British Journal of Mathematical and Statistical Psychology (2014)

The British Psychological Society

www.wileyonlinelibrary.com

A generalized linear factor model approach to the hierarchical framework for responses and response times

Dylan Molenaar[1]*, Francis Tuerlinckx[2] and Han L. J. van der Maas[1]

[1]University of Amsterdam, The Netherlands

[2]University of Leuven, Belgium

We show how the hierarchical model for responses and response times as developed by van der Linden (2007), Fox, Klein Entink, and van der Linden (2007), Klein Entink, Fox, and van der Linden (2009), and Glas and van der Linden (2010) can be simplified to a generalized linear factor model with only the mild restriction that there is no hierarchical model at the item side. This result is valuable as it enables all well-developed modelling tools and extensions that come with these methods. We show that the restriction we impose on the hierarchical model does not influence parameter recovery under realistic circumstances. In addition, we present two illustrative real data analyses to demonstrate the practical benefits of our approach.

1. Introduction

Psychometrics is concerned with the specification of appropriate measurement models to link observed item scores to the underlying latent variable that the test purports to measure. In the case of a continuously distributed latent variable, various models have been proposed for dichotomously scored items, such as the Rasch model (Rasch, 1960) and the two- and three-parameter logistic models (Birnbaum, 1968); for ordinal scored items, such as the graded response model (Samejima, 1969), the ordinal factor model (Wirth & Edwards, 2007), and the partial credit model (Masters, 1982); and for (approximate) continuously scored items, such as the linear factor model (Mellenbergh, 1994b; Spearman, 1904; Thurstone, 1947). These models have been developed to analyse item responses gathered using traditional paper-and-pencil tests.

Nowadays, item responses are increasingly collected using computerized tests. In this setting, together with the item scores, response times on the test items have become available. Within the field of psychometrics, the question arises how to incorporate this additional information into the measurement model. Different approaches are possible (for an overview, see Tuerlinckx, Molenaar, & van der Maas, 2012). One possible approach is based on process models for decision-making from cognitive psychology (Tuerlinckx & De Boeck, 2005; Tuerlinckx *et al.*, 2012; van der Maas, Molenaar, Maris, Kievit, & Borsboom, 2011). In these models, the observed item scores and reaction times are assumed to have arisen from an underlying diffusion process (Ratcliff, 1978).

*Correspondence should be addressed to Dylan Molenaar, Psychological Methods, Department of Psychology, University of Amsterdam, Weesperplein 4, 1018 XA Amsterdam, The Netherlands (email: d.molenaar@uva.nl).

DOI:10.1111/bmsp.12042

Tuerlinckx and De Boeck (2005) showed that this postulation implies a two-parameter item response model for the item responses and a response time distribution according to a Wiener process (Cox & Miller, 1970).

Another approach is to use the response times as collateral information. For instance, the responses and response times have been modelled within a hierarchical model (van der Linden, 2007; Fox *et al.*, 2007; Klein Entink, Fox, and van der Linden, 2009; Glas and van der Linden, 2010). This model is hierarchical in the sense that it consists of two levels. At the first level, the observed responses are linked to the latent ability variable using a two- or three-parameter item response model (Birnbaum, 1968; Lord, 1952). The response times are linked to the latent speed variable through a lognormal model (Samejima, 1973). At the second level of the model, both measurement models are connected by means of linear correlations between their item and person parameters. The model is thus a hierarchical crossed random effects model, as the item and the person parameters are considered to be multivariate random variables.

Within the field of psychometrics, the hierarchical crossed random effects model has been the most dominant approach to model responses and response times. As discussed in van der Linden (2007), the model can be fitted to data using a Bayesian procedure. Most developments and applications of the hierarchical framework have focused on this approach (but see Glas & van der Linden, 2010, for a frequentist version of the hierarchical crossed random effects model). For instance, the Bayesian model has been used to test cognitive theory (Klein Entink, Kuhn, Hornke, & Fox, 2009), to detect collusion between test takers (Van der Linden, 2009), to detect aberrant response time patterns (Van der Linden & Guo, 2008), to enable multivariate structural modelling of the ability and speed parameters (Klein Entink, Fox *et al.*, 2009), to improve item selection in computerized adaptive testing (Van der Linden, 2008) and to account for differential speediness in multistage testing (Van der Linden, Breithaupt, Chuah, & Zhang, 2007).

The advantages of adopting a Bayesian framework over a frequentist framework are the following (see Wagenmakers, Lee, Lodewyckx, & Iverson, 2008): first, it does not depend on asymptotic theory; second, item parameters can be considered random effects;[1] third, non-nested models can be meaningfully compared using Bayes factors;[2] fourth, it naturally incorporates the principle of parsimony ('Occam's razor') by marginalization of the likelihood function; and finally, it is more robust in handling complex models (e.g., models with more than five random effects; Wood, Wilson, Gibbons, Schilling, Muraki, & Bock, 2002).

