# SAS macros for longitudinal IRT models

**Maja Olsbjerg** [1] , **Karl Bang Christensen** [Corresp. 1]

[1] Department of Public Health, Section of Biostatistics, University of Copenhagen, Copenhagen, Danmark

Corresponding Author: Karl Bang Christensen
Email address: kach@sund.ku.dk

IRT models are often applied when observed items are used to measure a unidimensional latent variable. Originally used in educational research, IRT models are now widely used when focus is on physical functioning or psychological well-being. Modern applications often need more general models, typically models for multidimensional latent variables or longitudinal models for repeated measurements. This paper describes a collection of **SAS** macros that can be used for fitting data to, simulating from, and visualizing longitudinal IRT models. The macros encompass dichotomous as well as polytomous item response formats and are sufficiently flexible to accommodate changes in item parameters across time points and local dependence between responses at different time points.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

# SAS Macros for Longitudinal IRT Models

Maja Olsbjerg[1] and Karl Bang Christensen[1]

[1]Section of Biostatistics, Department of Public Health, University of Copenhagen, Copenhagen, Denmark. [Corresponding author: kach@sund.ku.dk]

March 14, 2018

**Abstract**

IRT models are often applied when observed items are used to measure a unidimensional latent variable. Originally used in educational research, IRT models are now widely used when focus is on physical functioning or psychological well-being. Modern applications often need more general models, typically models for multidimensional latent variables or longitudinal models for repeated measurements. This paper describes a collection of `SAS` macros that can be used for fitting data to, simulating from, and visualizing longitudinal IRT models. The macros encompass dichotomous as well as polytomous item response formats and are sufficiently flexible to accommodate changes in item parameters across time points and local dependence between responses at different time points.

Keywords: polytomous IRT model, Rasch model, 1PL model, Birnbaum model, 2PL model, Partial Credit model, Generalized Partial Credit model, longitudinal IRT model, marginal maximum likelihood (MML) estimation, item parameter drift, response dependence, SAS.

## 1 Introduction

Item response theory (IRT) models were developed to describe probabilistic relationships between test items and latent traits (van der linden & Hambleton, 1997).



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

Originally developed and used in educational testing to describe how the probability of a correct answer to test item depends on ability, they are applicable whenever location of persons and items on an underlying latent scale is of interest. Traditional applications in education often use dichotomous (correct/incorrect) item scoring, but polytomous items are common in other applications. The use of IRT models in new research fields increases the need for implementation in standard statistical software like SAS. Estimation in IRT models using SAS has been the topic of several research papers (Rijmen et al., 2003; Smits & De Boeck, 2003; Nandakumar & Hotchkiss, 2012). In particular, implementation of polytomous Rasch models in SAS has been discussed (Christensen, 2006), and SAS macros that used these ideas are available (Christensen & Bjorner, 2003; Hardouin & Mesbah, 2007; Christensen & Olsbjerg, 2013; Christensen, 2013).

Many applications require more general models. Typically when multidimensional latent variables are considered or when repeated measurements are used. Longitudinal Rasch models were studied by Pastor and Beretvas (2006), who illustrated how these models can be seen as hierarchical generalized linear models and implemented in the software program `HLM` (Raudenbush et al., 2004) that uses penalized quasi-likelihood for estimation, but as noted by Pastor and Beretvas, various estimation procedures and software programs for these kinds of models exist. An example is the Random Weights Linear Logistic Test Model (Rijmen & De Boeck, 2002), which is a special case of the multidimensional random coefficients multinomial logit model (Adams et al., 1997a) implemented in the computer program ConQuest (Wu et al., 2007). IRT models are increasingly applied in health status measurement and evaluation of Patient Reported Outcomes (Reeve et al., 2007). The simplest IRT model, the Rasch (1960) model, (Fischer & Molenaar, 1995; Christensen et al., 2013), is increasingly used for validation of measurement instruments (Tennant & Conaghan, 2007) and has been shown to be superior to classical approaches (Blanchin et al., 2011). This paper describes `SAS` macros that are available from

https://github.com/KarlBangChristensen/LIRT

and can can be used to fit longitudinal IRT models with the possibility to model item parameter drift and response dependence across time points. Four macros: (i) `%LIRT_MML` that estimates item parameters, (ii) `%LIRT_PPAR` that estimates person parameters and changes in person parameters over time given estimated item parameters, (iii) `%LIRT_ICC` that plots item characteristic curves, and (iv) `%LIRT_SIMU` that generates data sets with responses simulated from the model are described in



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

detail and illustrated using data from a longitudinal study based on a HRQoL questionnaire applied to women screened for breast cancer.

## 2 IRT models for a single time point

IRT models describe responses to items $\boldsymbol{X} = (X_i)_{i \in I}$ measuring a latent variable $\theta \in \mathbb{R}$. For dichotomously scored items two IRT models have traditionally been applied: the Rasch or 1PL model (Rasch, 1960; Fischer & Molenaar, 1995) and the Birnbaum or 2PL model (Birnbaum, 1968). These, the simplest IRT models, are given by the probabilities

$$P(X_i = x_i|\theta) = \frac{\exp(x_i(\theta + \eta_i)}{1 + \exp(\theta + \eta_i)} \qquad (x_i = 0, 1) \tag{1}$$

and

$$P(X_i = x_i|\theta) = \frac{\exp(\alpha_i(x_i(\theta + \eta_i)))}{1 + \exp(\alpha_i(\theta + \eta_i))} \qquad (x_i = 0, 1) \tag{2}$$

for each item $i \in I$. The formula (1) defining the Rasch model and the formula (2) defining the Birnbaum model reveal that these are logistic regression models. When $\alpha_i$ does not differ across items the IRT models (2) is identical to (1). For ordinal item response formats (1) can be generalized to

$$P(X_i = x_i|\theta) = \frac{\exp(x_i\theta + \eta_{ix_i})}{\sum_{k=0}^{m_i} \exp(k\theta + \eta_{ik})} \qquad (x_i = 0, 1, \ldots, m_i) \tag{3}$$

and the model (2) can be generalized to

$$P(X_i = x_i|\theta) = \frac{\exp(\alpha_i(x_i\theta + \eta_{ix_i}))}{\sum_{k=0}^{m_i} \exp(\alpha_i(k\theta + \eta_{ik}))} \qquad (x_i = 0, 1, \ldots, m_i) \tag{4}$$

where for identification purposes $\eta_{i0} = 0$ for all $i \in I$. Note that for $m_i = 1$ this corresponds to setting $\eta_{i0} = 0$ and $\eta_{i1} = \eta_i$ in (1) and (2), respectively. The model (3) is called the partial credit model (PCM; Masters 1982), but was originally proposed without this name by Andersen (1977)), and the model (4) is called the generalized partial credit model (GPCM; Muraki 1992). In these models the parameter $\alpha_i$ and



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

the vector $\boldsymbol{\eta}_i = (\eta_{ik})_{k=1,\ldots,m_i}$ are parameters describing the item while $\theta$ is a parameter describing the person responding. In the psychometric literature $\alpha_i$ is referred to as an item discrimination parameter. Thus, both dichotomous and polytomous Rasch models appear as the special case where the item discrimination is constant

$$\alpha_i = \alpha \quad \text{for all} \quad i \in I. \tag{5}$$

An alternative parametrization of the $\eta$'s can be obtained by using the so-called item thresholds defined by $\beta_{i0} = 0$ and

$$\beta_{ik} = -(\eta_{ik} - \eta_{ik-1}) \qquad (k = 1, \ldots, m_i) \tag{6}$$

for each item $i \in I$. These are called thresholds because they correspond to the values of the latent variable where adjacent categories equally likely: $P(X_i = x|\theta = \beta_{ix}) = P(X_i = x - 1|\theta = \beta_{ix})$.

## 2.1 Interpretation of item parameters

The models all specify the conditional probabilities of each response option given the value $\theta$ of the latent variable. The item parameters can most easily be interpreted using plots of these conditional probabilities against $\theta$, i.e., plots of the functions

$$\theta \mapsto P(X_i = x|\theta) \qquad (\text{for } x = 0, 1, \ldots, m_i). \tag{7}$$

These are called item characteristic curves (ICC's). Figure 1 shows examples of ICC's for items with $(\alpha, \boldsymbol{\eta}) = (1, (1, -1))$ and $(\alpha, \boldsymbol{\eta}) = (2.5, (0, -1))$, respectively.

Note that the curves are steeper for the item with the highest value of the discrimination parameter $\alpha$.

