances of θ_1 and θ_2 are equal and unequal. Furthermore, for each condition of the simulation, b from Class 1 approximately equals minus b of Class 2. Thus, the classes are located symmetrically around the straight line through the origin with slope one, and the distances between the classes and the origin of the $\boldsymbol{\theta}$ space become smaller with an increasing correlation between θ_1 and θ_2.

INSERT FIGURE 4 ABOUT HERE

In sum, the relation between a multidimensional model and a continuous mixture of Rasch models seems to transfer to the case where the number of classes is only two, and the number of dimensions is two. The fact that the outer integral of Equation 6 is approximated by a sum over only two grid points has little or no effect on the parameter estimates.

Example: Verbal Aggression

The example is taken from a study of Vansteelandt (1999) on verbal aggression. The questions and response choices of a self-report study were designed with the expectation to indicate greater and lesser degree of verbal aggression, based on earlier theoretical and empirical work (Vansteelandt, 1999). The respondents had to indicate how they would react in four situations that were described in written format. Three types of verbally aggressive behavior were asked about (cursing, scolding, and shouting) for each situation. For each of the aggressive behaviors, respondents had to indicate both whether they wanted and whether they would actually show the verbally aggressive behavior. Hence, in total, there were 24 items. Three possible response categories were provided: “yes, “perhaps”, and “no”. We recoded “yes” and “perhaps” as “1”, and “no” as “0”. 316 first-year psychology students, 73 males and 243 females, participated in the study. A more detailed description of the study can be found in Vansteelandt (1999).

To determine whether all items belonged to the same scale or whether the “do” and “want” items formed separate subscales, both a Rasch model and a between-item two-dimensional model with separate dimensions for the “do” (θ_1) and “want” (θ_2) items were estimated. The estimations were conducted with the SAS NLMIXED procedure, a procedure to estimate nonlinear mixed models (SAS/STAT User’s guide, SAS Institute Inc., 1999; see Rijmen, Tuerlinckx, De Boeck, & Kuppens, 2003, for a nonlinear mixed model framework for IRT models). The loglikelihood was equal to -4036 and -3990, respectively for the Rasch model and the two-dimensional model. The Bayesian information criterion (BIC; Schwartz, 1978) was respectively equal to 8216 and 8135.

Hence, we concluded that the "do" and "want" items formed separate subscales. Table 4 shows the estimates of the item parameters for the 12 "do" items and the 12 "want" items (in that order), as well as the estimates of the variances of θ_1 ("do") and θ_2 ("want") and their covariance.

INSERT TABLE 3 ABOUT HERE

Subsequently, two mixture Rasch models with two latent classes were estimated. In the first model, no constraints were put on the parameters of the model. The loglikelihood and the BIC for this first model were equal to –3925 and 8143, respectively. In the second model, the item parameters were constrained according to the expectations based on the true underlying model being a two-dimensional model, with separate dimensions for the "do" and "want" items. As was explained before, these expectations result in the restrictions that the item parameters of the two classes are equal up to a shift parameter that is specific for the items belonging to the same subscale: $\beta_{i2} = \beta_{i1} + \eta_d$ for the "do " items, and $\beta_{i2} = \beta_{i1} + \eta_w$ for the "want " items. Hence, the following parameters were free parameters in the restricted mixture Rasch model: 24 item parameters for the first class, two shift parameters, one class weight (as the class weights have to sum to 1), and two variances. The loglikelihood and BIC for the restricted mixture Rasch model with two classes were equal to –3988 and 8143, respectively. According to the BIC, the restricted mixture Rasch model is as appropriate as the unrestricted. We continue with the restricted model as we expect it to be estimated more precisely because of the smaller amount of parameters. The item parameters of the second class were reconstructed as the sum of the item parameters of the first class and the shift parameters. The estimates for the latter amounted to η_d = -1.09, and η_w = .98, meaning that participants of the second class tended to answer "yes" more likely on the "do" items, and less likely on the "want" items in comparison with participants of the first class The estimates for the restricted mixture Rasch model, and the difference between the estimated (reconstructed) item parameters for respectively the mixture Rasch and the two-dimensional model are also given in Table 3. For the first class, the estimates of the "do" items are consistently smaller than the corresponding item parameters of the two-dimensional model, and the estimates of the "want" items are consistently larger. The reverse is observed for the second class. The order of the item locations within both the set of 12 "do" and the set of 12 "want" items is the same for both classes and for the two-dimensional model.