However, despite these advantages, the relatively complex mathematical underpinning of the Bayesian sampling methodology and the limited availability of statistical software to fit the model hampers application of the hierarchical crossed random effects model for responses and response times. In addition, comparison of the model to competing models that are not implemented in a Bayesian way (e.g., the diffusion model; Tuerlinckx & De Boeck, 2005; Tuerlinckx *et al.*, 2012) is

[1] Note that this is – in principle – also possible in a frequentist framework by specifying random item effects; see De Boeck (2008) and Glas and van der Linden (2010). However, this is not common practice within a frequentist framework.

[2] In a frequentist framework, non-nested models are commonly compared using fit indices such as AIC and BIC. However, these indices are in essence Bayesian as their computation implicitly involves a prior; see Burnham and Anderson (2002).

difficult as the different frameworks use different fit measures which hampers comparability.

To advance the application of the hierarchical framework for responses and response times among applied psychologists, we show in this paper that by restricting the hierarchical crossed random effects model to include only random person effects, the model can readily be fitted within a (frequentist) generalized linear factor modelling framework (Mellenbergh, 1994a; Skrondal & Rabe-Hesketh, 2004). By doing this, one can profit from all well-developed and flexible modelling tools and extensions that exists within this framework. To name a few possibilities: (i) Standard latent variable modelling software can be used to fit the hierarchical model, for example, Mplus (Muthén & Muthén, 2007), Lisrel (Jöreskog & Sörbom, 1993), Amos (Arbuckle, 1997), Mx (Neale, Boker, Xie, & Maes, 2006), SAS (SAS Institute Inc, 2011), and OpenMX (Boker *et al.*, 2010). (ii) One is not limited to the two- or three-parameter logistic model for the item scores. The measurement model for the response data can be any model of choice (as long as it is implemented in the software that is used), for example a graded response model or a latent class model. (iii) The model could be extended to include multiple latent variables to account for multidimensionality in the case of, for instance, an intelligence test battery, or a test battery with item bundles that share a common property (testlet effects). (iv) It is straightforward to incorporate multi-level and multi-group components (as in Klein Entink, Fox, *et al.*, 2009). (v) It enables structural modelling of the speed and or ability factor in a traditional factor analysis manner. (vi) Time limits can be modelled using, for instance, truncation or censoring of the response time part of the model (Dolan, van der Maas, & Molenaar, 2002). (vii) Various types of well-developed model selection tools become available, such as likelihood-ratio tests, power analysis, modification indices, model fit statistics (as also argued by Glas & van der Linden, 2010), and bootstrap procedures. (viii) Differential item functioning can be investigated easily on both the responses and the response times.

Some of the possibilities described above are also possible within the Bayesian or frequentist hierarchical crossed random effects model in principle. However, necessary model developments either are not implemented in existing software – e.g., in *cirt* (Fox *et al.*, 2007), a package for the statistical software program R (R Development Core Team, 2007) – or require advanced statistical and/or programming knowledge – e.g., to implement it in WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000) or R. Therefore, insight into how the hierarchical model for responses and response times can be studied from the well-known and well-developed generalized linear factor model framework seems valuable.

The correspondence between the hierarchical modelling approach of van der Linden (2007) and the present generalized linear modelling approach was previously noted by Ranger and Ortner (2012). The present study adds to this study in three ways. First, we give the relation between both approaches formally. Results show, *inter alia*, that the hierarchical model includes random item effects while the generalized linear model does not. This issue was beyond the scope of Ranger and Ortner (2012). Second, we study the effect of the omission of the random item effects from the hierarchical crossed random effects model in a simulation study. Third, we give two illustrative examples of the benefits of the generalized linear modelling approach.

The outline of this paper is as follows. First, we present the Bayesian hierarchical crossed random effects model and discuss how it simplifies into a frequentist generalized linear factor model by excluding the random item effects. Next, we show by means of a simulation study that omitting the random item effects does not result in parameter bias under realistic circumstances, even though the true item effects are random. We then present two real data applications to illustrate the benefits of the present approach. We conclude with a general discussion.

2. The hierarchical crossed random effects model

As discussed above, the hierarchical crossed random effects model consists of two levels. At the first level, the observed data are linked to the latent speed and ability variables using the appropriate measurement models. For the responses, this is generally a two- or three-parameter item response model. Here we focus on the two-parameter normal ogive model (Lord, 1952),

$$\mathrm{P}(X_{pi} = 1|\theta_p) = \Phi[\alpha_i(\theta_p - \beta_i)], \tag{1}$$

where X_{pi} is the dichotomously coded response of person p for item i, $\Phi(\cdot)$ is the standard normal cumulative distribution function, θ_p denotes the latent ability variable, and α_i and β_i are the discrimination and difficulty parameters, respectively.