## 2.2 Joint likelihood and identification

IRT models share the technical assumption that items are locally independent, that is the vector $\boldsymbol{X}$ of item responses satisfies



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

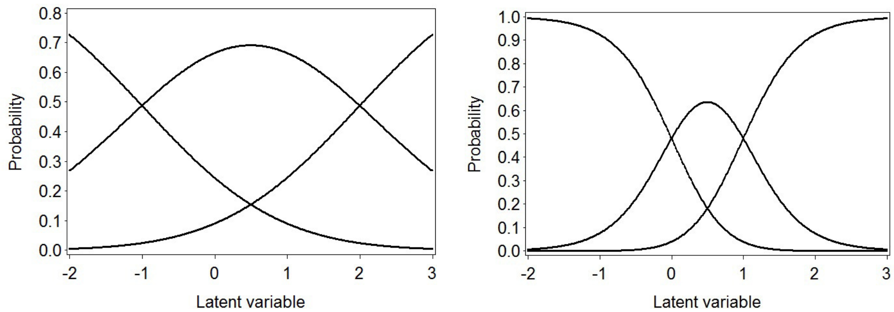


Figure 1: Item characteristic curves for an item with $\alpha = 1$ and $\boldsymbol{\eta} = (1, -1)$, i.e. $\boldsymbol{\beta}) = (-1, 2)$ (left panel) and for an item with $\alpha = 2.5$ and $\boldsymbol{\eta}) = (0, -1)$, i.e. $\boldsymbol{\beta} = (0, 1)$ (right panel).

$$P(\boldsymbol{X} = \boldsymbol{x}|\,\theta) = \prod_{i \in I} P(X_i = x_i|\theta) \text{ for all } \theta \in \mathbb{R}. \tag{8}$$

For a sample of persons $v = 1, \ldots, N$, where subject $v$ responds to items in a subset $I_v \subset I$ the joint likelihood implied by (8) is given by

$$L(\boldsymbol{\alpha}, \boldsymbol{\eta_i}, \boldsymbol{\theta}) = \prod_{v=1}^{N} P(\boldsymbol{X}_v = \boldsymbol{x}_v|\theta_v) = \prod_{v=1}^{N} \prod_{i \in I_v} P(X_{vi} = x_{vi}|\theta_v) \tag{9}$$

where $\boldsymbol{\theta} = (\theta_v)_{v=1,\ldots,N}$, $\boldsymbol{\alpha} = (\alpha_i)_{i \in I}$, and $\boldsymbol{\eta} = (\eta_{ik})_{i \in I, k=1,\ldots,m_i}$, for $i \in I$. Inserting the probabilities (4) in (9) yields

$$L(\boldsymbol{\alpha}, \boldsymbol{\eta_i}, \boldsymbol{\theta}) = \frac{\exp\left(\sum_{v=1}^{N} \sum_{i \in I_v} \theta_v \alpha_i x_{vi} + \sum_{v=1}^{N} \sum_{i \in I_v} \alpha_i \eta_{i x_{vi}}\right)}{\prod_{v=1}^{N} \prod_{i \in I_v} \sum_{k=0}^{m_i} \exp(\alpha_i(k\theta_v + \eta_{ik}))]}. \tag{10}$$

The model is only identified if restrictions are placed on either item or person parameters, since the reparametrizations

$$\alpha^* = \gamma\alpha \quad \text{and} \quad \theta^* = \frac{1}{\gamma}\theta \tag{11}$$

or

$$\eta^*_{ik} = \eta_{ik} - \omega k \quad \text{and} \quad \theta^* = \theta + \omega \tag{12}$$

for arbitrary $\gamma, \omega > 0$ yield the same probabilities. An often used restriction is

$$\prod_{i \in I} \alpha_i = 1 \tag{13}$$

for the discrimination parameters, and

$$\sum_{i=1}^{I} \eta_{i,m_i} = 0 \tag{14}$$

or equivalently

$$\sum_{i=1}^{I} \sum_{h=0}^{m_i} \beta_{ih} = 0 \tag{15}$$

for the thresholds, but restrictions on the person parameters can also be imposed. If the mean of $\theta$ is restricted then all $\eta$'s can be estimated, and if the variance of $\theta$ is restricted then all $\alpha$'s can be estimated.

## 2.3 Item parameter estimation

Estimation based on the likelihood (10) leads to inconsistent estimates (Neyman & Scott, 1948). In the special case of the Rasch model, conditional maximum likelihood (CML) estimation (Andersen, 1973) can be used for item parameter estimation, but in general marginal maximum likelihood (MML) estimation (Bock & Aitkin, 1981; Thissen, 1982; Zwinderman & van den Wollenberg, 1990) is used. This estimation method is based on a distributional assumption about the latent variable, typically that $\theta_1, \ldots, \theta_n$ are iid and normally distributed. If the item parameters are restricted as in (13) and (15) the mean and variance of this normal distribution can be estimated. Alternatively, assuming that $\theta \sim N(0, 1)$ all item parameters can be estimated. The marginal likelihood is given as

$$L_M(\boldsymbol{\alpha},\boldsymbol{\eta}) = \prod_{v=1}^{N} \int P(\boldsymbol{X}_v = \boldsymbol{x}_v|\theta)\varphi(\theta)d\theta \tag{16}$$

and random effects models like this are easily implemented in `PROC NLMIXED` in `SAS`. The implementation assumes that $\theta \sim N(0,1)$ and estimates all $\alpha$'s and $\eta$'s. If the restriction (13) is imposed on the $\alpha's$ we can assume that $\theta \sim N(0,\sigma^2)$ and estimate $\sigma^2$. From (11) we see that

$$\sigma = \left(\prod_{i\in I} \alpha_i\right)^{\frac{1}{I}}.$$

If the restriction (15) is imposed on the threshold parameters, we can assume $\theta \sim N(\mu,1)$ and estimate the mean $\mu$. From (12) we see that

$$\mu = -\frac{1}{\sum_{i\in I} m_i}\left(\sum_{i\in I}\sum_{h=1}^{m_i} \beta_{ih}\right).$$

### 2.4 Person parameter estimation

It is usually of interest to estimate individual values of the latent variable. This can be done by substituting item parameters by their MML estimates resulting in the likelihood function

$$L_P(\cdot) = L(\widehat{\boldsymbol{\alpha}}, (\widehat{\boldsymbol{\eta}}, \cdot). \tag{17}$$

Estimates of $\theta_v$ can then be obtained by numerical optimization of this likelihood.

## 3 Longitudinal IRT models

For time points $t = 1,2$ let $\boldsymbol{X}_t = (X_{it})_{i\in I}$ be a set of items measuring a value $\theta_t \in \mathbb{R}$ of the latent variable. In a situation like this it is natural to assume that $\mathrm{Corr}(\theta_{t_1},\theta_{t_2}) > 0$ and thus that item responses from the same person at two time points $t_1$ and $t_2$ are positively correlated. For simplicity we consider the situation where $T = 2$ and where for each person we are interested in baseline measurements



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

$\theta_1$, follow-up measurements $\theta_2$ and measurement of change $\theta_2 - \theta_1$. Note that even though the total item set $I$ is the same across items administered at only one time point can still be included since we assume that for each person $v \in \{1, \ldots, N\}$ we observe a response vector $\boldsymbol{X}_{v1} = (X_{vi1})_{i \in I_{v1}}$ at time 1 and $\boldsymbol{X}_{v2} = (X_{vi2})_{i \in I_{v2}}$ at time 2, for $I_{v1}, I_{v2} \subset I$. We will assume all items fit the model (4) for each time point $t = 1, 2$

$$P(X_{it} = x_{it} | \theta_t) = \frac{\exp(\alpha_{it}(x_{it}\theta_t + \eta_{ix_{it}t}))}{\sum_{k=0}^{m_i} \exp(\alpha_{it}(k\theta_t + \eta_{ikt}))} \tag{18}$$

and that the assumption of local independence (8) holds within each time point

$$P(\boldsymbol{X}_t = \boldsymbol{x}_t | \, \theta_t) = \prod_{i \in I} P(X_{it} = x_{it} | \theta) \text{ for } t = 1, 2 \text{ and all } \theta \in \mathbb{R}. \tag{19}$$

However, further assumptions about the dependence structure are needed to fully specify the model.

