Research Article

Statistics in Medicine

Received 4 June 2012, Accepted 6 November 2012 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5692

Longitudinal measurement in health-related surveys. A Bayesian joint growth model for multivariate ordinal responses

Josine Verhagen*† and Jean-Paul Fox

Longitudinal surveys measuring physical or mental health status are a common method to evaluate treatments. Multiple items are administered repeatedly to assess changes in the underlying health status of the patient. Traditional models to analyze the resulting data assume that the characteristics of at least some items are identical over measurement occasions. When this assumption is not met, this can result in ambiguous latent health status estimates. Changes in item characteristics over occasions are allowed in the proposed measurement model, which includes truncated and correlated random effects and a growth model for item parameters. In a joint estimation procedure adopting MCMC methods, both item and latent health status parameters are modeled as longitudinal random effects. Simulation study results show accurate parameter recovery. Data from a randomized clinical trial concerning the treatment of depression by increasing psychological acceptance showed significant item parameter shifts. For some items, the probability of responding in the middle category versus the highest or lowest category increased significantly over time. The resulting latent depression scores decreased more over time for the experimental group than for the control group and the amount of decrease was related to the increase in acceptance level. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: Bayesian hierarchical modeling; IRT; Latent variable models; longitudinal data; MCMC; measurement invariance; mental health; survey data

1. Introduction

Repeatedly administered questionnaires are increasingly used in clinical studies to assess the effects of interventions on or to track changes in reported physical or mental health status. Examples are quality of life questionnaires for patients with chronic diseases, psychiatric questionnaires to follow patients with psychiatric disorders, and cognitive ability questionnaires to track the onset of Alzheimer's disease. In this type of studies, a questionnaire is administered at each occasion to measure the underlying health status. The resulting data can be characterized as longitudinal multivariate response data designed to track intra-individual changes in latent physical or mental health status.

When the outcome consists of multiple item responses, it is common to use a latent variable to model the dependency between the item responses. The multivariate item responses are assumed to be conditionally independently distributed given a common latent variable. The latent variable and item characteristics are used to specify the probability of each item response, which defines the measurement model. Examples are common factor models for continuous responses (e.g. [1]), multiple indicators and multiple causes (MIMIC) models (e.g. [2]), item response models for discrete responses (e.g. [3]), and latent class models for discrete latent variables (e.g. [4, 5]). In this paper, we will focus mainly on item response theory (IRT) models, which are becoming more and more popular for health questionnaire data (e.g. [6–8])

Mixed random effects models have been proposed for modeling individual change over time [9–11]. The specification of random effects accounts for the dependency between the within-subject measurements. Furthermore, random effects provide a very flexible way of handling missing data within

Department of Research Methodology, Measurement and Data Analysis, Faculty of Behavioral Sciences, University of Twente, Drienerlolaan 5, Enschede, The Netherlands

**Correspondence to: Josine Verhagen, Department of Research Methodology, Measurement and Data Analysis, Faculty of Behavioral Sciences, University of Twente, Drienerlolaan 5, Enschede, The Netherlands.*

†E-mail: a.j.verhagen@gw.utwente.nl

subjects, as the number of measurement occasions per subject is allowed to vary. In addition, individual differences in the onset of the measurements and in the time span between measurement occasions can be taken into account [12].

In longitudinal surveys consisting of multiple administered items per measurement occasion, latent growth models are often used to model longitudinal growth in the latent variable using mixed random effects [13–16]. In latent growth curve models, for example, the random effects structure allows individual growth trajectories to vary randomly around a general growth trend. Longitudinal item response models (e.g. [17, 18]) have appeared particularly useful for analyzing longitudinal health-related questionnaires. Using a latent variable has the additional advantage that not all respondents necessarily have to answer all items at each occasion.

A common assumption in latent growth models is that the parameters of the measurement part, which connects the latent variable to the observed responses, are invariant over occasions. However, due to either the process of test administration, developmental changes between measurement occasions, or other occasion-dependent influences, the characteristics of the test can change over time. As a result, the relation between the observed responses and the latent variable will differ per occasion. Consequently, a constant level of underlying health status leads to different response probabilities at sequential occasions.

Invariance over measurement occasions of item characteristics relating the latent variable to the response probabilities will be referred to as longitudinal measurement invariance, a form of measurement invariance [19]. In a factor analysis measurement model, for example, this implies that intercepts and factor loadings are invariant over occasions [20]. In the field of IRT, non-invariance of discrimination and difficulty parameters over measurement occasions is known as item parameter drift [21, 22]. It is well known that longitudinal measurement invariance is not self-evident and should be tested for (e.g. [12, 20, 23]). Nevertheless, in most applications longitudinal measurement invariance of at least some items is assumed.

Implications of an erroneous assumption of longitudinal measurement invariance are that latent variable means and variances can be affected, rendering estimates of changes in latent health status ambiguous to interpret. To avoid bias in latent growth estimates due to violations of measurement invariance assumptions, latent growth models for longitudinal (health) questionnaire data should allow for occasion-specific measurement characteristics.

Procedures have been developed to assess measurement invariance for cross-sectional response data (e.g. [24, 25]), which are usually based on fixed item parameters. In the present approach, longitudinal measurement invariance will be assessed using a random item effects structure that allows the modeling of longitudinal growth in item characteristics. A nonlinear model for longitudinal multivariate responses will be introduced, which models growth in both the parameters of the measurement part and in latent health status simultaneously. The random item effects multilevel IRT model developed by [26] (see also [27–29]) will be extended with multivariate and covarying random item parameter effects as well as with growth structures to incorporate change over time in both latent health status and item parameters. In this model, the measurement and latent health parameters will be treated as crossed random effects (e.g. [30, 31]). Moreover, it is possible to include time-varying, person or item level covariates to explain variation in health status and item parameters. MCMC methods will be used for inference.

Advantages of this approach are that it will enable the growth modeling of latent health given a very flexible occasion-specific measurement model, not assuming item parameters to be invariant over measurement occasions. The more realistic occasion-specific measurement models will increase the accuracy of the latent health status estimates. In addition, information acquired about item parameter shifts will provide more insight in measuring latent health status longitudinally. As a case study, data from a randomized clinical trial concerning the treatment of depression will be analyzed.

In Section 2, we describe the model in more detail. In Section 3, we describe the estimation procedure and ways to explore longitudinal invariance. In Section 4, we show the results of a simulation study to assess parameter recovery and present the results of the randomized clinical trial concerning the treatment of depression. We give a discussion in Section 5.

2. A joint random effects growth model for longitudinal multivariate discrete responses

Health-related questionnaires often consist of a set of items with categorical or ordinal response categories. It will be assumed that the items under analysis measure one unidimensional latent construct, which represents some form of physical or mental health status.

Statistics in Medicine

The probability of each response Y_{ijk} can be perceived as a function of the latent health variable θ_{ij} of person $i = 1, \ldots, I$ at occasion $j = 1, \ldots, J$ and of the parameters of the measurement part of the model $\tilde{\boldsymbol{\xi}}_{kj}$ for item $k = 1, \ldots, K$ at occasion j.

To account for the cross-classified nested structure of the data, where occasions are nested in persons and items, random effects structures will be imposed to model the dependency structure. This results in two cross-cutting hierarchies: A three-level structure of observations Y_{ijk} within occasions j within items k referred to as the measurement part of the model and another three-level structure of observations Y_{ijk} within occasions j within persons i, referred to as the latent variable part of the model. Within these hierarchical structures, growth models can be implemented, as well as fixed or random covariates to explain variance within and between persons, items, or occasions (Figure 1).

2.1. *Occasion-specific measurement models for categorical responses*

The basis of the measurement model presented here was developed by [32] and [33] for educational data and can be seen as an extension to the polytomous logistic model for multinomial responses or a discrete choice model [34]. The probability that a subject with latent health status θ responds with category c depends on threshold parameters of the item categories and on a discrimination parameter, which can be unique for each item.

In this generalized partial credit model (GPCM), the conditional probability that person i chooses the cth category over the $c-1$st category is modeled with a dichotomous logistic model (see also [7]). In a longitudinal framework, the GPCM can be extended by including occasion j:

$$\frac{P(Y_{ijk} = c \mid Z_{ijck})}{P(Y_{ijk} = c \mid Z_{ijck}) + P\left(Y_{ijk} = (c-1) \mid Z_{ijck}\right)} = \frac{\exp^{Z_{ijck}}}{1 + \exp^{Z_{ijck}}},$$

which can also be written as

$$\frac{P(Y_{ijk} = c \mid Z_{ijck})}{P\left(Y_{ijk} = (c-1) \mid Z_{ijck}\right)} = \exp^{Z_{ijck}}.$$

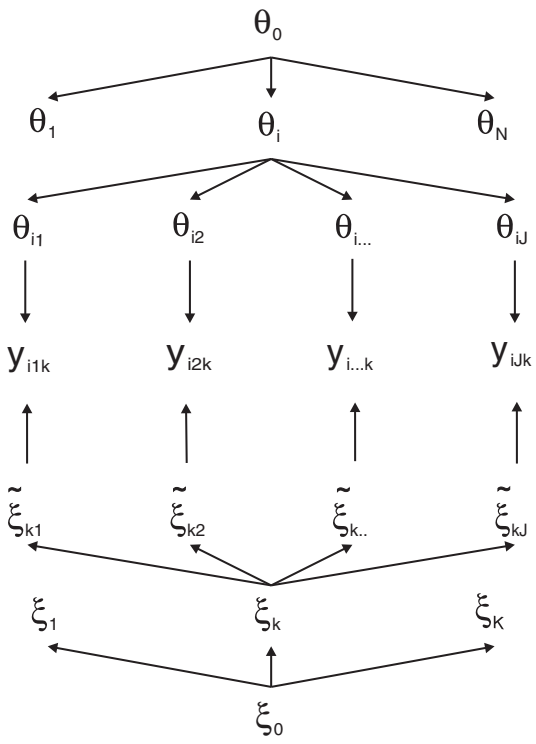

Figure 1. Directed acyclic graph of the joint random effects growth model. The observed scores y_{ijk} for person i on item k at occasion j are a function of the unobserved latent person scores θ_{ij} and item characteristics $\tilde{\xi}_{kj}$.