For the response times, the observed data, T_{pi}, are linked to the underlying latent speed variable, τ_p, through a lognormal model, using time intensity parameters, λ_i, and time discrimination parameters, ϕ_i. Here we focus on the specification used by Fox *et al.* (2007), Klein Entink, Fox, *et al.* (2009), and Fox (2010, Chapter 8) in which the time discrimination parameter is modelled as a slope parameter for the latent speed variable. Van der Linden (2007) and Glas and van der Linden (2010) use a slightly different specification in which the time discrimination parameter, ϕ_i, is the precision of the lognormal distribution of the response times. We focus on the former specification as this is the version of the model that is implemented in *cirt* (Fox *et al.*, 2007). In this specification, the lognormal model is given by

$$T_{pi} = \exp(\omega_{pi} + \lambda_i - \phi_i \tau_p), \tag{2}$$

where ω_{pi} is a normally distributed variable with variance $\sigma^2_{\omega_i}$. Within equation (2), the van der Linden (2007) and Glas and van der Linden (2010) notation can be obtained by fixing ϕ_i to 1 for all i. Then, $\sigma^2_{\omega_i}$ is interpreted as the time discrimination parameter.

At the second level, item and person parameters are allowed to covary. Denoting log(x) as x^*, the item parameter vector, $[\alpha^*_i \quad \beta \quad \phi^*_i \quad \lambda_i]$, is assumed to have a multivariate normal distribution with mean vector and covariance matrix

$$\boldsymbol{\mu}_I = [\mu_{\alpha^*} \quad \mu_\beta \quad \mu_{\varphi^*} \quad \mu_\lambda] \quad \text{and} \quad \boldsymbol{\Sigma}_I = \begin{bmatrix} \sigma^2_{\alpha^*} & & & \\ \sigma_{\beta\alpha^*} & \sigma^2_\beta & & \\ \sigma_{\varphi^*\alpha^*} & \sigma_{\varphi^*\beta} & \sigma^2_{\varphi^*} & \\ \sigma_{\lambda\alpha^*} & \sigma_{\lambda\beta} & \sigma_{\lambda\varphi^*} & \sigma^2_\lambda \end{bmatrix}, \tag{3}$$

where α^* and ϕ^* are log(α) and log(ϕ), respectively. In addition, the person parameter vector, $[\theta_p \quad \tau_p]$, is assumed to have a multivariate normal distribution with mean vector and covariance matrix

$$\boldsymbol{\mu}_P = [\,\mu_\theta \quad \mu_\tau\,] \quad \text{and} \quad \boldsymbol{\Sigma}_P = \begin{bmatrix} \sigma_\theta^2 & \\ \sigma_{\theta\tau} & \sigma_\tau^2 \end{bmatrix}, \tag{4}$$

where

$$\rho_{\theta\tau} = \frac{\sigma_{\theta\tau}}{\sigma_\theta \sigma_\tau}$$

gives the correlation between θ_p and τ_p. To enable Bayesian parameter estimation of the model in equations (1)–(4), priors need to be specified on the parameters of the model; see van der Linden (2007) and Fox *et al.* (2007). However, as we adopt a frequentist framework as in Glas and van der Linden (2010), we do not need to specify these. The model in equations (1)–(4) will be called the hierarchical model in the remainder of the paper.

3. A generalized linear factor modelling approach to the hierarchical model

With some minor simplification at the second level, the hierarchical model given by equations (1)–(4) can be specified in a generalized linear factor model framework (Mellenbergh, 1994a; Skrondal & Rabe-Hesketh, 2004).

To start, at the first level, the probability model in equation (1) is written by assuming that the discrete data in X arose because of the categorization of an underlying standard normally distributed variable Z_{pi} at specific thresholds, β'_i (see Takane & de Leeuw, 1987; Wirth & Edwards, 2007).[3] If we submit Z_{pi} to a linear factor model, that is,

$$Z_{pi} = \alpha'_i \theta_p = \varepsilon_{pi},$$

we obtain

$$\mathrm{P}(X_{pi} = 1 | \theta_p) = \Phi\left(\frac{\alpha'_1 \theta_p - \beta'_i}{\sigma_{\varepsilon_i}}\right), \tag{5}$$

which can be written as equation (1) by substituting $\alpha'_i = \alpha_i \times \sigma_{\varepsilon_i}$ and $\beta'_i = \beta_i \times \sigma_{\varepsilon_i}$. Throughout this paper we fix σ_{ε_i} to 1 for identification purposes; however, in case of polytomous data, it is possible to estimate σ_{ε_i} by fixing two adjacent thresholds (Mehta, Neale, & Flay, 2004; see also Molenaar, Dolan, & de Boeck, 2012).

Next, we note that equation (2) can be written as a linear factor model on $\log(T_{pi})$, that is,

$$\log(T_{pi}) = \lambda'_i + \varphi'_i \tau_p + \omega_{pi}. \tag{6}$$

When substituting $\lambda'_i = \lambda_i$ and $\phi' = -\phi_i$, the lognormal model in equation (2) arises.

[3] For the fixed effects parameters in the generalized linear factor model, we use the same symbols for those parameters that correspond to parameters from the hierarchical crossed random effects model. However, we put a prime (′) on them to distinguish them. Some parameters need to be transformed first to ensure a one-to-one correspondence with the parameters from the hierarchical model, for instance, for the discrimination parameter $\alpha = \alpha'/\sigma_\varepsilon$.