### 3.1 Simple model

The simplest specification of the longitudinal model is obtained by extending the independence assumption (8) to hold for all pairs of items $(i, i')$ and all time points $t$ and $t'$

$$P(X_{i1} = x_{i1}, X_{i'2} = x_{i'2} | \, \theta_1, \theta_2) = P(X_{i1} = x_{i1} | \, \theta_{t_1}) P(X_{i'2} = x_{i'2} | \, \theta_2) \tag{20}$$

and assuming that the item parameters are constant over time, i.e., that for all $i \in I$

$$\alpha_{i1} = \alpha_{i2} \quad \text{and} \quad \boldsymbol{\eta}_{i1} = \boldsymbol{\eta}_{i2}. \tag{21}$$

For the special case of the dichotomous Rasch model the longitudinal model imposed by the assumptions (20) and (21) was discussed by Andersen (1985) and Embretson (1991). Using these assumptions the contribution to the joint likelihood for person $v$ becomes



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

$$
\begin{aligned}
L(\boldsymbol{\eta}, \boldsymbol{\theta}_v) &= \prod_{t=1}^{2} P(\boldsymbol{X}_{vt} = \boldsymbol{x}_{vt} | \theta_{vt}) \\
&= \prod_{t=1}^{2} \prod_{i \in I_{vt}} \frac{\exp(\alpha_i(x_{vit}\theta_{vt} + \eta_{ix_{vit}}))}{\sum_{k=0}^{m_i} \exp(\alpha_i(k\theta_{vt} + \eta_{ik}))}
\end{aligned} \tag{22}
$$

where $\boldsymbol{\theta}_v = (\theta_1, \theta_2)$. Assuming that the vectors $\boldsymbol{\theta}_v$) are iid from a 2-dimensional normal distribution $\mathcal{N}(\mu, \Sigma)$ item and population parameters can be estimated using MML estimation. Several papers have considered this specification of the model in the special case of the Rasch model Andersen (1977); Embretson (1991); Adams et al. (1997a,b).

## 3.2 Allowing for local dependence across time points

When an item is used both time points the extended assumption of local independence (20) might not be justified. In that case we should stick to what we know, namely that

$$
P(X_{i1} = x_{i1}, X_{i2} = x_{i2} | \, \theta_1, \theta_2) = P(X_{i1} = x_{i1} | \, \theta_1) P(X_{i2} = x_{i2} | \, X_{i1} = x_{i1}; \, \theta_2) \tag{23}
$$

Taking account of local dependence is then a matter of choosing a suitable model for the conditional probabilities in (23). One option is stick to the GPCM assuming that

$$
P(X_{i2} = x_{i2} | X_{i1} = x_{i1}; \theta_2) = \frac{\exp[x_{i2}\alpha^*_{i2}(x_{i1})(\theta_2 + \eta^*_{ix_{i2}2}(x_{i1}))]}{\sum_{k=0}^{m_i} \exp[\alpha^*_{i2}(x_{i1})(k\theta_2 + \eta^*_{ik2}(x_{i1}))]} \tag{24}
$$

where $\alpha^*_{i2}(x_{i1})$ and $\eta^*_{i2}(x_{i1})$ are item parameters depending on the response observed at time 1. If item $i$ meets the extended assumption of local independence (20) then $\alpha^*_{i2}(x)$ and $\boldsymbol{\eta}^*_{i2}(x)$ are constant across $x \in \{0, 1, ..., m_i\}$. Henceforth, we assume that this is the case for some items forming the subset $I_0 \subseteq I$.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

### 3.3 Allowing for item parameter drift

Another way of adding flexibility to (22) is by allowing item parameters to change over time. Clearly change in a person is easiest to evaluate when the measurement instrument does not change over time. However, it is feasible to evaluate changes in the person parameters $\theta$ as long as a subset $J_0 \subset I_0$ of the items satisfy (21).

### 3.4 General model

A flexible model that extends (22) to a model with local dependence and change in tem parameters across time points, for some items, can be formulated by specifying the item subsets $I_0$ and $J_0$. The contribution to the joint likelihood for person $v$ with response vector $\boldsymbol{X}_{vt} = (X_{vit})_{i \in I_{vt}}$ becomes

$$\begin{aligned}
L(\boldsymbol{alpha}, \boldsymbol{\eta}, \boldsymbol{\theta}_v) &= \prod_{i \in I \bigcap I_{v1}} P(X_{vi1} = x_{vi1} | \theta_{v1}) \\
&\quad \prod_{i \in I_0 \bigcap I_{v2}} P(X_{vi2} = x_{vi2} | \theta_{v2}) \\
&\quad \prod_{i \in (I \setminus I_0) \bigcap I_{v2}} P(X_{vi2} = x_{vi2} | X_{vi1} = x_{vi1}; \theta_{v2}) \\
&= \prod_{i \in I \bigcap I_{v1}} \frac{\exp(\alpha_{i1}(x_{vi1}\theta_{v1} + \eta_{ix_{vi1}1}))}{\sum_{k=0}^{m_i} \exp(\alpha_{i1}(k\theta_{v1} + \eta_{ik1}))} \\
&\quad \prod_{i \in J_0 \bigcap I_{v2}} \frac{\exp(\alpha_{i1}(x_{vi2}\theta_{v2} + \eta_{ix_{vi2}1}))}{\sum_{k=0}^{m_i} \exp(\alpha_{i1}(k\theta_{v2} + \eta_{ik1}))} \\
&\quad \prod_{i \in (I_0 \setminus J_0) \bigcap I_{v2}} \frac{\exp(\alpha_{i2}(x_{vi2}\theta_{v2} + \eta_{ix_{vi2}2}))}{\sum_{k=0}^{m_i} \exp(\alpha_{i2}(k\theta_{v2} + \eta_{ik2}))} \\
&\quad \prod_{i \in (I \setminus I_0) \bigcap I_{v2}} \frac{\exp[\alpha^*_{i2}(x_{vi1})(x_{vi2}\theta_{v2} + \eta^*_{ix_{vi2}2}(x_{vi1}))]}{\sum_{k=0}^{m_i} \exp[\alpha^*_{i2}(x_{vi1})(k\theta_{v2} + \eta^*_{ik2}(x_{vi1}))]}.
\end{aligned} \tag{25}$$

Parameter restrictions are needed in order for the model to be identified.

### 3.5 Item parameter estimation

Joint estimation based on (25) leads to inconsistent estimates. Instead we can turn to MML estimation assuming that

$$\begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix} \sim \mathcal{N}_2\left(\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}, \begin{bmatrix}\sigma_1^2 & \sigma_1\sigma_2\rho\\ \sigma_1\sigma_2\rho & \sigma_2^2\end{bmatrix}\right) \tag{26}$$

where $\rho = \mathrm{Corr}(\theta_1, \theta_2)$ is the latent correlation. An alternative parametrization obtained by restricting the latent distribution at time 1 leads to a parametrization of the longitudinal IRT model in terms of change in the mean and variance, which is often the main interest. This can be expressed

$$\begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix} \sim \mathcal{N}_2\left(\begin{bmatrix}0\\ \mu\end{bmatrix}, \begin{bmatrix}1 & \sigma\rho\\ \sigma\rho & \sigma^2\end{bmatrix}\right) \tag{27}$$

For either of the choices (26) or (27) the item parameters and the parameters of the latent distribution can be estimated by numerical optimization of an approximation to the marginal likelihood

$$L(\boldsymbol{alpha}, \boldsymbol{\eta}) = \int_{\mathbb{R}^2} L(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\theta})\varphi_{\mu,\Sigma}(\boldsymbol{\theta})d\boldsymbol{\theta} \tag{28}$$

obtained by integrating out the random effects.

#### 3.5.1 Estimation of person parameters

It is usually of interest to estimate change at the individual level. In the previous section we described how the item parameters can be estimated by assuming a certain distribution for the latent vector and then maximizing an approximation to the marginal likelihood. Estimation of the person parameters can be done in a similar by assuming a one point distribution for the item parameters. More specifically, we substitute the item parameters in (25) by their MML estimates resulting in the likelihood function

$$L_P(\cdot) = L(\widehat{\boldsymbol{\alpha}}, (\widehat{\boldsymbol{\eta}}_i)_{i\in I}, \cdot) \tag{29}$$



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

yielding

$$
\begin{aligned}
L_P(\boldsymbol{\theta}) \quad = \quad & \prod_{i \in I \bigcap I_{v1}} \frac{\exp(\widehat{\alpha}_{i1}(x_{vi1}\theta_{v1} + \widehat{\eta}_{ix_{vi1}1}))}{\sum_{k=0}^{m_i} \exp(\widehat{\alpha}_{i1}(k\theta_{v1} + \widehat{\eta}_{ik1}))} \\
& \prod_{i \in J_0 \bigcap I_{v2}} \frac{\exp(\widehat{\alpha}_{i1}(x_{vi2}\theta_{v2} + \widehat{\eta}_{ix_{vi2}1}))}{\sum_{k=0}^{m_i} \exp(\widehat{\alpha}_{i1}(k\theta_{v2} + \widehat{\eta}_{ik1}))} \\
& \prod_{i \in (I_0 \setminus J_0) \bigcap I_{v2}} \frac{\exp(\widehat{\alpha}_{i2}(x_{vi2}\theta_{v2} + \widehat{\eta}_{ix_{vi2}2}))}{\sum_{k=0}^{m_i} \exp(\widehat{\alpha}_{i2}(k\theta_{v2} + \widehat{\eta}_{ik2}))} \\
& \prod_{i \in (I \setminus I_0) \bigcap I_{v2}} \frac{\exp[\widehat{\alpha}^*_{i2}(x_{vi1})(x_{vi2}\theta_{v2} + \widehat{\eta}^*_{ix_{vi2}2}(x_{vi1}))]}{\sum_{k=0}^{m_i} \exp[\widehat{\alpha}^*_{i2}(x_{vi1})(k\theta_{v2} + \widehat{\eta}^*_{ik2}(x_{vi1}))]}. \qquad (30)
\end{aligned}
$$

Estimates of $\boldsymbol{theta}_v$ can then be obtained by numerical optimization of (30).