Conclusion

We compared two generalizations of the Rasch model. In the multidimensional extension of the Rasch model, more than one latent variable is incorporated. We focused on a between-item multidimensional model, in which each item belongs to precisely one dimension. Such a model is appropriate if a test consists of several subtests or *groups of items* that each can be modeled with a Rasch model. In the mixture Rasch model on the other hand, the existence of several *subgroups of participants* is assumed. Within each of the subgroups, the responses are modeled with the Rasch model. The mixture Rasch model is appropriate if all items of a test are measuring a single construct within each subgroup of participants, but the particular construct that is measured differs across subgroups. Both extensions of the Rasch model are conceptually quite distinct, addressing subgroups of items versus subgroups of participants, or in other words, addressing heterogeneity in item content versus qualitative heterogeneity between participants (Reise & Gomel, 1995). In spite of their different conceptual underpinning, a formal relation exists between both types of models. Specifically, we showed that a multidimensional model can be expressed in terms of a continuous mixture of Rasch models, that is, a mixture Rasch model with an infinite number of classes. In the mixture Rasch model, the item parameters within each class are equal to the item parameters of the multidimensional model up to a shift parameter that is specific for the dimension an item is belonging to in the multidimensional model. Consequently, the item parameters for items belonging to the same dimension in the between-item multidimensional model are equal across classes in the continuous mixture of Rasch models up to a class-specific shift parameter. A mixture Rasch model with this kind of structure on the item parameters is called a Saltus model (Mislevy & Wilson, 1996; Wilson, 1989). The Saltus model was developed in a developmental context. Our theoretical results indicate that a Saltus model (with an infinite number of classes) can also mimic a between-item multidimensional model. Note that the reverse, that every Saltus model corresponds to a between-item multidimensional model, is not necessarily true.

The formal relation between both types of models holds irrespective of the distribution of the latent variables. However, when a multivariate normal distribution is assumed, more precise expressions can be derived for the parameters characterising the mixture Rasch model.

The simulation study suggests that the relation between the two kinds of models also holds approximately for a two-dimensional model with a bivariate normal distribution for the latent variables when the number of classes is as small as two.

Our results indicate that it might be the case that, in a particular application, the resulting estimates from a mixture Rasch model do not represent the item parameters for qualitatively different subgroups, but that the solution is merely an approximation of a multidimensional model instead, in which all individual differences are quantitative differences. Specifically, a mixture Rasch model solution in which the items can be divided into subgroups for which the item parameters are constant across classes up to a shift parameter common to all items within a subgroup, is a strong indication for a between-item multidimensional model, with as many dimensions as there are different subgroups of items. However, such a mixture Rasch model, without restrictions on the item parameters, contains a lot of parameters, so that a pattern in the item parameters might be blurred because of imprecise estimates. Alternatively, when a particular between-item multidimensional model is a plausible model a priori, the item parameters of the mixture Rasch model can be constrained accordingly, leading to a Saltus model (Mislevy & Wilson, 1996). We followed the latter approach in the example study on verbal aggression.

References

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

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

De Leeuw, J., & Verhelst, N. (1986). Maximum likelihood estimation in generalized Rasch models. *Journal of Educational Statistics, 11*, 183-196.

Lindsay, B., Clogg, C.C., & Grego, J. (1991). Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association, 86*, 96-108.

Mislevy, R.J., & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika, 55,* 195-215.

Mislevy, R.J., & Wilson, M. (1996). Marginal maximum likelihood estimation for a psychometric model of discontinuous development. *Psychometrika, 61*, 41-71.

Reise, S.P., & Gomel, J.N. (1995). Modeling qualitative variation within latent trait dimensions: Application of mixed-measurement to personality assessment. *Multivariate Behavioral Research, 30*, 341-358.

Rijmen, F., Tuerlinckx, F., De Boeck, P., & Kuppens, P. (2003). A nonlinear mixed models framework for IRT models. *Psychological Methods, 8*, 185-205.