As $\sum_0^C P(Y_{ijk} = c|Z_{ijck})$ equals 1, the probability of response c given Z_{ijck} is given by

$$P(Y_{ijk} = c \mid Z_{ijck}) = \frac{\exp\left(\sum_0^c (Z_{ijck})\right)}{\sum_0^C \exp\left(\sum_0^c (Z_{ijck})\right)}. \tag{1}$$

The Z_{ijck} will be modeled in such a way that respondents with a higher latent score θ_{ij} are more likely to score in a higher response category. Furthermore,

$$Z_{ijck} = \tilde{a}_{kj}\left(\theta_{ij} - \tilde{b}_{ckj}\right). \tag{2}$$

The threshold parameters $\tilde{b}_{ckj}$ are the points on the latent scale at which the category response functions of $P(Y_{ijk} = c|\theta)$ and $P(Y_{ijk} = (c-1)|\theta)$ intersect (Figure 2). When the latent health level increases beyond this point, the probability of responding with c becomes higher than the probability of responding with $(c-1)$. The discrimination parameter $\tilde{a}_{kj}$ is related to the discrimination between categorical responses as the latent health level changes. The higher the discrimination parameter, the steeper the category response functions, and the crisper the response categories discriminate between higher and lower latent health levels.

To account for changes in the measurement characteristics over time, Equation (2) contains occasion-specific item parameters. For each item, these parameters will be modeled as random effects varying around general longitudinally invariant item parameters. In this way, the parameters of all items are allowed to vary over occasions. The occasion-specific item parameters are exchangeable given the general item parameters and form a second level between the observed responses (level 1) and the invariant general item parameters (level 3).

Because they relate to the same item, the occasion-specific parameters per item are assumed to be correlated. The random item parameters will be assumed to be multivariate normally distributed with an unstructured covariance matrix. This novel way of modeling the GPCM parameters captures within-item correlations. It follows that the random occasion-specific item parameters $\tilde{\boldsymbol{\xi}}_{kj} = (\tilde{a}_{kj}, \tilde{b}_{0kj}, \ldots, \tilde{b}_{Ckj})$ are assumed to be multivariate normally distributed given the general item parameters $\boldsymbol{\xi}_k = (a_k, b_{0k}, \ldots, b_{Ck})$:

$$\tilde{\boldsymbol{\xi}}_{kj} \mid \boldsymbol{\xi}_k, \boldsymbol{\Sigma}_{\tilde{\xi}_k} \sim \mathcal{N}\left(\boldsymbol{\xi}_k, \boldsymbol{\Sigma}_{\tilde{\xi}_k}\right), \tag{3}$$

where

$$\boldsymbol{\Sigma}_{\tilde{\xi}_k} = \begin{bmatrix} \sigma^2_{a_k} & \sigma_{a_k b_{0k}} & \cdots & \sigma_{a_k b_{Ck}} \\ \sigma_{b_{0k} a_k} & \sigma^2_{b_{0k}} & \cdots & \sigma_{b_{0k} b_{Ck}} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{b_{Ck} a_k} & \sigma_{b_{Ck} b_{0k}} & \cdots & \sigma^2_{b_{Ck}} \end{bmatrix}. \tag{4}$$

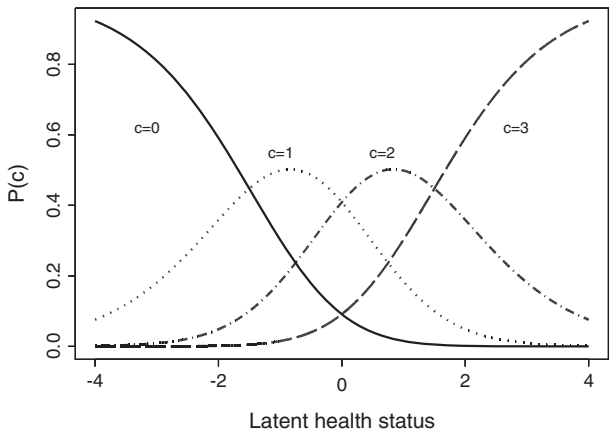

Figure 2. Item category response curves: probability of a response c $(c = 0, 1, 2, 3)$ given latent health status θ $(\theta \in [-4, 4])$.

Following a hierarchical prior structure, at the third level the general item parameters are assumed to be multivariate normally distributed given mean parameters $\boldsymbol{\mu}_\xi = (a_0, b_{00}, \ldots, b_{C0})$:

$$\boldsymbol{\xi}_k \mid \boldsymbol{\mu}_\xi, \boldsymbol{\Sigma}_\xi \sim \mathcal{N}(\boldsymbol{\mu}_\xi, \boldsymbol{\Sigma}_\xi). \tag{5}$$

The extension of the GPCM with occasion-specific random item parameters following a multivariate hierarchical structure is a natural extension in this Bayesian modeling framework. For model identification additional restrictions are necessary, which will be described in Section 2.4. When the item parameters are longitudinally invariant, the occasion-specific parameters are all restricted to be equal to the general item parameters. In that case the item characteristics are a priori distributed according to Equation (5).

2.2. *A growth model for item characteristic change*

It is natural to assume that change in item parameters over occasions is not completely random, but follows a growth pattern. This change can be modeled with a random linear time effect of the average time passed at occasion j over subjects. Depending on the total number of occasions available, characteristics can be added to explain linear and higher order change in item parameters over occasions, for example, through polynomial time effects. Let $\mathbf{v}_j$ denote a vector of explanatory information at occasion j, and $\boldsymbol{\delta}_k$ a vector of item-specific coefficients predicting the occasion-specific item parameters $\tilde{\boldsymbol{\xi}}_{kj}$. Then, a conditional growth model can be specified:

$$\tilde{\boldsymbol{\xi}}_{kj} = \boldsymbol{\xi}_k + \mathbf{v}_j \boldsymbol{\delta}_k + \boldsymbol{\epsilon}_{\xi_{kj}}, \; \boldsymbol{\epsilon}_{\xi_{kj}} \sim \mathcal{N}\left(0, \boldsymbol{\Sigma}_{\tilde{\xi}_k}\right), \tag{6}$$

where $\boldsymbol{\Sigma}_{\tilde{\xi}_k}$ is given by Equation (4). Its diagonal elements denote the conditional variance in occasion-specific item parameters given the explanatory information. The off-diagonal terms denote the within-item covariance between item characteristics.

2.3. *A growth model for latent health status*

The change over occasions in latent health status of person i can be modeled by a random intercept (β_{0i}) and random linear time effect (β_{1i}) of the time passed since the individual's starting point t_{i0},

$$\theta_{ij} = \beta_{0i} + (t_{ij} - t_{i0})\beta_{1i} + e_{ij}.$$

The growth model can be easily extended by time-varying covariates with random effects (including polynomial effects) denoted as $\mathbf{x}_{ij}$, time-varying covariates with fixed effects (including main time effects) denoted as $\mathbf{s}_{ij}$, and person-level covariates that do not vary over occasions, denoted as $\mathbf{w}_i$. Then, the more general representation of the latent variable growth part of the model is given by

$$\begin{aligned} \theta_{ij} &= \mathbf{x}_{ij}^t \boldsymbol{\beta}_i + \mathbf{s}_{ij}^t \boldsymbol{\zeta} + e_{ij}, e_{ij} \sim \mathcal{N}(0, \sigma_j) \\ \boldsymbol{\beta}_i &= \mathbf{w}_i^t \boldsymbol{\gamma} + \boldsymbol{\upsilon}_i, \boldsymbol{\upsilon}_i \sim \mathcal{N}(0, \mathbf{T}). \end{aligned} \tag{7}$$

An occasion-specific residual variance parameter was specified to model the unexplained variability per measurement occasion. This residual variance parameter will be included in identifying restrictions described in Section 2.4.

2.4. *Model identification*

Two issues need to be addressed. First, the latent scales for the different occasions need to be linked. Separate models for each occasion result in incomparable scales between occasions. Second, there is no unique solution to Equation (2), as multiple parameter combinations result in the same likelihood. For each occasion, a shift in the latent mean μ_j results in the same expected response probabilities as a shift of all threshold parameters $\tilde{b}_{ckj}$ in the opposite direction. A similar identification problem exists for the discrimination parameters $\tilde{a}_{kj}$ and the occasion-specific residual variance of the latent variable σ_j.

Assuming that an overall shift in the health responses over measurement occasions is more likely the result of a change in latent health status than of a mean change in threshold parameters over occasions, the following restriction is imposed: $\sum_k \sum_c \tilde{b}_{ckj} = 0$ for each j. For each occasion, the mean of the thresholds $\sum_k b_{kj}$ is fixed to the arbitrary value of 0, where $b_{kj} = \sum_c \tilde{b}_{ckj}$. Constraining the sum of

the threshold parameters to be equal in both groups links the occasion-specific scales, while fixing this sum to 0 identifies the model. For each occasion, the product of the discrimination parameters is fixed to 1: $\prod_k \tilde{a}_{kj} = 1$ for each j. This expresses the assumption that it is more likely for the latent health variance to change over time than for all items to discriminate equally more or less. In addition to these elementary identification constraints, the random effects occasion-specific item parameters will shrink toward the general item parameters. The variance of item parameters over occasions will indicate the degree to which the items are non-invariant.

The identification constraints deviate from the usual constraints for the GPCM for incomplete designs. Traditionally, so-called 'anchor' items that have identical parameters over occasions are used to link the scales. A reference occasion with fixed mean and variance is used to identify the scales. Restricting the sum of the thresholds and the product of the discrimination parameters is a very natural way of identification in a random effects framework, however, as it allows the variance components to be estimated freely. This is not possible when restricting specific distributional parameters. Problems would arise with the specification of proper prior distributions and the results would be hard to re-scale. In addition, unrestricted covariance components make it much easier to extend the model to include, for example covariates, to explain variance in both the latent variable and the item parameters.