# 4 Implementation in SAS

Our implementation consists of four `SAS` macros that provides a framework for fitting and comparing different IRT models. The general specification longitudinal GPCM represented by (25) can be fitted, and using different choices of the item subsets $I_0$ and $J_0$ different models can be specified. The macros handle the longitudinal case with two time points as described, and also the special case where only a single time point is considered. The macro `%LIRT_MML` estimates parameters using `PROC NLMIXED` that fits nonlinear mixed models (Rijmen et al., 2003; Smits & De Boeck, 2003) and is very flexible because the conditional distribution given the random effects can be specified to be a general distribution using `SAS` programming statements. The procedure maximizes an approximation to the likelihood integrated over the random effects. Different integral approximations are available, the principal one being adaptive Gaussian quadrature. The macro `%LIRT_PPAR` estimates the latent variable(s) for each person, while `%LIRT_ICC`, plots item characteristic curves (ICC's) for given item parameters. Finally, the macro `%LIRT_SIMU` simulates responses from a given model within the framework. The macros use different input data sets and options listed below. One important input data set is similar for all macros, the data set `names`, that determines whether there is item parameter drift and local dependence across time points for any of the items. For `%LIRT_MML` it also specifies which items are to be modeled according to the 1PL model and the 2PL model respectively. For `%LIRT_ICC` and `%LIRT_SIMU` additional variables holding the item parameter values are required. Furthermore, a couple more variables with information about the local dependence structure are needed for `%LIRT_SIMU`. Input data sets and options for the macros are listed below.

### Input data sets and options for `%LIRT_MML`

- `data`: data set with item responses. Each person should be represented by one record and each item by one variable.
- `names`: a model-specifying data set with information about the items.
- `dim`: dimension of the latent variable (1 or 2).
- `out`: prefix for output data sets.
- `delete`: indicating whether temporary data sets created by the macro should be deleted (Y) or not (N).

### Input data sets and options for `%LIRT_PPAR`

- `data`: data set with item responses. Each person should be represented by one record and each item by one variable.
- `names`: a model-specifying data set with information about the items and estimates of the item parameters.
- `dim`: dimension of the latent variable (1 or 2).
- `id`: variable in `data` with unique person ID's.
- `out`: prefix for output data sets.
- `delete`: indicating whether temporary data sets created by the macro should be deleted (Y) or not (N).

### Input data sets and options for `%LIRT_ICC`



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

- `names`: a model-specifying data set with information about the items and estimates of the item parameters.
- `dim`: dimension of the latent variable (1 or 2).
- `out`: prefix for output data sets.
- `delete`: indicating whether temporary data sets created by the macro should be deleted (Y) or not (N).

## Input data sets and options for `%LIRT_SIMU`

- `names`: a model-specifying data set with information about the items and estimates of the item parameters.
- `dim`: dimension of the latent variable (1 or 2).
- `ndata`: number of data sets.
- `npersons`: number of persons.
- `pdata`: data set specifying the normal distribution of the latent variable.
- `out`: prefix for output data sets.
- `delete`: indicating whether temporary data sets created by the macro should be deleted (Y) or not (N).

The variables required in the data set `names` differ depending on whether one or two time points are considered. It should become clear from the data example below what they should be. All of the data sets created by the macros are given names starting with an underscore. Such data sets are the only ones deleted with the option `delete=Y`.
As for the output data sets an important one of `%LIRT_MML` is `OUT_names` which is in essence a copy of the input data set `names`, but with scores, item parameter estimates and their standard errors added. This data set can be used directly as the input data set `names` for the macro `%LIRT_ICC`. Another output data set from `%LIRT_MML`



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

is `OUT_logl` containing the likelihood value for the fitted model. This makes it straightforward to compare various models with likelihood ratio tests. Goodness of fit can be further evaluated (graphically) by comparing observed data to data simulated under the model using `%LIRT_simu`. The outputs are listed below. For `%LIRT_ICC` the main output is of course plots of the ICC's.

### Output data sets of `%LIRT_MML`:

- `OUT_names`: a model-specifying data set holding information about the items, including the estimated item parameters.
- `OUT_disc`: data set with estimated item discriminations.
- `OUT_disc_std`: data set with standardized estimates of item discriminations.
- `OUT_ipar`: data set with estimated item parameters ($\eta's$).
- `OUT_thres`: data set with estimated item thresholds ($\beta's$).
- `OUT_poppar`: data set with estimated parameters of the 2-dimensional normal distribution of the latent variable (only available when `dim=2`).
- `OUT_logl`: the value of the loglikelihood for the fitted model.
- `OUT_conv`: data set with the convergence status of the numerical estimation.

### Output data sets of `%LIRT_PPAR`:

- `OUT_ppar`: data with estimated person parameters.
- `OUT_logl`: the value of the loglikelihood for the fitted model.
- `OUT_conv`: data set with the convergence status of the numerical estimation.

### Output data sets of `%LIRT_ICC`:



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

- OUT_plot: a data set relevant information on the ICC's.

Output data sets of %LIRT_SIMU:

- OUT_simu1, OUT_simu2,... : simulated data sets with item responses. As many as the specified with the option ndata.
- OUT_names: a model-specifying data set holding information about the items.

### 4.1 Additional SAS macros

Some additional SAS macros facilitating the use of the implementation are available. The SAS macro %lirt_split can be used for splitting locally dependent items. More generally it can be used to split any variable in a given input data set according to another. This means that it can be used for recoding items in analyses of differential item functioning (DIF). Two SAS macros %LIRT_SIMU_names and %lirt_pdata, generate input data sets for %LIRT_SIMU. The first one creates the input data set names and is particularly convenient when local dependence across time points is to be specified for some items. The second one create the input data set pdata specifying the distribution of the latent variable. Finally the SAS macro %LIRT_MML_names generates the input data set names for the SAS macro %LIRT_MML.

## 5 Example: longitudinal data from the COS-BC

Many women participating in screening mammography experience a false positive result. Most of these women will experience negative psychosocial consequences. The Psychological Consequences Questionnaire (PCQ) Cockburn et al. (1992) is a questionnaire designed to measure psychological consequences of screening mammography. The Consequences Of Screening in Breast Cancer (COS-BC) Brodersen & Thorsen (2008) is an adaptation and translation of this instrument to a Danish setting with subscales measuring among others anxiety, sense of dejection and sleep.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

We consider responses to four polytomous items intended to measure sleep problems collected at two time points in women who underwent screening for breast cancer. The data set SLEEP contains responses to the four sleep items at two time points. The first 10 records look as follows

```
 idnr    sleepba1   fallasl1   wokenup1   awake1   sleepba2   fallasl2   wokenup2   awake2

110001       0          0          0         0         1          1          1         1
110002       0          0          0         0         0          0          0         0
110003       0          0          0         0         0          0          0         0
110004       1          1          1         2         2          1          1         1
110005       0          0          0         0         0          0          0         0
110006       0          1          2         0         .          .          .         .
110007       0          0          0         0         0          0          0         0
110008       2          2          2         0         .          .          .         .
110009       2          1          0         0         1          0          0         0
110010       3          3          0         3         3          3          3         2
```

Note that not everyone participates at time 2. The ID variable idnr illustrates that each person is represented by one record only, but the variable itself is not required in any of the input data sets of the SAS macros. The response options are $0, 1, 2$ and 3 (where 0 is 'Not at all' and 3 is 'A lot'). The item wording and the marginal item frequencies at the two time points are shown in Table 1.

| Item wording | | Response options | | | | | |
|---|---|---|---|---|---|---|---|
| (SAS variable name) | Time | not at all | a bit | quite a bit | a lot | Total | Missing |
| I have been awake | 1 | 1096 (85.0%) | 112 (8.7%) | 57 (4.4%) | 24 (1.9%) | 1289 | 29 |
| most of the night | 2 | 932 (86.5%) | 90 (8.4%) | 36 (3.3%) | 19 (1.8%) | | 241 |
| (awake) | | | | | | | |
| It has take me a | 1 | 938 (72.9%) | 180 (14.0%) | 90 (7.0%) | 78 (6.1%) | 1286 | 32 |
| long time to fall | 2 | 852 (81.1%) | 114 (10.9%) | 52 (5.0%) | 32 (3.1%) | | 268 |
| asleep (fallasl) | | | | | | | |
| I have slept | 1 | 918 (71.2%) | 196 (15.2%) | 106 (8.2%) | 69 (5.4%) | 1289 | 29 |
| badly | 2 | 825 (78.6%) | 135 (12.9%) | 59 (5.6%) | 31 (3.0%) | | 268 |
| (sleepba) | | | | | | | |
| I have woken up far too | 1 | 1005 (77.7%) | 152 (11.8%) | 76 (5.9%) | 61 (4.7%) | 1294 | 24 |
| early in the morning | 2 | 876 (81.5%) | 117 (10.9%) | 41 (3.8%) | 41 (3.8%) | | 243 |
| (wokenup) | | | | | | | |

Table 1: Item wording and marginal distribution at the two time points for the four sleep items.