Together, equations (5) and (6) form the first level of the model, which is equivalent to a two-factor model with discrete indicators for the first factor, and continuous indicators for the second factor. As the transformed responses, X^*, and the transformed response times, $\log(T_{pi})$, follow a linear form, the model belongs to the family of generalized linear factor models. Within this generalized linear framework it is straightforward to introduce the second-level person parameter covariances from equation (4). θ_p and τ_p can be allowed to correlated with correlation parameter $\rho_{\theta\tau}\prime$, that is,

$$\boldsymbol{\mu}_P = [\,\mu_\theta \quad \mu_\tau\,] \quad \text{and} \quad \boldsymbol{\Sigma}_P \begin{bmatrix} \sigma_\theta^2 & \\ \sigma'_{\theta\tau} & \sigma_\tau^2 \end{bmatrix} \tag{7}$$

with

$$\rho'_{\theta\tau} = \frac{\sigma'_{\theta\tau}}{\sigma_\theta \sigma_\tau}.$$

Note that because of the inverse relation between φ'_i and ϕ_i in equation (6), we have that $\sigma'_{\theta\tau} = -\sigma_{\theta\tau}$ and thus that $\rho'_{\theta\tau} = -\rho_{\theta\tau}$.

Traditionally, in the (generalized) linear factor analysis framework, the item parameters are assumed to be fixed and not random. Therefore, we do not include a multivariate distribution for the item parameters at the second level (with item parameter variances and covariances), for the following reasons: first, to remain in the framework of traditional factor analysis; second, because, as we will show in this paper, neglecting item parameter correlations will not cause bias in the results of the analysis; third, to enable the use of a large variety of existing, well-developed, and flexible modelling tools (e.g., modification indices); and finally, to arrive at a relatively simple and well-understood model. Note, however, that introducing random item effects within a frequentist framework is possible in principle; see De Boeck (2008) and Glas and van der Linden (2010). In addition, Mplus 7 enables the possibility of factor analysis with random item effects (see Asparouhov & Muthén, 2012). However, such models are computationally demanding, and relatively less flexible in the sense that not all modelling tools from the generalized linear factor framework are developed in a random item framework.

3.1. Identification

The hierarchical crossed random effects model in equations (1)–(4) is commonly identified by fixing $\mu_\theta = 0$, $\sigma_\theta^2 = 1$ and $\Pi_{i=1}^k \phi_i = 1$ (see Fox *et al.*, 2007). Identification of the generalized linear factor model from equations (5)–(7) can be done by either $\mu_\theta = 0$, $\sigma_\theta^2 = 1$, $\mu_\tau = 0$, and $\sigma_\tau^2 = 1$; or $\lambda'_i = 1$ for an arbitrary item, and $\alpha'_i = 1$ for an arbitrary item (see Bollen, 1989, p. 238). These different identification constraints cause σ_τ^2 to be a free parameter in the hierarchical model but not in the generalized factor model. Models are, however, still comparable using

$$\varphi'_i = -\varphi_i \times \sigma_\tau.$$

The generalized linear factor model from equations (5)–(7) can readily be fitted to data using the joint modelling facilities for continuous and discrete data of software packages OpenMx (Boker *et al.*, 2010) or Mplus (Muthén & Muthén, 2007), for instance. Example scripts can be found in Appendix A for OpenMx and in Appendix B for Mplus.

4. Simulation study

We show the viability of the present approach by comparing the Bayesian hierarchical crossed random effects model (BHM) for responses and response times to the frequentist generalized linear factor model (GFM) in a simulation study. Specifically, we want to demonstrate that parameter estimates are very similar between both approaches, and that omitting the random item effects does not result in bias or inefficiency (at least under realistic circumstances). To this end we simulated data according to the BHM and fitted the GFM with marginal maximum likelihood (Bock & Aitkin, 1981), using Mplus (Muthén & Muthén, 2007), and the BHM, using *cirt* (Fox *et al.*, 2007) to these data. We manipulated two aspects in the data: we used either 20 or 40 items because we think that this approximate number is representative of the number of items encountered for a subscale in practice; and we systematically varied the correlation between the item parameters (α_i, β_i, λ_i, and ϕ_i) from small to relatively large. Specifically, we chose the standardized off-diagonal elements of $\boldsymbol{\Sigma}_I$ to be either 0, .2, .4, or .6, that is, within each condition all item parameters have the same correlation.