3. Estimation and inference

Combining all parts of the model described previously, the implied conditional model can be defined by inserting Equation (6) and (7) into Equations (2) and (1). Hence, the likelihood model can be represented by

$$P(Y_{ijk} = c|\boldsymbol{\xi}_{kj}, \theta_{ij}) = \frac{\exp\left(\sum_0^c (Z_{ijck})\right)}{\sum_0^C \exp\left(\sum_0^c (Z_{ijck})\right)},$$

where

$$Z_{ijck} = \left(a_k + \mathbf{v}_j \boldsymbol{\delta}_{ak} + \epsilon_{a_{kj}}\right)\left((\mathbf{x}_{ij}\boldsymbol{\beta}_i + \mathbf{s}_{ij}\boldsymbol{\zeta} + e_{ij}) - \left(b_{ck} + \mathbf{v}_j \boldsymbol{\delta}_{ck} + \epsilon_{b_{ckj}}\right)\right).$$

3.1. Estimation

In the hierarchical modeling approach, the parameters at each level are conditionally independent given the parameters on the higher level. The resulting full posterior is therefore a product of the likelihood and the hierarchical priors at each level:

$$p\left(\boldsymbol{\theta}, \tilde{\boldsymbol{\xi}}, . \mid \mathbf{Y}\right) \propto \left[\prod_i \left[\prod_j \left[\prod_k p\left(y_{ijk} \mid \tilde{\boldsymbol{\xi}}_{kj}, \theta_{ij}\right) p\left(\tilde{\boldsymbol{\xi}}_{kj} \mid .\right)\right] p\left(\theta_{ij} \mid .\right)\right]\right],$$

with conditional hierarchical priors for the latent variables $p(\theta_{ij} \mid .)$ and for the random item parameters $p\left(\tilde{\boldsymbol{\xi}}_{kj} \mid .\right)$. The hierarchical prior incorporating the mixed effects model on the latent variable is constructed as follows:

$$p(\theta_{ij} \mid .) = p(\theta_{ij} \mid \mathbf{x}_{ij}, \beta_i, \mathbf{s}_{ij}, \zeta, \sigma_j) p(\boldsymbol{\beta}_i \mid \mathbf{w}_i, \boldsymbol{\gamma}, \mathbf{T}) p(\boldsymbol{\zeta} \mid \boldsymbol{\mu}_\zeta, \boldsymbol{\Sigma}_\zeta) p(\sigma_j),$$

and the hierarchical prior defining the growth model for the item parameter

$$p\left(\tilde{\boldsymbol{\xi}}_{kj} \mid .\right) = p\left(\tilde{\boldsymbol{\xi}}_{kj} | \boldsymbol{\xi}_k, \boldsymbol{\delta}_k, \mathbf{v}_j, \boldsymbol{\Sigma}_{\tilde{\xi}_k}\right) p\left(\boldsymbol{\xi}_k \mid \boldsymbol{\mu}_\xi, \boldsymbol{\Sigma}_\xi\right) p\left(\boldsymbol{\delta}_k \mid \boldsymbol{\mu}_\delta, \boldsymbol{\Sigma}_\delta\right) p\left(\boldsymbol{\Sigma}_{\tilde{\xi}_k}\right).$$

To estimate the model parameters, an MCMC sampling method will be used. The first sampling step for both person and item parameters will be a Metropolis–Hastings step, which samples correlated and truncated group-specific item parameters. The hierarchical prior parameters will be sampled with a Gibbs sampler. Conjugate hyperprior distributions will be used, which define a hierarchy of normal distributions. The means and variances of the mean parameters $\boldsymbol{\mu}_\zeta, \boldsymbol{\gamma}, \boldsymbol{\mu}_\xi$ and $\boldsymbol{\mu}_\delta$ will be drawn from normal-inverse Wishart distributions. The hyperpriors for the variance terms σ_j will be drawn from inverse gamma distributions and the hyperpriors for the covariance matrices $\boldsymbol{\Sigma}_{\tilde{\xi}_k}$ from inverse Wishart distributions. The full sampling scheme can be found in Appendix B. Software in the form of Splus or R code combined with Fortran.dll file is available on request from the corresponding author.

3.2. Exploring longitudinal invariance

The proposed model offers many opportunities to explore whether the item parameters are longitudinally invariant. We will focus on two tests of longitudinal measurement invariance, a Bayes factor test focused on invariance of the separate items, and the deviance information criterion (DIC) for the comparison of models with and without invariance restrictions.

3.2.1. Bayes factor. The variance and covariance components of $\boldsymbol{\Sigma}_{\tilde{\xi}_k}$ can be used to evaluate whether the occasion-specific parameters are invariant over measurement occasions, and whether the discrimination and threshold parameters covary over time. To test longitudinal invariance of each item parameter, a Bayes factor can be computed to compare the marginal likelihood of the nested models with and without invariance of the parameter (see also [35]).

In case of nested models, the Bayes factor reduces to the ratio of the density (region) in accordance with the null hypothesis under the prior and posterior distribution for the most general model [36, 37]. Under the conditions that both models share the same conditional distribution of observed data and the parameter space associated with the prior under the null hypothesis, denoted as Θ_0, is a subset of the parameter space associated with the prior under the alternative hypothesis, Θ_1, the Bayes factor can be evaluated as the posterior expectation of the ratio of prior densities.

Let $p\left(\sigma^2_{b_k} \mid H_0\right), \sigma^2_{b_k} \in \Theta_0$ and $p\left(\sigma^2_{b_k} \mid H_1\right), \sigma^2_{b_k} \in \Theta_1$, where $\Theta_0 \subseteq \Theta_1$, denote the prior under the null hypothesis and the prior under the alternative hypothesis, respectively. The Bayes factor in favor of the null hypothesis can be expressed as [38]

$$\begin{aligned}
\mathrm{BF} &= \frac{\int_{\Theta_0} p\left(\sigma^2_{b_k} \mid H_0\right) p\left(\mathbf{y} \mid \sigma^2_{b_k}\right) \mathrm{d}\sigma^2_{b_k}}{\int_{\Theta_1} p\left(\sigma^2_{b_k} \mid H_1\right) p\left(\mathbf{y} \mid \sigma^2_{b_k}\right) \mathrm{d}\sigma^2_{b_k}} = \frac{\int_{\Theta_0} p\left(\sigma^2_{b_k} \mid H_0\right) p\left(\mathbf{y} \mid \sigma^2_{b_k}\right) \mathrm{d}\sigma^2_{b_k}}{p\left(\mathbf{y} \mid H_1\right)} \\
&= \int_{\Theta_1} \left[\frac{p\left(\sigma^2_{b_k} \mid H_0\right)}{p\left(\sigma^2_{b_k} \mid H_1\right)}\right] \frac{p\left(\sigma^2_{b_k} \mid H_1\right) p\left(\mathbf{y} \mid \sigma^2_{b_k}\right)}{p\left(\mathbf{y} \mid H_1\right)} \mathrm{d}\sigma^2_{b_k} \\
&= \int_{\Theta_1} \left[\frac{p\left(\sigma^2_{b_k} \mid H_0\right)}{p\left(\sigma^2_{b_k} \mid H_1\right)}\right] p\left(\mathbf{y} \mid \sigma^2_{b_k}, H_1\right) \mathrm{d}\sigma^2_{b_k} \\
&= E\left[\frac{p\left(\sigma^2_{b_k} \mid H_0\right)}{p\left(\sigma^2_{b_k} \mid H_1\right)} \mid \mathbf{y}, \right].
\end{aligned} \tag{8}$$

The ratio of prior densities is evaluated using MCMC samples from the marginal posterior density of $\sigma^2_{b_k}$ under the alternative hypothesis, given that the ratio is bounded on Θ_1. Because the specific null hypothesis of a variance component of zero is on the boundary of the parameter space, the null hypothesis will be specified as $H_0 : \sigma^2 < \delta$, where δ is a very small number which is chosen to represent a small difference in the context of the problem under analysis [39–41].

The Bayes factor is the ratio of the support for the two models given the data and the prior information. A Bayes factor higher than 1 indicates more evidence for the null hypothesis, and a Bayes factor lower than 1 indicates more support for the alternative hypothesis.

3.2.2. DIC. The DIC of [42] will be used to compare the fit of models with and without invariant item parameters. The deviance function will be defined as

$$D\left(\boldsymbol{\Lambda}\right) = -2 \log \left[\prod_i \left[\prod_j \left[\prod_k p\left(Y_{ijk} \mid \tilde{\boldsymbol{\xi}}_{kj}, \theta_{ij}\right)\right]\right]\right]. \tag{9}$$

The DIC consists of the posterior mean of the deviance corrected for the number of parameters in the model. In hierarchical models, the number of parameters is hard to determine, which is solved in the DIC by computing the effective number of parameters p_D. The p_D is computed by subtracting the deviance at the posterior means from the posterior mean of the deviance.

The DIC is given by

$$\begin{aligned} \text{DIC} &= \overline{D\left(\boldsymbol{\Lambda}\right)} + \overline{D\left(\boldsymbol{\Lambda}\right)} - D\left(\hat{\boldsymbol{\Lambda}}\right) \\ &= \overline{D\left(\boldsymbol{\Lambda}\right)} + p_D, \end{aligned}$$

where $\overline{D\left(\boldsymbol{\Lambda}\right)}$ is the posterior mean deviance and $D\left(\hat{\boldsymbol{\Lambda}}\right)$ the estimated deviance at the posterior estimate of $\hat{\Lambda}$.

Longitudinal measurement invariance can be tested by comparing two of the models with occasion-specific item parameters with the DIC of the measurement invariant model. Both models have to be estimated.

4. Results

4.1. Simulation study: parameter recovery

To check whether parameter recovery is accurate under the proposed estimation procedure, we performed a simulation study. We generated a data set containing 800 cases with 10 measurement occasions each from a model with both latent (Equation (7)) and item parameter (Equation (6)) growth. We generated measurement occasion times for each case, with varying starting points and time intervals between the occasions.