We start by carrying out one-dimensional analyses of each time point separately, investigating whether the PCM (3 or the GPCM (4) is more appropriate, and as an initial evaluation if item parameters are stable across the two time points. Subsequently, a longitudinal model is considered.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

## 5.1 Separate analyses for each time point

The data set `names` specifying the one-dimensional 2PL model should contain the variables name, max and disc_yn. It can be created using the SAS code

```
data names1;
    input name $ max disc_yn $;
    datalines;
    awake1   3  Y
    fallasl1 3  Y
    sleepba1 3  Y
    wokenup1 3  Y
   ;
run;
```

here disc_yn=Y for all items meaning that they are modeled as GPCM items. The model is fitted using the macro call

```
%LIRT_MML(  DATA=SLEEP,
            NAMES=names1,
            DIM=1,
            OUT=gpcm1);
```

The maximum likelihood value attained is stored in the data set gpcm1_logl. The item parameters estimates are also stored in data sets, the discrimination parameters are in data sets gpcm1_disc and model1_disc_std and the threshold parameters in the data set model1_thres. These data sets are printed in Appendix A1. Letting disc_yn=N for all items specifies the Rasch model, and a comparison of the estimated likelihood values yields a likelihood ratio test of the Rasch model against the more general GPCM model. Thus we adjust the names data set

```
data names1_rm;
    set names1;
    disc_yn='N';
run;
```

and fit the PCM



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

```
%LIRT_MML(  DATA=SLEEP,
            NAMES=names1_rm1,
            DIM=1,
            OUT=rm1);
```

and we can test whether this simpler model is justified for the items using a likelihood ratio test

```
proc sql;
    select Value into :_llrm from rm1_logl;
    select Value into :_ll from gpcm1_logl;
quit;
data _lrt;
    lrt=-(&_ll-&_llrm);
    df=3;
    p=1-cdf('chisquared',lrt,df);
run;
proc print data=_lrt round noobs;
run;
```

The PMC is clearly rejected, with a $\chi^2$ value of 72.9 on 3 degrees of freedom and we stick to the GPCM specified by names1. Similar analyses for time point two also rejected the PCM (results not shown). The estimated item parameters are shown in Tables 2 and 3 together with item parameters from a separate analysis of the time 2 data.

for all thresholds we see a very small difference (of the magnitude 0.1 to 0.2) across the two time points, whereas the discrimination parameters appear to change for two of the four items. This is confirmed by a visual comparison of ICC's. When item parameters have been estimated these can be plotted using the SAS macro %LIRT_ICC and for this purpose we use the data set gpcm1_names in which the item names and estimated parameters are stored and plot the ICC's with the macro call

```
%LIRT_icc(  NAMES=gpcm1_names,
            DIM=1,
            OUT=icc1);
```

| | | Time 1 | | | Time 2 | | |
|---|---|---|---|---|---|---|---|
| Item | | Est. | 95% CI | Std. est. | Est. | 95% CI | Std. est. |
| awake | $\beta_1$ | 1.19 | (1.08, 1.30) | -0.11 | 1.22 | (1.12, 1.32) | -0.23 |
| | $\beta_2$ | 1.51 | (1.38, 1.64) | 0.21 | 1.62 | (1.48, 1.76) | 0.17 |
| | $\beta_3$ | 2.06 | (1.86, 2.26) | 0.76 | 2.02 | (1.82, 2.23) | 0.57 |
| fallas | $\beta_1$ | 0.71 | (0.63, 0.80) | -0.59 | 0.98 | (0.89, 1.06) | -0.48 |
| | $\beta_2$ | 1.19 | (1.09, 1.29) | -0.11 | 1.43 | (1.32, 1.54) | -0.03 |
| | $\beta_3$ | 1.49 | (1.37, 1.61) | 0.19 | 1.78 | (1.63, 1.93) | 0.33 |
| sleepba | $\beta_1$ | 0.62 | (0.55, 0.70) | -0.67 | 0.85 | (0.78, 0.93) | -0.60 |
| | $\beta_2$ | 1.15 | (1.06, 1.24) | -0.15 | 1.37 | (1.27, 1.47) | -0.08 |
| | $\beta_3$ | 1.57 | (1.45, 1.69) | 0.27 | 1.76 | (1.62, 1.90) | 0.30 |
| wokenup | $\beta_1$ | 1.13 | (0.99, 1.28) | -0.16 | 1.23 | (1.08, 1.38) | -0.22 |
| | $\beta_2$ | 1.35 | (1.19, 1.50) | 0.05 | 1.60 | (1.42, 1.78) | 0.14 |
| | $\beta_3$ | 1.60 | (1.41, 1.79) | 0.30 | 1.59 | (1.38, 1.81) | 0.14 |
| | $\mu$ | 0 | | -1.30 | 0 | | -1.46 |

Table 2: Standardized and unstandardized threshold parameter estimates estimated separately at the two time points.

| | | time 1 | | | time 2 | | |
|---|---|---|---|---|---|---|---|
| Item | | Est. | 95% CI | Std. est. | Est. | 95% CI | Std. est. |
| awake | $\alpha$ | 4.03 | (3.23, 4.84) | 0.98 | 5.06 | (3.86, 6.26) | 0.97 |
| fallas | $\alpha$ | 4.68 | (3.74, 5.62) | 1.14 | 6.38 | (4.77, 7.99) | 1.55 |
| sleepba | $\alpha$ | 6.73 | (4.92, 8.54) | 1.64 | 9.18 | (5.87,12.48) | 2.23 |
| wokenup | $\alpha$ | 2.25 | (1.88, 2.61) | 0.55 | 2.47 | (2.01, 2.94) | 0.60 |
| | $\sigma$ | 1 | | 4.11 | 1 | | 5.20 |

Table 3: Standardized and unstandardized item discrimination parameter estimates estimated separately at the two time points.

using a data set gpcm1_names created by the macro `%LIRT_MML`. This yields ICC's that are shown in Figure 2 together with ICC's from time two obtained by a similar macro call for time 2

We can estimate the person locations for the ten first records using the macro call

```
%LIRT_ppar( DATA=SLEEP(where=(idnr<110011)),
            NAMES=gpcm1_names,
            DIM=1,
            OUT=pp1,
            QPOINTS=30);
```

the estimated person locations are saved in a data set pp1_latent. Edited output looks like this

| | TIME 1 | | | | TIME 2 | | | |
|---|---|---|---|---|---|---|---|---|
| Parameter | Estimate | Standard Error | Lower | Upper | Estimate | Standard Error | Lower | Upper |
| theta1 | | | | | 1.2328 | 0.1480 | 0.9312 | 1.5343 |
| theta2 | | | | | | | | |
| theta3 | | | | | | | | |
| theta4 | 1.1398 | 0.1619 | 0.8125 | 1.4671 | 1.4062 | 0.1314 | 1.1385 | 1.6740 |
| theta5 | | | | | | | | |
| theta6 | 0.7253 | 0.2005 | 0.3202 | 1.1305 | | | | |
| theta7 | | | | | | | | |
| theta8 | 1.2802 | 0.1564 | 0.9641 | 1.5963 | | | | |
| theta9 | 1.0407 | 0.1716 | 0.6940 | 1.3874 | 0.8760 | 0.1627 | 0.5445 | 1.2075 |
| theta10 | 1.8475 | 0.2300 | 1.3828 | 2.3123 | 2.0641 | 0.2163 | 1.6235 | 2.5047 |

### 5.1.1 Tests of fit

Beyond the test likelihood ratio of the GPCM against the PCM further tests of fit can be done.