Taken together, we have a design with $2 \times 4 = 8$ conditions. All other aspects of the data were kept constant across these conditions,: $\mu_\alpha = \mu_\beta = \mu_\phi = \mu_\lambda = 0$, $\sigma^2_\beta = \sigma^2_\lambda = 1$, $\sigma^2_\theta = \sigma^2_\tau = 1$, $\sigma^2_\omega = 1$, and $\rho = .5$. For α and ϕ we chose $\sigma^2_{\alpha^*} = \sigma^2_{\omega^*} = 0.3$, which gives var ($\alpha$) = var($\phi$) ≈ 0.472 after taking the exponent. Doing so results in values for α and ϕ between roughly 0 and 6, which we think are realistic. For each condition, we conducted 100 replications. For the GFM, the model was identified by fixing $\sigma^2_{\epsilon_i} = \sigma^2_\theta = \sigma^2_\tau = 1$ and $\mu_\theta = \mu_\tau = 0$. Next, estimates of α'_i and β'_i of the GFM needed to be transformed to enable comparison to the BHM estimates as *cirt* uses a normal-ogive item response model, while Mplus uses a logistic approximation. Transformation involved dividing the α'_i and β'_i estimates by 1.749 (Savalei, 2006). Note that when one uses, for instance, OpenMX (Boker *et al.*, 2010) to estimate the GFM, this is not necessary as OpenMx uses a normal-ogive model as well.

4.1. Results

To start, we note that the correlations between the BHM and GFM estimates were all close to 1 for all item parameters (i.e., correlations were roughly between .97 and .99 when aggregating over all conditions). This does not, however, rule out possible bias on the parameter estimates of the GFM. We therefore first considered the condition in the simulation study with the largest number of items and the largest item parameter correlation – the condition with 40 items and a correlation of .6 between the item parameters. For this condition, we plotted the parameter estimates against the true parameter values in Figure 1 (for the GFM) and in Figure 2 (for the BHM) across the 100 replications for the first item. As, for this condition, the data contain a large (.6) correlation between the item parameters, the possible effects of neglecting the random item effect in the GFM should be clear in these figures. First, by comparing Figures 1 and 2, it can be seen that both approaches recover the true values well, that is, the estimates show no systematic departures from the solid line (which represents a one-to-one correspondence). This suggests for the GFM that neglecting the random item effects does not result in bias.

With respect to the efficiency of the estimates, it can be seen that for the GFM in Figure 1, estimates are approximately as close to the solid line as for the BHM, indicating that neglecting the random item effects hardly affects the efficiency of the estimates. For

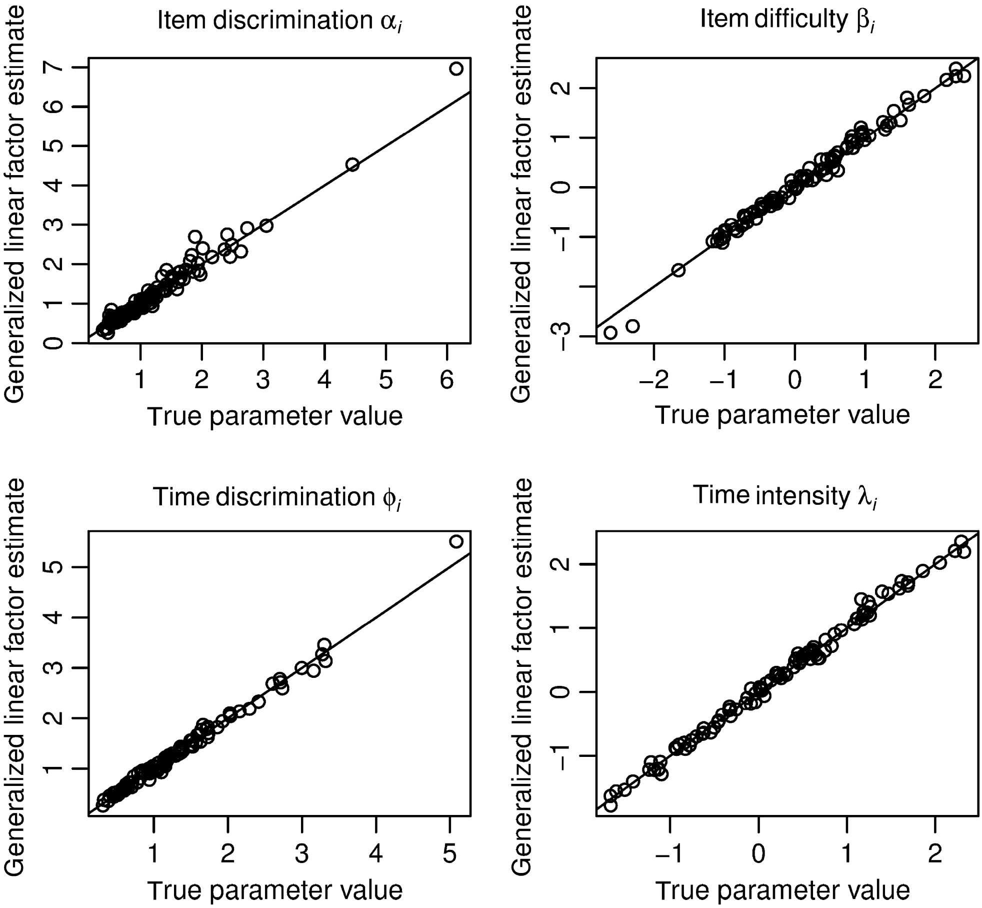

Figure 1. For the generalized linear factor model: estimated item parameters against their true values for the first item over the 100 replications of the most extreme condition in the simulation study (40 items; item parameter correlation. 6). The straight lines represent a one-to-one correspondence.

the time discrimination parameter, ϕ_i, the GFM estimates show even a somewhat better efficiency as compared to the BHM. These results can also be seen in more depth across all conditions in Tables 1 and 2. In Table 1, the mean true and estimated parameters are depicted for all conditions for the – arbitrarily chosen – first item (the results for the other items are similar). The difference between the mean true value and the average estimate gives an indication of the bias in the estimator. Generally, the estimates of the BHM and the GFM are close to the mean true value of the item parameters.[4] No systematic effects can be seen of the increasing item parameter correlations.