We used normal distributions to draw person-specific means ($N(0, 0.5)$) and person-specific latent time effects ($N(0.2, 0.5)$). We drew within-person latent variable values from normal distributions with the person-specific means and person specific variances from an inverse gamma distribution $IG(15, 1/15)$. We also drew the general item parameters from normal distributions, with means equal to 1 for the discrimination parameters and means equal to $\pm 2/3$ for the threshold parameters. We simulated the occasion-specific item parameters to be normally distributed around the general item parameters with variances $\sigma^2_{b_{ckj}} = 0.02$ and $\sigma^2_{a_{kj}} = 0.04$. We randomly assigned time effects to each item parameter, varying between 0 and 0.5. We ran a single long MCMC chain with 1000 burn-in iterations and 10 000 final iterations.

In Table I, an illustration of the results can be found. For all general item parameters except for the lower category of item 18, the true value fell within two standard errors of the posterior mean and the posterior means did not differ systematically from the true values. The correlations between the true values and the posterior means of the occasion-specific parameters, as well as the correlations between the true values and the posterior means of the growth parameters were all above 0.94.

4.2. Application: intervention effects on depression level

We applied the proposed model to a study on the effects of guided self-help based on Acceptance Commitment Therapy (ACT) [43]. We recruited participants through advertisements in Dutch newspapers requesting people who want more out of their life but are hindered by depressive or anxiety symptoms. We excluded respondents with very few symptoms, as well as respondents with a severe disorder, respondents already receiving treatment, and respondents with a high risk of suicide. We randomly assigned the remaining 376 participants to one of three conditions: ACT intervention with minimal or extensive email support (250) and a waiting list (116) condition.

In the two experimental conditions, we administered questionnaires at five moments during the study: at the start of the study and after 3, 6, 9, and 20 weeks. The respondents in the control condition only answered the questionnaires at the start of the study and after 9 weeks. We assumed that missing occasion measurements (2 for 12 experimental and 1 for 21 experimental and 3 control group members) were missing randomly.

The aim of the data collection was to investigate whether the ACT intervention reduced depression and anxiety. To measure depression, we used the Center for Epidemiologic Studies Depression (CES-D) questionnaire [44], which consists of 20 items measuring symptoms of depression experienced in the last week on a four-point scale (seldom or never, i.e. less than 1 day; sometimes, i.e. 1–2 days; often, i.e. 3–4 days; almost always, i.e. 5–7 days). For many items, we observed few or no responses in the highest answer category. Therefore, for all items, we collapsed this category with the third category. This last category now indicates the experience of the symptom for more than 3 days in the past week. One of the

Table I. Simulation study: true values and posterior means and standard errors of general item parameters ξ_k and correlations between the true values and posterior means of occasion-specific ($a_{kj}, b_{ckj}, \theta_{ij}$) and growth ($\delta_k, \beta_i$) parameters.

Item	a_k	EAP(a_k)	SE a_k	b_{1k}	EAP(b_{1k})	SE b_{1k}	b_{2k}	EAP(b_{2k})	SE b_{2k}
1	1.20	1.16	0.10	−0.82	−0.79	0.08	0.52	0.51	0.08
2	1.20	1.24	0.06	−0.13	−0.11	0.06	1.00	1.02	0.08
3	0.98	1.04	0.07	−1.05	−1.12	0.09	−0.73	−0.79	0.06
4	0.80	0.76	0.08	−0.11	−0.06	0.07	1.37	1.33	0.08
5	0.98	0.93	0.06	−1.50	−1.52	0.07	1.58	1.61	0.07
6	0.80	0.87	0.08	−1.26	−1.25	0.09	−0.76	−0.87	0.06
7	1.04	1.09	0.08	−0.86	−0.81	0.06	1.55	1.56	0.07
8	0.98	0.94	0.08	−0.08	0.02	0.08	0.11	0.02	0.07
9	0.85	0.79	0.05	−0.53	−0.51	0.06	0.04	0.07	0.05
10	0.99	0.82	0.10	−1.54	−1.44	0.07	0.56	0.54	0.06
11	1.16	1.17	0.08	0.05	0.07	0.07	0.67	0.67	0.08
12	0.70	0.69	0.05	−0.08	−0.10	0.07	0.59	0.55	0.07
13	1.19	1.15	0.06	−0.24	−0.13	0.07	1.79	1.77	0.08
14	1.02	0.93	0.06	−0.74	−0.70	0.06	0.69	0.70	0.06
15	1.02	1.02	0.09	−1.24	−1.19	0.07	0.49	0.53	0.06
16	0.92	0.90	0.06	−0.69	−0.58	0.07	0.27	0.28	0.07
17	1.20	1.18	0.07	0.76	0.82	0.07	1.34	1.26	0.08
18	0.99	0.95	0.07	−1.37	−1.20	0.07	−0.12	−0.13	0.05
19	0.98	0.98	0.05	−0.47	−0.37	0.06	0.73	0.76	0.06
20	1.20	1.16	0.07	−0.26	−0.18	0.07	0.49	0.46	0.07
Correlations between true values and posterior means									
$\rho_{a_{kj}}$			0.94						
$\rho_{b_{ckj}}$			0.99						
ρ_{δ_k}			0.94						
$\rho_{\theta_{ij}}$			0.94						
ρ_{β_i}			0.97						

items generated almost no responses in the lowest category, and we removed this item for the analysis without threatening the test validity. The content of the CES-D items is given in Appendix A.

In addition to the effect of the ACT intervention, we investigated the hypothesis that the process of decreasing depression was mediated by a higher acceptance level [45, 46]. Acceptance is characterized by patients becoming more able to embrace and accept negative personal experiences instead of avoiding them [43]. Therefore, we measured the construct acceptance using the Acceptance and Action Questionnaire [47].

Besides the hypotheses concerning change in depression level, we tested the hypotheses regarding measurement invariance for each item. First, we determined whether the hypothesis of invariance held for the item parameters, as described in Section 3.2.1. Then, we investigated the hypotheses that the change in the non-invariant item parameters was described by a growth model (Section 2.2).

For all models, we ran a single long MCMC chain of 50 000 iterations, with a burn-in of 5000 iterations. The trace plots showed good convergence characteristics, and the convergence statistics (Geweke Z, autocorrelations) were satisfactory. We modeled the responses at the five measurement occasions with the occasion-specific GPCM (Equation (2.1)).

4.2.1. Latent growth model. First, the model component for the latent variable depression was modeled conditional on invariant item parameters, reducing the item parameter structure to the general item parameters (Equation (5)). To evaluate whether the depression level of the experimental group members decreased over time, and whether this decrease was related to the change in acceptance, three models were compared.

In the first model, denoted as M1, a basic discrete time effect on depression was specified. The common structure on the latent variable for the experimental and control group consisted of a random intercept and fixed occasion means,

$$\theta_{ij} = \beta_{0i} + \zeta_j Occasion_j + e_{ij},$$

where $e_{ij} \sim N(0, \sigma_j)$ and $\beta_{0i} \sim \mathcal{N}(\gamma_{00}, T_{00})$. The random intercept varied over individuals and specified the between-subject variability in average depression levels conditional on occasion-specific mean levels. An occasion-specific residual variance parameter was specified to model the unexplained variability per measurement occasion.

In the second model, denoted as M2, a latent growth model was implemented. For individuals measured on more than two occasions, the latent growth model included a random intercept β_{0i} and a random slope for a first (β_{1i}) and second (β_{2i}) order polynomial time effect,

$$\theta_{ij} = \beta_{0i} + \beta_{1i} Time_{ij} + \beta_{2i} Time_{ij}^2 + e_{ij}, \text{ with } e_{ij} \sim N(0, \sigma_j).$$

The random effects were assumed to be multivariate normally distributed, where the random intercepts and first order slopes were defined conditional on membership of the control ($Experimental = 0$) or experimental ($Experimental = 1$) group,

$$\begin{aligned} \beta_{0i} &= \gamma_{00} + \gamma_{01} Experimental_i + \upsilon_{0i} \\ \beta_{1i} &= \gamma_{10} + \gamma_{11} Experimental_i + \upsilon_{1i} \\ \beta_{2i} &= \gamma_{20} + \upsilon_{2i}, \end{aligned} \tag{10}$$

with $\boldsymbol{\upsilon}_i \sim N(0, \mathbf{T})$. The time-point zero corresponds with the first measurement occasion such that the random intercept variance specifies the between-subject variation in depression levels at the first measurement occasion. Furthermore, between-subject variation was specified over the subject-specific linear trend variable and quadratic time variable. The random effects were allowed to correlate using a common covariance matrix. For individuals assessed at just two measurement occasions (i.e. the control group), the latent trajectory was specified without the subject-specific quadratic time effect.

In the third model, denoted as M3, the subject-specific difference in acceptance between the first and fourth measurement occasion was used as an explanatory third-level variable, denoted as *Acceptance*. Subsequently, the random subject effects at level 3 are given by

$$\begin{aligned} \beta_{0i} &= \gamma_{00} + \gamma_{01} Experimental_i + \gamma_{02} Acceptance_i + \upsilon_{0i} \\ \beta_{1i} &= \gamma_{10} + \gamma_{11} Experimental_i + \gamma_{12} Acceptance_i + \upsilon_{1i} \\ \beta_{2i} &= \gamma_{20} + \upsilon_{2i}, \end{aligned} \tag{11}$$

with $\boldsymbol{\upsilon}_i \sim N(0, T)$.

In the upper part of Table II, the DIC of the three models is compared. The object is to select the best model given invariant item parameters such that in a next stage, the assumption of longitudinal measurement invariance can be evaluated. Conditional on invariant item parameters, the second and third models have a substantially lower DIC than the first model, which indicates that the subject-specific latent variable trajectory significantly improves the model fit. According to the DIC, the fit of model M3 does not increase relative to model M2. However, the DIC measures the fit of the random effects at level 1, which does not really change by adding acceptance as an explanatory variable. The common effect of acceptance on the linear trend parameter of the latent variable trajectory of depression is significant and large, around 1.06 (0.28), which shows that participants increasing their level of acceptance show a significant decrease in their level of depression. As a result, model M3 will be used to evaluate the longitudinal measurement invariance assumptions.

Table II. Deviance information criteria for models M1 to M6.