Tests of Differential item Functioning (DIF) ) by splitting an item.

Christensen and Olsbjerg (2013) proposed a visual evaluation of item fit based on comparison of observed and expected item means across values of the sum of all items. An example of the rsultaing GOF plot is illustrated fir the longitudinal model.

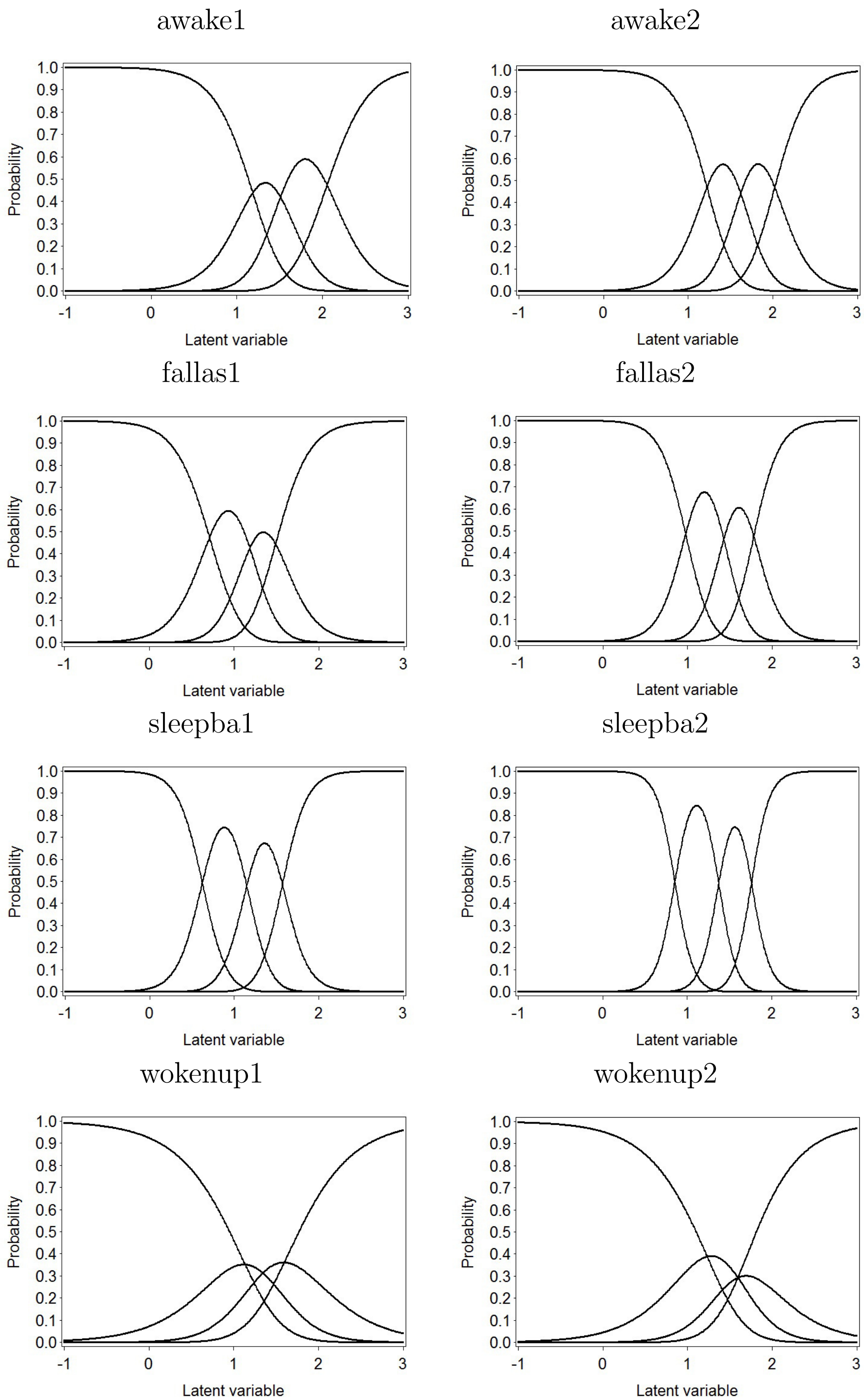


Figure 2: Item characteristic curves, thresholds and discrimination parameters are estimated separately at the two time points.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

## 5.2 Longitudinal analysis

We now take into account the longitudinal nature of the data and specify the longitudinal GPCM (18) for the four sleep items

```
data names3;
    input name1 $ name2 $ max disc_yn $;
    datalines;
    awake1 awake2 3 Y
    fallasl1 fallasl2 3 Y
    sleepba1 sleepba2 3 Y
    wokenup1 wokenup2 3 Y
    ;
run;
```

Note that for longitudinal data the columns representing items at time point 1 and 2 are called name1 and name2, respectively. The SAS macro identifies items by their order in this names data set. This is a model without any item parameter drift or local dependence across time points, i.e., $I_0 = I$ and $J_0 = I$. Models allowing for local dependence across time points and/or item parameter drift can be specified by changing the names data set. We fit the simple longitudinal model with no drift or dependence with the macro call

```
%LIRT_MML(  DATA=SLEEP,
            NAMES=names5,
            DIM=2,
            OUT=LGPCM);
```

Estimates of the item parameters (the $\eta$'s) stored in the data set LGPCM_ipar (printed in Appendix A) and the item thresholds (the $\beta$'s) stored in the data set LPGPCM_thres. The item discriminations (the $\alpha$'s) stored in model3_disc

| Parameter | Estimate | Standard Error | Lower | Upper |
|---|---|---|---|---|
| awake1 | 4.2444 | 0.3468 | 3.5641 | 4.9246 |
| awake2 | 4.2444 | 0.3468 | 3.5641 | 4.9246 |
| fallasl1 | 4.9996 | 0.4089 | 4.1973 | 5.8018 |
| fallasl2 | 4.9996 | 0.4089 | 4.1973 | 5.8018 |



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

| | | | | |
|---|---|---|---|---|
| sleepba1 | 6.3798 | 0.6002 | 5.2024 | 7.5572 |
| sleepba2 | 6.3798 | 0.6002 | 5.2024 | 7.5572 |
| wokenup1 | 2.3170 | 0.1569 | 2.0093 | 2.6247 |
| wokenup2 | 2.3170 | 0.1569 | 2.0093 | 2.6247 |

and finally the parameters of the latent distribution stored in LGPCM_poppar

| Parameter | Estimate | Standard Error | Lower | Upper |
|---|---|---|---|---|
| mu | -0.3428 | 0.07291 | -0.4858 | -0.1998 |
| rho | 0.8279 | 0.01896 | 0.7907 | 0.8651 |
| sigma2 | 1.1219 | 0.06464 | 0.9950 | 1.2487 |

telling us that the mean decreases and that the latent correlation $\rho = \mathrm{Corr}(\theta_1, \theta_2) = 0.83$ is substantial.

### 5.2.1 Model allowing for local dependence across time points

A model allowing for local dependence between the items awake1 and awake2 across time points (i.e., $I_0 = \{\mathtt{sleepba}, \mathtt{fallas}, \mathtt{wokenup}\}$ can be fitted using by splitting the item fallasl2 for dependence. This can be done using the 'help' macro %LIRT_SPLIT that requires the simple data set names3 as input. The macro call

```
%LIRT_SPLIT(DATA=SLEEP,
            NAMES=names_2pl,
            INDEP=awake1,
            DEP=awake2)
```

splits the item(s) in DEP according to the item(s) in INDEP in the order they appear. The macro call creates a data set called SLEEP_split that is a copy of the original data with the addition of the items resulting from the item split. The first ten records in this data set that includes the new items cat0_awake2, cat1_awake2, cat2_awake2 and cat3_awake2 looks as follows

| idnr | sleepba1 | fallasl1 | wokenup1 | awake1 | sleepba2 | fallasl2 | wokenup2 | awake2 | cat0_ awake2 | cat1_ awake2 | cat2_ awake2 | cat3_ awake2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 110001 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | . | . | . |
| 110002 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | . | . | . |
| 110003 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | . | . | . |



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 110004 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | . | . | 1 | . |
| 110005 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | . | . | . |
| 110006 | 0 | 1 | 2 | 0 | . | . | . | . | . | . | . | . |
| 110007 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | . | . | . |
| 110008 | 2 | 2 | 2 | 0 | . | . | . | . | . | . | . | . |
| 110009 | 2 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | . | . | . |
| 110010 | 3 | 3 | 0 | 3 | 3 | 3 | 3 | 2 | . | . | . | 2 |

using the data set SLEEP_split we can specify the model using

```
data names4;
    input name1 $ name2 $ max disc_yn $;
    datalines;
    awake1      .           3  Y
    .           cat0_awake2 3  Y
    .           cat1_awake2 3  Y
    .           cat2_awake2 3  Y
    .           cat3_awake2 3  Y
    fallasl1    fallasl2    3  Y
    sleepba1    sleepba2    3  Y
    wokenup1    wokenup2    3  Y
    ;
run;
```

Note that the original item awake2 is not included. With the dependence assumption built into names4 we simply fit the model with the macro call

```
%LIRT_MML(  DATA=SLEEP,
            NAMES=names4,
            DIM=2,
            OUT=split);
```

and as before the maximum likelihood value, the discrimination parameters, and the threshold parameters are stored in data sets. We can test for local dependence across time points by comparing the likelihood values in the two data sets split_logl and lgpcm_logl.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

### 5.3 Model allowing for item parameter drift

Another alternative to the simple longitudinal GPCM model specified using names3 is a model allowing for item parameter drift for the item awake, i.e.,

$$J_0 = \{fallas, sleepba, wokenup\}$$

can be specified using

```
data names5;
    input name1 $ name2 $ max disc_yn $;
    datalines;
    awake1 . 3 Y
    . awake2 3 Y
    fallasl1 fallasl2 3 Y
    sleepba1 sleepba2 3 Y
    wokenup1 wokenup2 3 Y
    ;
run;
```

We fit this model with the macro call

```
%LIRT_MML(  DATA=SLEEP,
            NAMES=names5,
            DIM=2,
            OUT=drift);
```

using the original data set and we can test for local dependence across time points by comparing the likelihood values in the two data sets drift_logl and lgpcm_logl

As for the unidimensional ICC curves are plotted for all items with the call

```
%LIRT_MML(  NAMES=model3_names,
            DIM=2,
            OUT=model3);
```

For items with no drift the ICC is only printed once. For the item fallasl the ICC's are shown in Figure 3

As before a data set, model3_plot, with the corresponding data points is available.

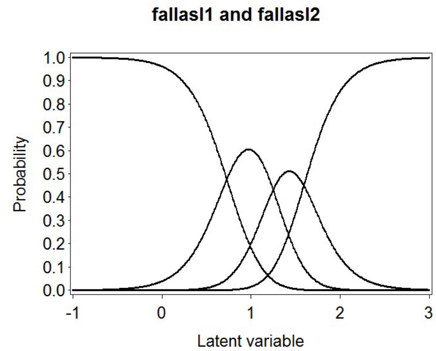


Figure 3: ICC's for an anchored item in the longitudinal model.

### 5.4 Models allowing for local dependence across time points and item parameter drift

More general models can be specified and fitted in order to assess possible item drift and local dependence across time points by adjusting the data set names, by changing the structure of the variables name1 and name2. For instance a model with local dependence across time points for the item awake and item drift for the item sleepba can be specified in the names data set

```
data names;
    input name1 $ name2 $ max disc_yn $;
    datalines;
    awake1     .          3 Y
    .        cat0_awake2 3 Y
    .        cat1_awake2 3 Y
    .        cat2_awake2 3 Y
    .        cat3_awake2 3 Y
    fallasl1 fallasl2    3 Y
    sleepba1 .           3 Y
    .        sleepba2    3 Y
    wokenup1 wokenup2    3 Y
    ;
run;
```

followed by the macro call



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

```
%LIRT_MML(  DATA=SLEEP,
            NAMES=names,
            DIM=2,
            OUT=model4)
```

### 5.5 Tests of fit

Beyond the likelihood ratio tests testing local dependence over time points and item parameter drift further tests of individual fit can be done by comparing observed and expected mean scores

The mean of the item awake1 across values of the sum score can be computed using the statements

```
%let it1=awake1 fallasl1 sleepba1 wokenup1;
data SLEEP;
 set SLEEP;
 score1=sum(of awake1 fallasl1 sleepba1 wokenup1);
run;
proc means data=SLEEP;
 var awake1;
 class score1;
 output out=means mean=mean;
run;
```

and the a plot of the observed item mean across the values of the sum of all items can be created using the SAS code

```
axis1 order=0 to 12 by 1 value=(H=2) minor=NONE label=(H=2);
axis2 value=(H=2) minor=NONE label=(H=2 A=90);
proc gplot data=means;
 plot mean*score1 / haxis=axis1 vaxis=axis2;
 symbol v=none i=join w=3 l=1 color=black;
run;
```

this plot is shown in Figure 4. In order to simulate expected values under the simple longitudinal model with the assumptions (20) and (21) we specify



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

```
data pd;
    set lgpcm_names3;
    LD_GROUP=.;
    LD_ITEM='';
run;
data pdata;
    input PARAMETER $ ESTIMATE;
    datalines;
    mu1 0
    mu2 -0.3094
    sigma1 1
    sigma2 1.0830
    rho 0.8133
    ;
run;
```

and simulate four data sets using the SAS macro `%LIRT_SIMU`

```
%LIRT_SIMU( NAMES=PD,
            DIM=2,
            NDATA=4,
            NPERSONS=1289,
            PDATA=PDATA,
            OUT=s);
```

this macro call creates four data sets s_simu1, s_simu2, s_simu3 and s_simu4. An item mean plot in first simulated data set can be obtained using

```
data gof1;
 set s_simu1;
 score1=sum(of &it1);
run;
proc means data=gof1;
 var awake1;
 class score1;
 output out=means mean=mean;
run;
```



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

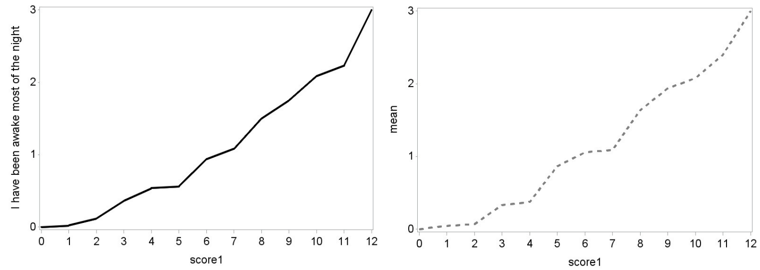


Figure 4: The observed item mean plot and a single simulated simulated curve.

```
axis1 order=0 to 12 by 1 value=(H=2) minor=NONE label=(H=2);
axis2 value=(H=2) minor=NONE label=(H=2 A=90);
proc gplot data=means;
 plot mean*score1 / haxis=axis1 vaxis=axis2;
 symbol v=none i=join w=3 l=33 color=grey;
run;
```

This plot is shown in Figure 4.

A GOF plot can be obtained using

```
data s0; set SLEEP; dataset=0; run;
data s1; set s_simu1; dataset=1; score1=sum(of &it1); run;
data s2; set s_simu2; dataset=2; score1=sum(of &it1); run;
data s3; set s_simu3; dataset=3; score1=sum(of &it1); run;
data s4; set s_simu4; dataset=4; score1=sum(of &it1); run;
data gof; set s0-s4; run;
proc means data=gof;
 var awake1;
 class score1 dataset;
 output out=means mean=mean;
run;
axis1 order=0 to 12 by 1 value=(H=2) minor=NONE label=(H=2);
axis2 value=(H=2) minor=NONE label=(H=2 A=90);
proc gplot data=means(where=(dataset ne .));
 plot mean*score1 = dataset / haxis=axis1 vaxis=axis2;
 symbol1 v=none i=join w=3 l=1 color=black;
```



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

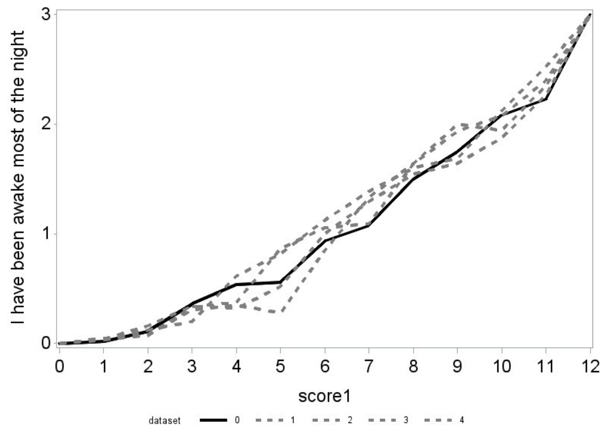


Figure 5: The GOF plot comparing the observed item means with four curves based on data sets simulated under the model.

```
 symbol2 v=none i=join w=3 l=33 color=grey r=4;
run;
```

This plot is shown in Figure 5.