Table 2 shows the standard deviations of the true parameters and the estimated parameters in the BHM and GFM for the first item for all conditions. Ideally, the standard

[4] Note that we speak of 'mean true value' because within a given condition of the simulation study, the true value of the item parameter changes for each replication as the item parameters are random.

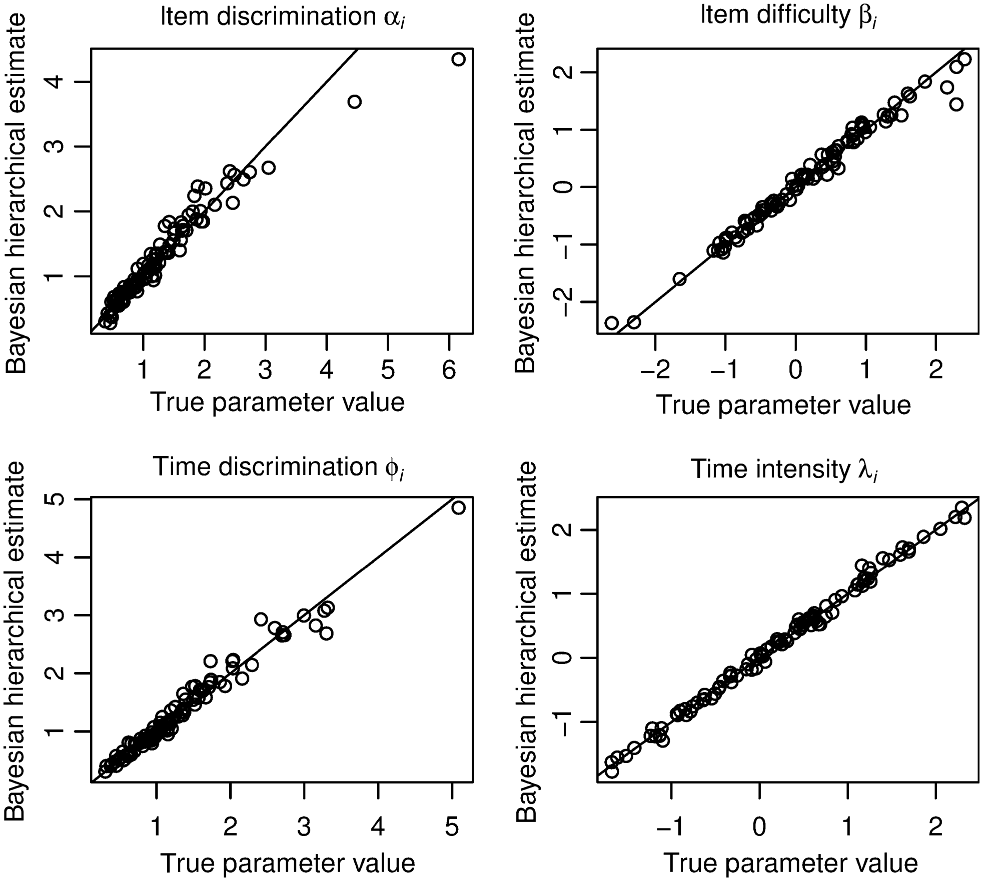

Figure 2. For the Bayesian hierarchical crossed random effects model: estimated item parameters against their true values for the first item over the 100 replications of the most extreme condition in the simulation study (40 items; item parameter correlation. 6). The straight lines represent a one-to-one correspondence.

deviations of the parameter estimates of the BHM and GFM are the same as the standard deviations of the true values indicating efficiency of the estimates. In the table it can be seen that the standard deviations are nearly the same for parameters λ_i and ϕ_i. For parameters α_i and β_i, standard deviations tend to be slightly larger for the GFM, but this is all within an acceptable range. In addition, no systematic deviations could be observed.