Model specification	$\hat{D}$	$\bar{D}$	p_D	DIC
Investigating latent growth model (fixed item parameters over occasions)				
M1: Discrete latent occasion effects	40191	41885	1694	43580
M2: Quadratic latent growth model with fixed condition effect	40116	41081	965	42047
M3: M2 corrected for change in Acceptance	40120	41082	962	42044
Investigating longitudinal measurement invariance (conditional on latent growth model M3)				
M4: M3 allowing item parameter change for all items	40228	41234	1007	42241
M5: M4 with items indicated as invariant ($BF > 3$) restricted	39968	40992	1024	42015
M6: M5 with growth model on non-invariant item parameters	39909	40976	1066	42041

4.2.2. Investigating longitudinal measurement invariance. In model M3, item parameters were restricted to be measurement invariant over occasions, reducing the item parameter structure to the general item parameters (Equation (5)). Given the latent growth structure of depression in model M3, this model was generalized by assuming all item parameters to be measurement non-invariant using the random item effects specification over time (Equation (3)), which will be referred to as model M4. A restricted version of model M4, denoted as model M5, consisted of some items restricted to be invariant and was used to evaluate partial longitudinal measurement invariance. In model M6, latent trajectories were added to the identified non-invariant item characteristics of model M5.

In the lower part of Table II, the DICs of the generalized models from the full invariant model M3 are given. The data do not support the assumption of full longitudinal measurement non-invariance, because the DIC of model M4 is lower than that of model M3. Subsequently, the measurement invariance assumption of each item was evaluated. Therefore, the multiple marginal null hypotheses of longitudinal measurement invariance were investigated using the Bayes factor specified in Equation (8) and by examining the item parameter variances over time.

In Table III, for each item the estimated posterior standard deviation over time of the three item parameters are given under the label σ_{a_k}, $\sigma_{b_{1k}}$, and $\sigma_{b_{2k}}$ respectively. A high Bayes factor value supports the null hypothesis of longitudinal measurement invariance. For items 5 and 19, the Bayes factors showed substantially more evidence for longitudinal variance in the highest item threshold parameters than for longitudinal invariance, as indicated by a Bayes factor lower than 0.33. The measurement invariance assumption was at least three times more likely than measurement non-invariance for the parameters of items 2, 8–10, and 12–16, for the threshold parameters of items 17 and 18, and for the discrimination parameters of items 1–3, 5, and 19.

Model M5 was defined as a partially restricted measurement invariant model, where the measurement invariant item parameters (i.e. tested to be invariant under model M4) were restricted to be invariant over time. This model M5 showed a better fit to the data than both the full invariant (M3) and the full non-invariant (M4) models according to the DICs represented in Table II. The test result supports the joint hypothesis of partial measurement invariance.

In model M6, latent trajectories with a linear trend component were defined for the non-invariant item characteristics. According to Equation (6), a latent trajectory was defined with a random intercept, defining the mean item characteristic level over time, and a time variable that models the trend over time.

Table III. For model M4, posterior variance estimates of random item characteristics over time and Bayes factor estimates concerning the marginal measurement invariance null hypotheses and for model M6, latent trajectory parameter estimates of the non-invariant item parameters.

	Model M4						Model M6					
Item	σ_{a_k}	BF	$\sigma_{b_{1k}}$	BF	$\sigma_{b_{2k}}$	BF	δ_{a_k}	$SE_{\delta_{a_k}}$	δ_{1k}	$SE_{\delta_{1k}}$	δ_{2k}	$SE_{\delta_{2k}}$
1	0.15	5.18	0.14	6.58	0.21	1.45			−0.01	0.05	0.09	0.06
2	0.14	6.23	0.14	6.59	0.20	2.89			0.02	0.05	0.06	0.06
3	0.18	3.87	0.17	4.69	0.20	2.89			−0.06	0.05	0.03	.06
4	0.14	6.77	0.16	4.73	0.15	4.78						
5	0.16	4.02	0.16	4.41	0.24	**0.32**			−0.03	0.06	0.09	0.06
6	0.29	1.88	0.41	1.03	0.23	3.30	0.13	0.08	**−0.28**	0.10	0.12	0.08
7	0.20	2.43	0.17	4.69	0.18	2.64	0.08	0.06	**−0.16**	0.08	0.05	0.06
8	0.14	6.77	0.15	6.17	0.14	6.79						
9	0.17	4.21	0.14	6.38	0.17	4.00						
10	0.14	7.30	0.13	8.02	0.17	4.18						
11	0.16	2.59	0.18	3.53	0.18	2.34	0.02	0.03	−0.10	0.06	0.06	0.05
12	0.15	5.27	0.16	4.79	0.19	3.10						
13	0.16	4.60	0.19	3.06	0.16	4.50						
14	0.14	7.18	0.14	6.43	0.15	6.06						
15	0.17	3.62	0.21	3.45	0.14	7.39						
16	0.16	4.66	0.14	6.46	0.19	3.55						
17	0.19	2.91	0.15	6.28	0.16	5.12	−0.02	0.05				
18	0.18	1.83	0.14	6.98	0.15	5.54	**−0.07**	0.03				
19	0.16	4.48	0.27	0.62	0.27	**0.16**			−0.08	0.05	**0.11**	0.06

The estimated posterior mean trend effects and the standard errors can be found in Table III under model M6. The discriminating effect of item 18 decreased significantly over time, whereas for items 6 and 7 the probability of responding with the second over the first category increased over time, and for item 19 the probability of responding with the third over the second category decreased over time. In Figure 3, the significant shift in probabilities for the answer categories over time is illustrated for item 19, where the category bounds as a function of depression level become more bold over time. This illustrates the trend of the middle category to become more probable over time, as the lower threshold shifts downward and the upper threshold shifts upward for each level of depression.

4.2.3. Latent developmental trajectory of depression. In Model M6, the subject-specific latent trajectory of depression was modeled conditional on the partial measurement invariant item parameters. Model M7 corresponds with model M6 but the explanatory variable change in *Acceptance* was excluded from

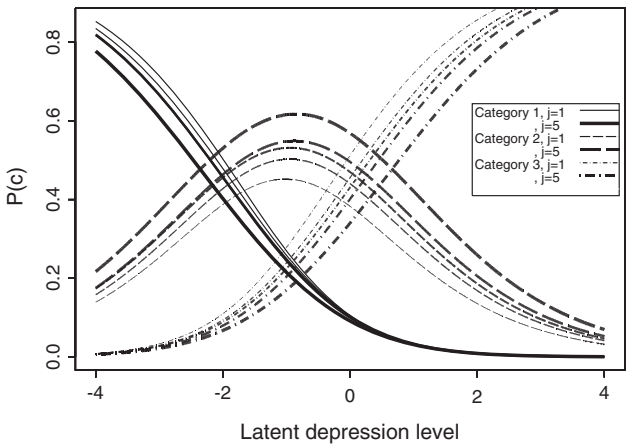

Figure 3. Item category response probabilities for item 19 on five measurement occasions.

Table IV. Conditional on partial longitudinal measurement invariance, population parameter estimates, and standard errors of the mean latent trajectory of depression with and without variable *Acceptance*.

		Model M7		Model M6	
	Parameter	EAP	SE	EAP	SE
Fixed effects					
Intercept	γ_{00}	0.13	0.07	0.13	0.07
	γ_{10}	**−0.28**	0.09	−0.15	0.09
	γ_{20}	**0.17**	0.02	**0.17**	0.02
Experimental	γ_{01}	−0.05	0.08	−0.06	0.08
	γ_{11}	**−0.84**	0.11	**−0.78**	0.11
Acceptance	γ_{02}			−0.03	0.23
	γ_{12}			**−1.06**	0.28
Random effects					
Residual variance level 2					
	σ_1	0.04	0.01	0.04	0.01
	σ_2	0.10	0.03	0.10	0.03
	σ_3	0.09	0.03	0.09	0.03
	σ_4	0.04	0.01	0.04	0.01
	σ_5	0.05	0.02	0.06	0.02
Residual variance level 3					
	T_{00}	0.52	0.05	0.52	0.05
	T_{11}	0.98	0.13	0.83	0.13
	T_{22}	0.12	0.03	0.11	0.03

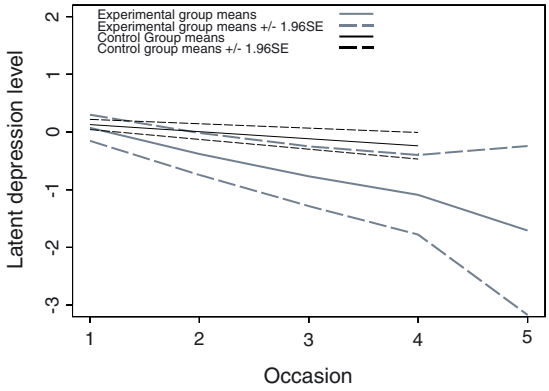

Figure 4. Mean latent growth patterns of depression for the experimental and control group.

the latent trajectory function. The mean latent trajectory parameter estimates of both models can be found in Table IV.

Model M7 shows a non-zero negative linear trend of depression for subjects in the control group ($\gamma_{10} = -0.28, SE = 0.09$), with a steeper negative mean trend for subjects in the experimental group ($\gamma_{11} = -0.84, SE = 0.11$). For the experimental group members, the negative linear mean trend is decelerated by a second order time effect ($\gamma_{20} = 0.17, SE = 0.02$). The estimated random effects variances show that there is heterogeneity in depression levels at the start of the study and significant between-subject variation in trend effects and in the decelerating effects of the squared time variable.

The inclusion of the level-3 variable 'change in acceptance' to explain heterogeneity in the subject-specific trajectory parameters (intercept and trend) annihilates the first order time effect for the control group ($\gamma_{10} = -0.15, SE = 0.09$), which is no longer significantly different from 0. It also slightly attenuates the negative linear mean trend in depression for the experimental group ($\gamma_{11} = -0.78, SE = 0.11$). The decrease of both effects indicates that part of the decrease of depression over time can be explained by an increase in level of change in acceptance. The same effect is also indicated by a decrease in the residual variance in first order slopes from 0.98 to 0.83, when conditioning on the change in acceptance. A strong effect was found of positive change in acceptance on decrease in depression ($\gamma_{12} = -1.06, SE = 0.28$). The other variance components were unaffected by the inclusion of acceptance. The small residual variances at level 2 show that models M6 and M7 explain most of the heterogeneity per measurement occasion.