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement, 14,* 271-282.

SAS Institute Inc. (1999). *SAS OnlineDoc (Version 8)* [software manual on CD-ROM]. Cary, NC: SAS Institute Inc.

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*, 461-464.

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

Vansteelandt, K. (1999). A formal model for the competency-demand hypothesis. *European Journal of Personality, 13*(5, Spec Issue): 429-442.

Wilson, M.(1989). Saltus: A psychometric model of discontinuity in cognitive development. *Psychological Bulletin, 105*, 276-289.

Figure captions

Figure 1. *Geometrical representation of a between-item two-dimensional model as a continuous mixture of Rasch models.*

Figure 2. *Difference between the estimated mixture Rasch model parameters and the generating item parameters of the between-item two-dimensional model for the condition with* $\rho_{\theta_1\theta_2} = .25$, $\sigma^2_{\theta_1} = 1$, *and* $\sigma^2_{\theta_2} = 2$ *(100 datasets and 1000 subjects per dataset). Points resulting from the same dataset are connected with a line.*

Figure 3. *Box plots of the estimated variances, the class weights, and the difference between the estimated mixture Rasch model item parameters and the generating item parameters of the two-dimensional model(100 datasets and 1000 subjects per dataset for each condition).*
(a) $\rho_{\theta_1\theta_2} = 0; \sigma^2_{\theta_1} = \sigma^2_{\theta_2} = 1$; *(b)* $\rho_{\theta_1\theta_2} = 0; \sigma^2_{\theta_1} = 1, \sigma^2_{\theta_2} = 2$;
(c) $\rho_{\theta_1\theta_2} = .25; \sigma^2_{\theta_1} = \sigma^2_{\theta_2} = 1$; *(d)* $\rho_{\theta_1\theta_2} = .25; \sigma^2_{\theta_1} = 1, \sigma^2_{\theta_2} = 2$;
(e) $\rho_{\theta_1\theta_2} = .5; \sigma^2_{\theta_1} = \sigma^2_{\theta_2} = 1$; *(f)* $\rho_{\theta_1\theta_2} = .5; \sigma^2_{\theta_1} = 1, \sigma^2_{\theta_2} = 2$;
(g) $\rho_{\theta_1\theta_2} = .75; \sigma^2_{\theta_1} = \sigma^2_{\theta_2} = 1$; *(h)* $\rho_{\theta_1\theta_2} = .75; \sigma^2_{\theta_1} = 1, \sigma^2_{\theta_2} = 2$.

Figure 4. *Location of the two classes of the mixture Rasch model (based on the mean estimated values of the item parameters) as lines in the two-dimensional space of the generating between-item two-dimensional model. For the latter, the .05 density contour is drawn.*
(a) $\rho_{\theta_1\theta_2} = 0; \sigma^2_{\theta_1} = \sigma^2_{\theta_2} = 1$; *(b)* $\rho_{\theta_1\theta_2} = 0; \sigma^2_{\theta_1} = 1, \sigma^2_{\theta_2} = 2$;

(c) $\rho_{\theta_1\theta_2} = .25; \sigma^2_{\theta_1} = \sigma^2_{\theta_2} = 1$; *(d)* $\rho_{\theta_1\theta_2} = .25; \sigma^2_{\theta_1} = 1, \sigma^2_{\theta_2} = 2$;

(e) $\rho_{\theta_1\theta_2} = .5; \sigma^2_{\theta_1} = \sigma^2_{\theta_2} = 1$; *(f)* $\rho_{\theta_1\theta_2} = .5; \sigma^2_{\theta_1} = 1, \sigma^2_{\theta_2} = 2$;

(g) $\rho_{\theta_1\theta_2} = .75; \sigma^2_{\theta_1} = \sigma^2_{\theta_2} = 1$; *(h)* $\rho_{\theta_1\theta_2} = .75; \sigma^2_{\theta_1} = 1, \sigma^2_{\theta_2} = 2$;

Table 1

Expected variances for the classes of the mixture Rasch model as a function of the covariance structure of the between-item two-dimensional model.