Not depicted in Tables 1 and 2 are the results concerning the residual variances, $\sigma^2_{\omega_i}$, the correlation between the person parameters, ρ, and the correlations among the item parameters. With respect to $\sigma^2_{\omega_i}$ and ρ, the results are similar to those above, that is, estimates are close to the true parameter values and show no large differences in efficiency. For the estimates of the correlations between the item parameters in the BHM, see Table 3. As can be seen, within the BHM, these parameters are underestimated, with more bias in the case of 20 items than for 40 items. This underestimation is to be expected as we used a limited number of items. In Bayesian statistics, it is known that with few data, the prior distribution has a more pronounced influence on the parameter estimates. In the present case, with a relatively small number of items and a prior that assigns most mass to a

Table 1. Means of the true and estimated parameters for the item parameters of item 1 in the simulation study

		α_i			β_i			ϕ_i			λ_i		
Items	Correlation	True	BHM	GLM	True	BHM	GLM	True	BHM	GLM	True	BHM	GLM
20	0	1.05	1.07	1.07	0.04	0.05	0.05	1.15	1.15	1.15	0.03	0.04	0.04
	.2	1.22	1.25	1.30	−0.01	−0.03	−0.04	1.21	1.20	1.22	0.00	0.00	0.00
	.4	1.05	1.09	1.10	−0.14	−0.13	−0.16	1.11	1.12	1.11	−0.16	−0.16	−0.16
	.6	1.42	1.39	1.48	0.09	0.08	0.10	1.33	1.32	1.34	0.19	0.19	0.19
40	0	1.02	1.03	1.02	−0.03	−0.05	−0.04	1.18	1.19	1.18	0.03	0.03	0.03
	.2	1.20	1.20	1.21	−0.15	−0.15	−0.15	1.24	1.25	1.24	0.00	−0.01	−0.01
	.4	1.14	1.17	1.17	0.12	0.12	0.14	1.21	1.22	1.21	0.07	0.06	0.07
	.6	1.27	1.28	1.30	0.12	0.12	0.14	1.33	1.34	1.34	0.19	0.20	0.21

Note. ‘Cor’ refers to the simulated correlation between the item parameters in the data. ‘True’ is the mean of the true value of the corresponding parameter, ‘BHM’ denotes the Bayesian hierarchical crossed random effects model, and ‘GLM’ denotes the generalized linear factor model.

Table 2. Standard deviations of the true and estimated parameters for the item parameters of item 1 in the simulation study

		α_i			β_i			ϕ_i			λ_i		
Items	Correlation	True	BHM	GLM	True	BHM	GLM	True	BHM	GLM	True	BHM	GLM
20	0	0.51	0.55	0.56	1.06	1.07	1.12	0.63	0.59	0.65	1.01	1.00	1.00
	.2	0.83	0.79	1.09	1.01	1.02	1.10	0.66	0.61	0.66	0.80	0.80	0.80
	.4	0.63	0.63	0.68	0.97	1.01	1.15	0.60	0.59	0.61	1.04	1.05	1.06
	.6	0.99	0.80	1.18	0.97	0.95	1.02	0.86	0.84	0.86	0.93	0.94	0.94
40	0	0.54	0.54	0.56	0.94	0.96	0.99	0.67	0.68	0.67	0.91	0.90	0.90
	.2	0.82	0.70	0.77	1.06	1.05	1.10	0.77	0.78	0.76	1.01	1.01	1.01
	.4	0.69	0.66	0.72	0.97	0.99	1.02	0.92	0.91	0.96	1.17	1.17	1.17
	.6	0.83	0.72	0.91	0.95	0.91	0.97	0.81	0.78	0.82	0.96	0.97	0.97

Note. 'True' is the standard deviation of the true item parameters. For the other abbreviations, see Table 1.

Table 3. Estimates of item correlations within the BHM

Items	Correlation	Estimates of item correlations					
		(α_i, β_i)	(α_i, ϕ_i)	(α_i, λ_i)	(β_i, ϕ_i)	(β_i, λ_i)	(ϕ_i, λ_i)
20	.0	.05	.08	.05	.06	.04	.04
	.2	.13	.15	.15	.15	.14	.13
	.4	.25	.22	.25	.26	.31	.25
	.6	.34	.30	.34	.35	.41	.36
40	.0	.03	.03	.03	.02	.00	.02
	.2	.16	.16	.18	.15	.18	.18
	.4	.28	.28	.27	.29	.33	.28
	.6	.41	.38	.40	.41	.49	.42

Note. 'Cor' refers to the true correlation. BHM denotes the Bayesian hierarchical crossed random effects model.

correlation of almost 0, this led to the underestimation of the item parameter correlations.[5] A large increase in the number of items (e.g., to 250) will avoid this problem; however, this is unrealistic and not necessary as the main conclusion of this simulation study will not change. That is, the correlations between the item parameters are still large enough to conclude that their presence will not bias results within the generalized linear factor framework, given the parameter ranges we used in this simulation study.

4.2. Conclusion

The simulation show that results of the BHM and GFM are very similar. Specifically, we showed that the first two moments (mean and standard deviation) of the distribution of the parameter estimates of the BHM are very similar to those of the GFM. Therefore we conclude that neglecting the item parameter correlations does not result in bias or inefficiency in the GFM parameter estimates as compared to the BHM.