Figure 4 shows the predicted occasion means of depression for the experimental and control group and 95% credible intervals. The mean depression level in the experimental group declined much steeper than that of the control group resulting in a lower average depression level at measurement occasion four.

5. Discussion

In this paper, we proposed a joint growth model with random occasion-specific parameters for both latent health and item parameters to measure changes in latent health status over time. Instead of assuming longitudinal invariance of the measurement model, we modeled variance in item parameters over measurement occasions with a growth model. The result is a multilevel growth model that reflects the effects of time characteristics on the occasion-specific item parameters and on the latent health construct simultaneously. Using the more flexible occasion-specific item parameters, change in the latent variable can be measured more accurately. In addition, the change in item parameters can provide insight into the process of longitudinal measurement. It can be used, for example, to detect and account for response shift (e.g. [29, 48]) by testing whether item parameters are changing over time and by explaining this shift with time-related variables.

A simulation study showed accurate parameter recovery for most of the general, the occasion-specific, and the growth parameters in both the latent and the item parameter growth models.

The analysis of a randomized trial on depression showed that the item parameters did change over measurement occasions. For some items, participants were more inclined to answer toward the middle category instead of the lowest or highest category over time. Conditional on this item parameter change,

there was an overall decrease in depression level for the experimental group, taking subject-specific change over time into account. The decrease in depression was stronger for the experimental group than for the control group. The change in depression level was steeper for the participants with a larger change in acceptance level.

A natural next step is to investigate what causes the shifts in item parameters and how to handle this in the best way. In addition, health questionnaires often consist of clusters of items, of which some may be more sensitive to change over time than others. To model this, it would be possible to include covariates with item characteristics for each item on the third level to predict differential time paths for items with, for example, specific content.

Appendix A. CES-D Questionnaire

Below is a list of the ways you might have felt or behaved. Please tell me how often you have felt this way during the past week:

Rarely or none of the time (less than 1 day)
Some or a little of the time (1–2 days)
Occasionally or a moderate amount of time (3–4 days)
Most or all of the time (5–7 days)

During the past week:

1. I was bothered by things that usually don't bother me.
2. I did not feel like eating; my appetite was poor.
3. I felt that I could not shake off the blues even with help from my family or friends.
4. I felt that I was just as good as other people.
5. I had trouble keeping my mind on what I was doing.
6. I felt depressed.
7. I felt that everything i did was an effort.
8. I felt hopeful about the future.
9. I thought my life had been a failure.
10. I felt fearful.
11. My sleep was restless.
12. I talked less than usual.
13. I felt lonely.
14. People were unfriendly.
15. I enjoyed life.
16. I had crying spells.
17. I felt sad.
18. I felt that people dislike me.
19. I could not get 'going'.

removed: I was happy.

Appendix B. MCMC Algorithm

The combination of truncated and correlated random effects with the identification restrictions described in Section 2.4 makes the model unfortunately not suitable for estimation in programs such as Winbugs or Mplus. The sampler was written in Fortran and can be called from Splus. It is available from the authors.

To sample the parameters of the joint growth model, an MCMC sampling scheme has been developed with a Metropolis–Hastings step for the sampling of the occasion-specific item and person parameters and a Gibbs sampler for the the higher level parameters.

At the $m+1$th iteration,

1. For each k, j, and c, sample a proposal $\tilde{\xi}^{*}_{ckj}$ $(c = 1,\ldots,C)$ from $N\left(\tilde{\xi}^{(m)}_{ckj}, \sigma^{2}_{mh\xi}\right)$, where $\sigma^{2}_{mh\xi}$ is tuned during the process to acquire an acceptance rate between 0.3 and 0.5. Let $\tilde{\boldsymbol{\xi}}^{-c}_{kj} = \left(\tilde{\xi}_{1kj},\ldots,\tilde{\xi}_{(c-1)kj},\tilde{\xi}_{(c+1)kj},\ldots,\tilde{\xi}_{Ckj}\right)$. The acceptance ratio R is the posterior probability ratio

of the proposed value $\tilde{\xi}^{*}_{ckj}$ and the previously sampled value $\tilde{\xi}^{(m)}_{ckj}$:

$$R = \frac{p\left(\mathbf{y}_{kj} \mid \tilde{\xi}^{*}_{ckj}, \boldsymbol{\theta}^{(m)}, \tilde{\boldsymbol{\xi}}^{(m)}_{kj}\right) p\left(\tilde{\xi}^{*}_{ckj} \mid \xi^{(m)}_{ck}, \delta^{(m)}_{ck}, \mathbf{v}_j, \tilde{\boldsymbol{\xi}}^{-c}_{kj}, \boldsymbol{\Sigma}^{(m)}_{\tilde{\xi}_k}\right)}{p\left(\mathbf{y}_{kj} \mid \tilde{\xi}^{(m)}_{ckj}, \boldsymbol{\theta}^{(m)}, \tilde{\boldsymbol{\xi}}^{(m)}_{kj}\right) p\left(\tilde{\xi}^{(m)}_{ckj} \mid \xi^{(m)}_{ck}, \delta^{(m)}_{ck}, \mathbf{v}_j, \tilde{\boldsymbol{\xi}}^{-c}_{kj}, \boldsymbol{\Sigma}^{(m)}_{\tilde{\xi}_k}\right)}.$$

A random uniform number u_{ckj} is drawn, and the proposal is accepted when $u_{ckj} \leq R$.

2. For each k and j and for each discrimination parameter $c = 0$, sample $\tilde{\xi}^{*}_{0kj}$ from $N\left(\tilde{\xi}^{(m)}_{0kj}, \sigma^2_{mh\xi}\right) I(\tilde{\xi}^{*}_{0kj} > 0)$. Because the proposal density is truncated, the acceptance ratio R is

$$R = \frac{p\left(\mathbf{y}_{kj} \mid \tilde{\xi}^{*}_{0kj}, \boldsymbol{\theta}^{(m)}, \tilde{\boldsymbol{\xi}}^{(m+1)}_{kj}\right) p\left(\tilde{\xi}^{*}_{0kj} \mid \xi^{(m)}_{ck}, \delta^{(m)}_{ck}, \mathbf{v}_j, \tilde{\boldsymbol{\xi}}^{(m+1)}_{kj}, \boldsymbol{\Sigma}^{(m)}_{\tilde{\xi}_k}\right)}{p\left(\mathbf{y}_{kj} \mid \tilde{\xi}^{(m)}_{0kj}, \boldsymbol{\theta}^{(m)}, \tilde{\boldsymbol{\xi}}^{(m+1)}_{kj}\right) p\left(\tilde{\xi}^{(m)}_{0kj} \mid \xi^{(m)}_{ck}, \delta^{(m)}_{ck}, \mathbf{v}_j, \tilde{\boldsymbol{\xi}}^{(m+1)}_{kj}, \boldsymbol{\Sigma}^{(m)}_{\tilde{\xi}_k}\right)} \frac{\boldsymbol{\Phi}\left(\frac{\tilde{\xi}^{(m)}_{0kj}}{\sigma_{mh}}\right)}{\boldsymbol{\Phi}\left(\frac{\tilde{\xi}^{*}_{0kj}}{\sigma_{mh}}\right)}.$$

A random uniform number u_{0kj} is drawn, and the proposal is accepted when $u_{0kj} \leq R$.

3. For each i, sample θ^{*}_{ij} from $N\left(\theta^{(m)}_{ij}, \sigma^2_{mh_\theta}\right)$. The acceptance ratio R is

$$R = \frac{p\left(\mathbf{y}_{ij} \mid \theta^{*}_{ij}, \tilde{\boldsymbol{\xi}}^{(m+1)}\right) p\left(\theta^{*}_{ij} \mid \mathbf{x}_{ij}, \boldsymbol{\beta}^{(m)}_i, \mathbf{s}_{ij}, \boldsymbol{\zeta}^{(m)}, \sigma^{(m)}_j\right)}{p\left(\mathbf{y}_{ij} \mid \theta^{(m)}_{ij}, \tilde{\boldsymbol{\xi}}^{(m+1)}\right) p\left(\theta^{(m)}_{ij} \mid \mathbf{x}_{ij}, \boldsymbol{\beta}^{(m)}_i, \mathbf{s}_{ij}, \boldsymbol{\zeta}^{(m)}, \sigma^{(m)}_j\right)}.$$

A random uniform number u_{ij} is drawn, and the proposal is accepted when $u_{ij} \leq R$.

4. For each k, sample the general item parameters $\boldsymbol{\xi}^{(m+1)}_k$ and time coefficients $\boldsymbol{\delta}^{(m+1)}_k$ from the full conditional

$$\text{vec}\left[\begin{pmatrix} \xi_{0k} \\ \boldsymbol{\delta}_{0k} \end{pmatrix}^{(m+1)} \quad \dots \quad \begin{pmatrix} \xi_{Ck} \\ \boldsymbol{\delta}_{Ck} \end{pmatrix}^{(m+1)}\right] \mid \cdot \sim \mathcal{N}\left(\boldsymbol{\mu}^{*}, \Omega^{*}\right),$$

where

$$\boldsymbol{\Omega}^{*-1} = \boldsymbol{\Sigma}^{-1}_{\tilde{\xi}_k} \otimes \left(\begin{bmatrix} \mathbf{1} & \mathbf{v} \end{bmatrix}^t \begin{bmatrix} \mathbf{1} & \mathbf{v} \end{bmatrix}\right)^{-1} + \boldsymbol{\Sigma}^{-1}_{\xi\delta},$$

$$\boldsymbol{\mu}^{*} = \boldsymbol{\Omega}^{*}\left(\boldsymbol{\Sigma}^{-1}_{\tilde{\xi}_k} \otimes \left(\begin{bmatrix} \mathbf{1} & \mathbf{v} \end{bmatrix}^t \begin{bmatrix} \mathbf{1} & \mathbf{v} \end{bmatrix}\right)^{-1} \begin{bmatrix} \hat{\xi_{0k}} \\ \hat{\boldsymbol{\delta}_{0k}} \\ \dots \\ \hat{\xi_{Ck}} \\ \hat{\boldsymbol{\delta}_{Ck}} \end{bmatrix} + \boldsymbol{\Sigma}^{-1}_{\xi\delta} \begin{bmatrix} \mu_{\xi_0} \\ \boldsymbol{\mu}_{\delta_0} \\ \dots \\ \mu_{\xi_C} \\ \boldsymbol{\mu}_{\delta_C} \end{bmatrix}\right),$$

$$\text{with} \begin{bmatrix} \hat{\boldsymbol{\xi}_k} \\ \hat{\boldsymbol{\delta}_k} \end{bmatrix} = \left(\begin{bmatrix} \mathbf{1} & \mathbf{v} \end{bmatrix}^t \begin{bmatrix} \mathbf{1} & \mathbf{v} \end{bmatrix}\right)^{-1} \begin{bmatrix} \mathbf{1} & \mathbf{v} \end{bmatrix}^t \tilde{\boldsymbol{\xi}}_k,$$

$$\text{and } \boldsymbol{\Sigma}_{\xi\delta} = \left(\sigma_{\xi_0} \oplus \boldsymbol{\Sigma}_{\delta_0}\right) \oplus \dots \oplus \left(\sigma_{\xi_C} \oplus \boldsymbol{\Sigma}_{\delta_C}\right).$$