## 6 Additional SAS macros

Some additional SAS macros for facilitating the use the implementation are available: (i) the SAS macro `%lirt_pdata` that generates the input data set pdata for the SAS macro `%LIRT_SIMU`; (ii) the SAS macro `%lirt_names_mml` that generates a NAMES data set that can be used as input data set for the SAS macro `%LIRT_MML`.

## Discussion

The proprietary software packages RUMM (Andrich et al., 2010) and WINSTEPS (Linacre, 2011) for fitting Rasch models are widely used. These fit unidimensional models only, even though many applications deal with multidimensional or longitudinal data.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

The twodimensional Rasch model as originally dicussed by Andersen (1985) and Embretson (1991) was formulated for longitudinal data.

To obtain consistent item parameter estimates marginal maximum likelihood estimation (Bock & Aitkin, 1981; Thissen, 1982; Zwinderman & van den Wollenberg, 1990) is used. This approach to item parameter estimation assumes that the latent variables are sampled from a population and introduces an assumption about the distribution of the latent variable.

The implementation allows the user to specify the model structure in a separate statements using the NAMES data sets.

The general implementation allows the user to specify models where the item parameters do change over time. The SAS macros can be used to test the assumption of item parameter invariance using likelihood ratio tests, thus adding to existing methods for detection of item parameter drift (Donoghue & Isham, 1998; DeMars, 2004; Galdin & Laurencelle, 2010).

The SAS macros also make it possible to study local dependence across time points, by splitting of the item at follow-up into new items according to the responses given at baseline (Olsbjerg & Christensen, 2013b). The macro %lrasch_mml makes it possible include splitted items and to test the assumption local independence across time points using likelihood ratio tests, thus adding to existing tests of this assumption (Olsbjerg & Christensen, 2013a).

For the PCM Christensen and Olsbjerg (2013) proposed a visual evaluation of item fit based on comparison of observed and expected item means across values of the sum of all items. Using the proposed SAS macros this idea is generalized in from the PCM to the GPCM and from unidimensional models for a single time point to longitudinal models.

# References

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

Adams, R. J., Wilson, M., & Wu, M. (1997b). Multilevel item response models: An approach to errors in variables regression. *Journal of Educational and Behavioral Statistics*, *22*, 47–76.

Andersen, E. (1977). Sufficient statistics and latent trait models. *Psychometrika*, *42*, 69–81.

Andersen, E. B. (1973). Conditional inference for multiple-choice questionnaires. *British Jornal of Mathematical and Statistical Psychology*, *26*, 31//44.

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

Andrich, D., Sheridan, B., & Luo, G. (2010). *RUMM2030 [Computer software and manual]*. RUMM Laboratory, Perth, Australia.

Birnbaum, A. L. (1968). Latent trait models and their use in inferring an examinee's ability. In F. M. Lord, & M. R. Novick (Eds.) *Statistical Theories of Mental Test*. Reading, MA: Addison-Wesley.

Blanchin, M., Hardouin, J., Le Neel, T., Kubis, G., Blanchard, C., Mirallié, E., & Sébille, V. (2011). Comparison of ctt and Rasch-based approaches for the analysis of longitudinal patient reported outcomes. *Statistics in Medicine*, *30*, 825–838.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: an application of an em algorithm. *Psychometrika*, *46*, 443–459.

Brodersen, J., & Thorsen, H. (2008). Consequences of screening in breast cancer (cos-bc): Development of a questionnaire. *Scandinavian Journal of Primary Health Care*, *26*, 251–256.

Christensen, K. B. (2006). Fitting polytomous Rasch models in sas. *Journal of Applied Measurement*, *7*, 407–417.

Christensen, K. B. (2013). Conditional maximum likelihood estimation in polytomous Rasch models using sas. *ISRN Computational Mathematics*, *Article ID 617475*, 8 pages.

Christensen, K. B., & Bjorner, J. B. (2003). Sas macros for Rasch based latent variable modelling (research report 03/13). Tech. rep., Department of Biostatistics, University Of Copenhagen.

Christensen, K. B., Kreiner, S., & Mesbah, M. (2013). *Rasch Models in Health*. Hoboken, NJ: John Wiley & Sons.

Christensen, K. B., & Olsbjerg, M. (2013). Marginal maximum likelihood estimation in polytomous Rasch models using sas. *Pub. Inst. Stat. Univ. Paris*, *57, fasc. 1-2*, 69–84.

Cockburn, J., De, L., Hurley, S., & Clover, K. (1992). Development and validation of the pcq: A questionnaire to measure the psychological consequences of screening mammography. *Soc Sci Med*, *34*, 1129–1134.

DeMars, C. E. (2004). Detection of item parameter drift over multiple test administrations. *Applied Measurement in Education*, *17*, 265–300.

Donoghue, J. R., & Isham, S. P. (1998). A comparison of procedures to detect item parameter drift. *Applied Psychological Measurement*, *22*, 33–51.

Embretson, S. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika*, *56*, 495–516.

Fischer, G. H., & Molenaar, I. W. (1995). *Rasch Models: Foundations, Recent Developments, and Applications*. New York: Springer-Verlag.

Galdin, M., & Laurencelle, L. (2010). Assessing parameter invariance in item response theory's logistic two item parameter model: A monte carlo investigation. *Tutorials in Quantitative Methods for Psychology*, *6*, 39–51.

Hardouin, J.-B., & Mesbah, M. (2007). The sas macro-program %anaqol to estimate the parameters of item responses theory models. *Communications in Statistics - Simulation and Computation*, *36*, 437–453.

Linacre, J. M. (2011). *WINSTEPS Rasch measurement computer program*. winsteps.com, Beaverton, Oregon.

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

Muraki, E. (1992). A generalized partial credit model: Application of an em algorithm. *Applied Psychological Measurement*, *16*, 159–176.

Nandakumar, R., & Hotchkiss, L. (2012). PROC NLMIXED: For estimating parameters of irt models. *Applied Psychological Measurement*, *36*, 60–63.

Neyman, J., & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrika*, *16*, 1–32.

Olsbjerg, M., & Christensen, K. B. (2013a). Marginal and conditional approaches to longitudinal Rasch models. *Pub. Inst. Stat. Univ. Paris*, *57, fasc. 1-2*, 109–126.

Olsbjerg, M., & Christensen, K. B. (2013b). Modeling local dependence in longitudinal IRT models. Research Report 14, Department of Statistics, University of Copenhagen.

Pastor, D. A., & Beretvas, S. N. (2006). Longitudinal rasch modeling in the context of psychotherapy outcomes assessment. *Applied Psychological Measurement*, *30*, 100–120.

Rasch, G. (1960). *Probabilistic Models for some Intelligence and Attainment Tests*. Copenhagen: Danish National Institute for Educational Research.

Raudenbush, S. W., Bryk, A. G., Cheong, Y. F., & Congdon, R. (2004). *HLM 6: Hierarchical linear and nonlinear modeling [Computer software]*. Scientific Software International, Chicago.

Reeve, B. B., Hays, R. D., JB, B., Cook, K. F., Crane, P. K., Teresi, J. A., Thissen, D., Revicki, D. A., Weiss, D. J., Hambleton, R. K., Liu, H., Gershon, R., Reise, S. P., Lai, J. S., Cella, D., & Group, P. C. (2007). Psychometric evaluation and calibration of health-related quality of life item banks. plans for the patient-reported outcomes measurement information system (promis). *Medical Care*, *45, Suppl 1*.

Rijmen, F., & De Boeck, P. (2002). The random weights linear logistic test model. *Applied Psychological Measurement*, *26*, 271–285.

Rijmen, F., Tuerlinckx, F., De Boeck, P., & Kuppens, P. (2003). A non-linear mixed model framework for item response theory. *Psychological Methods*, *8*, 185–205.

Smits, D., & De Boeck, P. (2003). Estimation of the mirid: A program and a sas-based approach. *Behavior Research Methods, Instruments and Computers*, *35*, 537–549.

Tennant, A., & Conaghan, P. (2007). The Rasch measurement model in rheumatology: What is it and why use it? when should it be applied, and what should one look for in a Rasch paper? *Arthritis & Rheumatism (Arthritis Care & Research)*, *57*, 1358–1362.

Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika*, *47*, 175–186.

van der linden, W. J., & Hambleton, R. K. (1997). *Handbook of Modern Item Response Theory*. New York: Springer-Verlag.

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). *ACER ConQuest Version 2: Generalised item response modelling software*. Australian Council for Educational Research, Camberwell.

Zwinderman, A. H., & van den Wollenberg, A. L. (1990). Robustness of marginal maximum likelihood estimation in the Rasch model. *Applied Psychological Measurement*, *14*, 73–81.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26740v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018