	$\sigma^2_{\theta_1} = \sigma^2_{\theta_2} = 1$	$\sigma^2_{\theta_1} = 1; \sigma^2_{\theta_2} = 2$
$\rho_{\theta_1\theta_2} = 0$	.5	.67
$\rho_{\theta_1\theta_2} = .25$	.63	.82
$\rho_{\theta_1\theta_2} = .5$	.75	.95
$\rho_{\theta_1\theta_2} = .75$	.88	1.00

Table 2

Expected item parameters for the classes of the mixture Rasch model as a function of the covariance structure of the between-item two-dimensional model.

	$\sigma^2_{\theta_1} = \sigma^2_{\theta_2} = 1$		$\sigma^2_{\theta_1} = 1; \sigma^2_{\theta_2} = 2$	
	Items of first dimension	Items of second dimension	Items of first dimension	Items of second dimension
$\rho_{\theta_1\theta_2} = 0$	$\beta_i + .5b$	$\beta_i - .5b$	$\beta_i + .33b$	$\beta_i - .67b$
$\rho_{\theta_1\theta_2} = .25$	$\beta_i + .5b$	$\beta_i - .5b$	$\beta_i + .28b$	$\beta_i - .72b$
$\rho_{\theta_1\theta_2} = .5$	$\beta_i + .5b$	$\beta_i - .5b$	$\beta_i + .18b$	$\beta_i - .82b$
$\rho_{\theta_1\theta_2} = .75$	$\beta_i + .5b$	$\beta_i - .5b$	$\beta_i - .07b$	$\beta_i - 1.07b$

Table 3. *Results of the study on verbal aggression. Dimension the item belongs to in the between-item two-dimensional model, parameter estimates for the between-item two-dimensional model, (reconstructed) parameter estimates for the restricted two-class mixture Rasch model, and the difference between the estimates of the parameters of the two-dimensional and mixture Rasch model.*

	Dimension of item in 2-dimensional model	2-dimensional model	Class 1 mixture Rasch model	Difference item parameters Class 1 and 2-dimensional model	Class 2 mixture Rasch model[b]	Difference item parameters Class 2 and 2-dimensional model
π_t			.57		.43	
$\boldsymbol{\Sigma}^a$		2.85 1.91	.93		4.72	
		1.91 2.11				
β_1	Do	1.33	1.81	.48	.72	-.61
β_2	Do	.41	.87	.46	-.22	-.62
β_3	Do	-.96	-.47	.50	-1.55	-.59
β_4	Do	.94	1.41	.47	.32	-.62
β_5	Do	-.08	.39	.47	-.70	-.62
β_6	Do	-1.62	-1.09	.53	-2.18	-.56
β_7	Do	-.25	.22	.47	-.86	.61
β_8	Do	-1.64	-1.11	.53	-2.20	-.56
β_9	Do	-3.21	-2.60	.61	-3.69	-.48
β_{10}	Do	.76	1.22	.46	.14	-.62
β_{11}	Do	-.44	.04	.48	-1.04	-.61
β_{12}	Do	-2.17	-1.61	.56	-2.70	-.53
β_{13}	Want	1.25	.99	-.27	1.96	.71
β_{14}	Want	.58	.33	-.25	1.31	.72

Table continues

Table 3 *(continued)*

β_{15}	Want	.09	-.16	-.25	.82	.73
β_{16}	Want	1.79	1.51	-.28	2.49	.70
β_{17}	Want	.73	.47	-.26	1.45	.72
β_{18}	Want	.02	-.23	-.25	.75	.73
β_{19}	Want	.55	.29	-.25	1.27	.72
β_{20}	Want	-.70	-.96	.26	.02	.72
β_{21}	Want	-1.56	-1.87	-.31	-.89	.67
β_{22}	Want	1.11	.85	-.26	1.83	.71
β_{23}	Want	-.35	-.60	-.25	.37	.73
β_{24}	Want	-1.06	-1.34	-.28	-.36	.70

[a]: $\boldsymbol{\Sigma}$ represents either the covariance matrix (between-item two-dimensional model) or the variance within a class (mixture Rasch model).

[b]: The item parameters of the second class were reconstructed as the sum of the item parameters of the first class and the shift parameters.