5. For each k, sample $\boldsymbol{\Sigma}^{(m+1)}_{\tilde{\xi}_k}$ from the full conditional

$$\boldsymbol{\Sigma}^{(m+1)}_{\tilde{\xi}_k} \mid \tilde{\boldsymbol{\xi}}^{(m+1)}, \boldsymbol{\xi}^{(m+1)}_k, \boldsymbol{\delta}^{(m+1)}_k, S_{0\xi_k}, n_0 \sim \mathcal{IW}\left((n_0 + J/2), (S_{0\xi_k} + S)\right),$$

$$\text{where } S = \left(\tilde{\xi}_k - (\xi_k + \mathbf{v}\boldsymbol{\delta}_k)\right)' \left(\tilde{\xi}_k - (\xi_k + \mathbf{v}\boldsymbol{\delta}_k)\right).$$

6. Sample $\boldsymbol{\mu}_{\xi}^{(m+1)}$ and $\boldsymbol{\Sigma}_{\xi}^{(m+1)}$ from the full conditionals

$$\begin{aligned}
\boldsymbol{\mu}_{\xi}^{(m+1)} \mid \boldsymbol{\xi}_k^{(m+1)}, \boldsymbol{\Sigma}_{\xi}^{(m)}, K_0 &\sim \mathcal{N}\left(\frac{K}{K_0+K}\overline{\boldsymbol{\xi}}, \frac{\boldsymbol{\Sigma}_{\xi}}{K_0+K}\right), \\
\boldsymbol{\Sigma}_{\xi}^{(m+1)} \mid \boldsymbol{\xi}_k^{(m+1)}, \nu, K_0, \mathbf{S}_{0_{\xi}}, &\sim \mathcal{IW}\left(K+\nu, \boldsymbol{\Sigma}^*\right),
\end{aligned}$$

with

$$\begin{aligned}
\overline{\boldsymbol{\xi}} &= \sum_k \boldsymbol{\xi}_k / K, \\
\boldsymbol{\Sigma}^* &= \mathbf{S}_{0_{\xi}} + K\left((\boldsymbol{\xi}_k - \overline{\boldsymbol{\xi}})(\boldsymbol{\xi}_k - \overline{\boldsymbol{\xi}})^t\right) + \frac{KK_0}{K+K_0}\overline{\boldsymbol{\xi}}\,\overline{\boldsymbol{\xi}}^t.
\end{aligned}$$

7. Sample $\boldsymbol{\mu}_{\delta}^{(m+1)}$ and $\boldsymbol{\Sigma}_{\delta}^{(m+1)}$ from the full conditionals

$$\begin{aligned}
\boldsymbol{\mu}_{\delta}^{(m+1)} \mid \boldsymbol{\delta}_k^{(m+1)}, \boldsymbol{\Sigma}_{\delta}^{(m)}, K_0 &\sim \mathcal{N}\left(\frac{K}{K_0+K}\overline{\boldsymbol{\delta}}, \frac{\boldsymbol{\Sigma}_{\delta}}{K_0+K}\right), \\
\boldsymbol{\Sigma}_{\delta}^{(m+1)} \mid \boldsymbol{\delta}_k^{(m)}, \nu, K_0, \mathbf{S}_{0\delta} &\sim \mathcal{IW}\left(K+\nu, \boldsymbol{\Sigma}^*\right),
\end{aligned}$$

with

$$\begin{aligned}
\overline{\boldsymbol{\delta}} &= \sum_k \boldsymbol{\delta}_k / K, \\
\boldsymbol{\Sigma}^* &= \mathbf{S}_{0\delta} + K\left((\boldsymbol{\delta}_k - \bar{\boldsymbol{\delta}})(\boldsymbol{\delta}_k - \bar{\boldsymbol{\delta}})^t\right) + \frac{KK_0}{K+K_0}\bar{\boldsymbol{\delta}}\bar{\boldsymbol{\delta}}^t.
\end{aligned}$$

8. For each i, sample $\boldsymbol{\beta}_i^{(m+1)}$ from the full conditional

$$\boldsymbol{\beta}_i^{(m+1)} \mid \boldsymbol{\theta}_i^{(m+1)}, \sigma_j^{(m)}, \mathbf{T}^{(m)}, \boldsymbol{\gamma}^{(m)}, \boldsymbol{\zeta}^{(m)} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\beta}, \boldsymbol{\Sigma}_{\beta}\right),$$

where

$$\begin{aligned}
\boldsymbol{\mu}_{\beta} &= \boldsymbol{\Sigma}_{\beta}\left(\sigma_j^{-2}\left(\mathbf{x}_i^t\mathbf{x}_i\right)\hat{\beta}_i + \mathbf{T}^{-1}\mathbf{w}_i^t\boldsymbol{\gamma}\right), \\
\boldsymbol{\Sigma}_{\beta} &= \left(\sigma_j^{-2}\left(\mathbf{x}_i^t\mathbf{x}_i\right) + \mathbf{T}^{-1}\right)^{-1}, \\
\text{with } \hat{\beta}_i &= \left(\mathbf{x}_i^t\mathbf{x}_i\right)^{-1}\left(\mathbf{x}_i^t(\boldsymbol{\theta}_i - \mathbf{s}_i\boldsymbol{\zeta})\right).
\end{aligned}$$

9. Sample $\boldsymbol{\gamma}^{(m+1)}$ from the full conditional

$$\boldsymbol{\gamma}^{(m+1)} \mid \boldsymbol{\beta}_i^{(m+1)}, \mathbf{T}^{(m)}, S_{0\gamma} \sim \mathcal{N}(\boldsymbol{\mu}_{\gamma}, \boldsymbol{\Sigma}_{\gamma}),$$

where

$$\begin{aligned}
\boldsymbol{\mu}_{\gamma} &= \boldsymbol{\Sigma}_{\gamma}\sum_i \mathbf{w}_i^t\mathbf{T}^{-1}\boldsymbol{\beta}_i, \\
\boldsymbol{\Sigma}_{\gamma} &= \left(\sum_i \mathbf{w}_i^t\mathbf{T}^{-1}\mathbf{w}_i + S_{0\gamma}^{-1}\right)^{-1}.
\end{aligned}$$

10. Sample $\mathbf{T}^{(m+1)}$ from the full conditional

$$\mathbf{T}^{(m+1)} \mid \boldsymbol{\beta}_i^{(m+1)}, \boldsymbol{\gamma}^{(m+1)}, S_{0T} \sim \mathcal{IW}\left((n_{0T}+I), \left(S_{0T} + \sum_i(\boldsymbol{\beta}_i - \mathbf{w}_i\boldsymbol{\gamma})(\boldsymbol{\beta}_i - \mathbf{w}_i\boldsymbol{\gamma})^t\right)\right).$$

11. Sample $\boldsymbol{\zeta}^{(m+1)}$ and $\boldsymbol{\Sigma}_{\zeta}^{(m+1)}$ from the full conditionals

$$\boldsymbol{\zeta}^{(m+1)} \mid \boldsymbol{\theta}^{(m+1)}, \boldsymbol{\beta}^{(m+1)}, \Sigma_{\zeta}^{(m)} \sim \mathcal{N}\left(\frac{K}{K_0+N}\hat{\boldsymbol{\xi}}, \frac{\boldsymbol{\Sigma}_{\xi}}{K_0+N}\right),$$

$$\boldsymbol{\Sigma}_{\zeta}^{(m+1)} \mid \boldsymbol{\zeta}^{(m)}, \nu, K_0, \mathbf{S}_{0\zeta} \sim \mathcal{IW}\left(K+\nu, \boldsymbol{\Sigma}^{*}\right),$$

$$\text{with } \hat{\boldsymbol{\xi}} = (\mathbf{s}^t\mathbf{s})^{-1}\left(\mathbf{s}^t(\boldsymbol{\theta}-\mathbf{x}\boldsymbol{\beta})\right),$$

$$\text{and } \boldsymbol{\Sigma}^{*} = \mathbf{S}_{0\zeta} + N\left((\boldsymbol{\theta}-\mathbf{x}\boldsymbol{\beta})-\mathbf{s}\boldsymbol{\zeta}\right)^t\left((\boldsymbol{\theta}-X\boldsymbol{\beta})-\mathbf{s}\boldsymbol{\zeta}\right) + \frac{NK_0}{N+K_0}\left(\hat{\boldsymbol{\xi}}\right)^t\left(\hat{\boldsymbol{\xi}}\right).$$

12. Sample $\sigma_j^{(m+1)}$ from the full conditional

$$\sigma_j^{(m+1)} \mid \boldsymbol{\theta}_j^{(m+1)}, \boldsymbol{\beta}^{(m+1)}, \boldsymbol{\zeta}^{(m+1)}, n_0, s_{0\sigma} \sim \mathcal{IG}\left(n_j+n_0, s^{*}+s_{0\sigma}\right),$$

$$\text{where } s^{*} = \sum_{1}^{n_j}\left((\theta_{ij}-\mathbf{x}_i\boldsymbol{\beta}_i)-\mathbf{s}_i\boldsymbol{\zeta}\right)^2.$$

Acknowledgements

The authors thank Martine Fledderus and Ernst Bohlmeijer for providing the data from the randomized trial on the treatment of depression. Funding for the project was provided by the Dutch scientific organization (NWO). *Conflict of interest*: None declared.

References

1. Spearman C. General intelligence, objectively determined and measured. *American Journal of Psychology* 1904; **15**:201–292. DOI: 10.2307/1412107.
2. Halberstadt SM, Schmitz KH, Sammel MD. A joint latent variable model approach to item reduction and validation. *Biostatistics* 2012; **13**:48–60. DOI: 10.1093/biostatistics/kxr018.
3. Lord FM. *Applications of Item Response Theory to Practical Testing Problems*. Lawrence Erlbaum: New Jersey, 1980.
4. Lazarsfeld PF. *Measurement and Prediction*. Princeton University Press: Princeton, NJ, 1950.
5. McHugh TB. Efficient estimation and local identification in latent class analysis. *Psychometrika* 1956; **21**:331–347. DOI: 10.1007/BF02296300.
6. Finkelman M, Green JG, Gruber MJ, Zaslavsky AM. A Zero- and K-Inflated mixture model for health questionnaire data. *Statistics in Medicine* 2001; **30**:1028–1043. DOI: 10.1002/sim.4217.
7. Li Y, Baser R. Using R and Winbugs to fit a generalized partial credit model for developing and evaluating patient-reported outcomes assessments. *Statistics in Medicine* 2012; **31**(18):2010–2026.
8. Klein Entink RH, Fox J-P, van den Hout A. A mixture model for the joint analysis of latent developmental trajectories and survival. *Statistics in Medicine* 2011; **30**:2310–2325. DOI: 10.1002/sim.4266.
9. Bryk AS, Raudenbush SW. Application of hierarchical linear models to assessing change. *Psychological Bulletin* 1987; **101**:147–158. DOI: 10.1037//0033-2909.101.1.147.
10. Goldstein H. Models for multilevel response variables with an application to growth curves. In *Multilevel Analysis of Educational Data*, Bock RD (ed.). Academic Press: San Diego, CA, 1989.
11. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; **38**:963–974. DOI: 10.2307/2529876.
12. Hertzog C, Nesselroade JR. Assessing psychological change in adulthood: an overview of methodological issues. *Psychology and Aging* 2003; **18**:639–657. DOI: 10.1037/0882-7974.18.4.639.
13. McArdle JJ. Latent variable growth within behavior genetic models. *Behavioral Genetics* 1986; **16**:163–200. DOI: 10.1007/BF01065485.
14. Meredith W, Tisak J. Latent curve analysis. *Psychometrika* 1990; **55**:107–122. DOI: 10.1007/BF02294746.
15. Muthén LK, Muthén BO. *Mplus*. Muthén and Muthén: Los Angeles, CA, 2001.
16. Blanchin M, Hardouin J-B, Le Neel T, Kubis G, Blanchard C, Miraillé E, Sébille V. Comparison of CTT and IRT based-approach for the analysis of longitudinal patient reported outcome. *Statistics in Medicine* 2011; **30**:825–838. DOI: 10.1002/sim.4153.
17. Liu LC. A model for incomplete multivariate ordinal data. *Statistics in Medicine* 2008; **27**:6299–6309. DOI: 10.1002/sim.3422.
18. Muthén B, Brown CH, Booiljo KM, Khoo S-T, Yang C-C, Wang C-P, Kellam SG, Carlin JB, Liao J. General growth mixture modeling for randomized preventative interventions. *Biostatistics* 2002; **3**:459–475. DOI: 10.1093/biostatistics/3.4.459.
19. Meredith W. Measurement invariance, factor analysis and factorial invariance. *Psychometrika* 1993; **58**:525–543. DOI: 10.1007/BF02294825.
20. Meredith W, Horn J. The role of factorial invariance in modeling growth and change. In *New Methods for the Analysis of Change: Decade of Behavior*, Collins LM, Sayer AG (eds). American Psychological Association: Washington, DC, 2001, DOI: 10.1037/10409-007.

21. Meade AW, Lautenschlager GJ, Hecht JE. Establishing measurement equivalence/invariance in longitudinal data with item response theory. *International Journal of Testing* 2005; **5**:279–301. DOI: 10.1207/s15327574ijt0503-6.
22. Millsap RE. Testing measurement invariance using item response theory in longitudinal data: an introduction. *Child Development Perspectives* 2010; **4**:5–9. DOI: 10.1111/j.1750-8606.2009.00109.x.
23. Horn JL, McArdle JJ. A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research* 1992; **18**:117–144. DOI: 10.1080/03610739208253916.
24. Crane PK, Gibbons LE, Jolley LMS, van Belle G. Differential item functioning analysis with ordinal logistic regression techniques: DIFdetect and difwithpar. *Medical Care* 2006; **44**:115–123. DOI: 10.1097/01.mlr.0000245183.28384.ed.
25. Crane PK, van Belle G, Larson EB. Test bias in a cognitive test: differential item functioning in the CASI. *Statistics in Medicine* 2004; **23**:241–256. DOI: 10.1002/sim.1713.
26. Fox J-P. *Bayesian Item Response Modeling: Theory and Applications*. Springer: New York, 2010. DOI: 10.1007/978-1-4419-0742-4.
27. De Jong MG, Steenkamp JBEM, Fox J-P. Relaxing measurement invariance in cross-national consumer research using a hierarchical IRT model. *Journal of Consumer Research* 2007; **34**:260–278. DOI: 10.1086/518532.
28. Fox J-P, Verhagen AJ. Random item effects modeling for cross-national survey data. In *Cross-cultural Analysis: Methods and Applications*, Davidov E, Schmidt P, Billiet J (eds). Routeledge Academic: London, 2010.
29. Howard GS, Dailey PR. Response-shift bias: a source of contamination of self-report measures. *Journal of Applied Psychology* 1979; **64**:144–150. DOI: 10.1037/0021-9010.64.2.144.
30. De Boeck P. Random item IRT models. *Psychometrika* 2008; **73**:533–559. DOI: 10.1007/s11336-008-9092-x.
31. Cho S-J, Rabe-Hesketh S. Alternating imputation posterior estimation of models with crossed random effects. *Computational Statistics and Data Analysis* 2011; **55**:12–25. DOI: 10.1016/j.csda.2010.04.015.
32. Masters GN. A Rasch model for partial credit scoring. *Psychometrika* 1982; **47**:149–174. DOI: 10.1007/BF02296272.
33. Muraki E. A generalized partial credit model: application of an EM algorithm. *Applied Psychological Measurement* 1992; **16**:159–176. DOI: 10.1177/014662169201600206.
34. Agresti A. *Categorical Data Analysis*. John Wiley and Sons: Hoboken, New Jersey, 2002. DOI: 10.1002/0471249688.
35. Verhagen AJ, Fox J-P. Bayesian tests of measurement invariance. *British Journal of Mathematical and Statistical Psychology* 2012. DOI: 10.1111/j.2044-8317.2012.02059.x.
36. Dickey JM. The weighted likelihood ratio, linear hypotheses on normal location parameters. *The Annals of Mathematical Statistics* 1971; **42**:204–223. DOI: 10.1214/aoms/1177693507.
37. Verdinelli I, Wasserman L. Computing Bayes factors using a generalization of the Savage–Dickey density ratio. *Journal of the American Statistical Association* 1995; **90**:614–618. DOI: 10.1080/01621459.1995.10476554.
38. Geweke J. *Contemporary Bayesian Econometrics and Statistics*. Wiley-Interscience: New York, 2005. DOI: 10.1002/0471744735.
39. Hoijtink HJA. *Informative Hypotheses. Theory and Practice for Behavioral and Social Scientists*. Chapman & Hall/CRC: Boca Raton, 2011. DOI: 10.1201/b11158.
40. Klugkist I. Competing theories based on (in) equality constraints. In *Bayesian Evaluation of Informative Hypotheses*, Hoijtink H, Klugkist I, Boelen PA (eds). Springer Verlag: New York, 2008, DOI: 10.1007/978-0-387-09612-4.
41. Klugkist I, Hoijtink H. The Bayes factor for inequality and about equality constrained models. *Computational Statistics & Data Analysis* 2007; **51**:6367–6379. DOI: 10.1016/j.csda.2007.01.024.
42. Spiegelhalter DJ, Best NG, Carlin BR, Van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 2002; **64**:583–616. DOI: 10.1111/1467-9868.00353.
43. Hayes SC, Luoma JB, Bond FW, Masuda A, Lillis J. Acceptance commitment therapy: model, process and outcomes. *Behaviour Research and Therapy* 2006; **44**:1–25. DOI: 10.1016/j.brat.2005.06.006.
44. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement* 1977; **1**:385–401. DOI: 10.1177/014662167700100306.
45. Bohlmeijer ET, Fledderus M, Rokx TAJJ, Pieterse ME. Efficacy of an early intervention based on acceptance and commitment therapy for adults with depressive symptomatology: evaluation in a randomized controlled trial. *Behavior Research and Therapy* 2011; **49**:62–67. DOI: 10.1016/j.brat.2010.10.003.
46. Forman EM, Herbert JD, Moitra E, Yeomans PD, Geller PA. A randomized controlled effectiveness trial of acceptance and commitment therapy and cognitive therapy for anxiety and depression. *Behavior Modification* 2007; **31**:772–799. DOI: 10.1177/0145445507302202.
47. Bond FW, Hayes SC, Baer RA, Carpenter KC, Guenole N, Orcutt HK, Waltz T, Zettle RD. Preliminary psychometric properties of the acceptance and action questionnaire-II: a revised measure of psychological flexibility and acceptance. *Behavior Therapy* 2011; **42**:676–688. DOI: 10.1016/j.beth.2011.03.007.
48. Sprangers MAG, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. *Social Science & Medicine* 1999; **48**:1507–1515. DOI: 10.1016/S0277-9536(99)00045-3.