5. Illustrations

In this section we present two applications of the generalized linear factor model to real data to demonstrate the possibilities and flexibility of this framework. Specifically, we want to illustrate two of the many modelling possibilities that are not feasible in the current implementation of the hierarchical approach of van der Linden (2007). Throughout the illustrations we use marginal maximum likelihood estimation of the model parameter. We assess comparative model fit using Akaike's information criterion (AIC), the Bayesian information criterion (BIC), and the sample size adjusted BIC (sBIC; Sclove, 1987). For these fit indices a smaller value indicates a better model fit. All models are fitted using Mplus (Muthén & Muthén, 2007).

[5] By default the cirt package uses an inverse-Wishart distribution for Σ_I with degrees of freedom equal to 4 and a scale matrix with 10s on the diagonal elements and 1s on the off-diagonal elements.

5.1. Using a latent class measurement model for the responses

In this application we illustrate how the measurement model for the responses could be replaced by a latent class model. If the responses are categorical and a nominal latent trait is assumed to underlie these responses, the probability of a correct response can be modelled in a latent class model by

$$\mathrm{P}(X_{pi} = 1|\theta_p = c) = \Phi(-\beta_{ic}),$$

where θ is the nominal latent class variable with levels $c = 1, \ldots, C$ and β_{ic} is the difficulty of item i in class c. For the log response times we again have the one factor model as discussed before, that is,

$$\ln T_{pi} = \lambda_i + \varphi_i \tau_p + \omega_{pi}.$$

In the present case, to model the association between τ_p and θ_p we cannot use a linear correlation as θ_p is nominal. We therefore estimate the expected value and the variance of τ_p in each class, that is,

$$E(\tau_p|\theta_p = c) = \mu_\tau|c,$$

$$\mathrm{var}(\tau_p|\theta_p = c) = \sigma_\tau^2|c.$$

This enables us to investigate how the average speed, τ_p, varies across classes, θ. To identify the model, $\mu_\tau \mid c$ and $\sigma_\tau^2|c$ can be fixed to 0 and 1 respectively in an arbitrary class. Given that the number of classes, C, is limited, the model is then identified. However, in the case of a large number of classes, the latent class part of the model needs to be restricted using either pragmatic or theoretically motivated restrictions. We will illustrate this below.

5.1.1. Data

We analysed the balance scale data from Van der Maas and Jansen (2003). The data used here comprise the responses and response times of 191 children to 70 balance scale problems. Specifically, each item displays a picture of a balance scale with equally heavy weights placed at pegs situated at equal distance from the fulcrum. The items differed in how many weights are placed at each arm; and at which pegs the weights were placed. The weights are configured to create seven item types (see below). For each item type, 10 items exist. The children had to decide whether the balance scale would tip to the left, tip to the right, or remain in balance.

A suitable model is formulated in accordance with the theory outlined in van der Maas and Jansen (2003). According to this theory, children use one of five possible solution strategies to solve the balance scale items. The success of the strategy being used will depend on the item type the item belongs to. Table 4 depicts the predicted result – 0 (fail), 1 (pass) or guess – for the different kinds of items and for each strategy. As can be seen, from the five strategies, strategy I is the worst as the use of this strategy is associated with the most fails. Specifically, this strategy only involves comparing the weights between the two arms of the balance scale, neglecting the distance of the weights to the fulcrum. In contrast, strategy IV is the best; it uses all available information correctly to make a

Table 4. Predicted result – 0 (fail), 1 (pass) or guess – when using one of the strategies for each item type

Item type	Strategy I	Strategy II	Strategy III	Strategy IV	Strategy V
Simple balance	1	1	1	1	1
Simple weight	1	1	1	1	1
Simple distance	0	1	1	1	1
Conflict balance A	0	0	Guess	1	1
Conflict balance B	0	0	Guess	0	1
Conflict distance	0	0	Guess	1	1
Conflict weight	1	1	Guess	0	1

decision on the balance scale item. Children using strategy III guess on so-called conflict items (i.e., they perform at chance level). While in the other suboptimal strategies (I, II, and IV) children will score below chance level as they think they know the correct answer while in fact it is wrong.

The model for the responses thus contains five latent classes (one for each solution strategy). Here we restrict the model to the probability structure in Table 4. In addition, the 10 items of each type are assumed to be homogeneous. That is, we assume the same difficulty parameter, β_{ic}, for items of the same item type. In addition, the thresholds of a given item type are assumed equal across classes. This reflects the postulation that the difficulty does not depend on the strategy you use. If there are multiple strategies that result in a correct response (e.g., strategies IV and V for a conflict balance item), the correct information is used in both strategies so there is no difference in item difficulty to be expected. See Table 5 for the resulting parameter configuration in the response time model.

In the measurement model for the response times, the intercepts and residual variances are assumed to be equal across classes and across the different item types. In addition, similarly to van der Linden (2007), we fixed the factor loadings, ϕ_i, to unity. The resulting response time model thus contains seven σ^2_ω parameters, seven ν_i parameters (one for each item type), six $\mu_\tau \mid c$ parameters and six $\sigma^2_\tau|c$ parameters (as we fixed the mean and variance in the first class to be equal to 0 and 1 respectively, as discussed above). The full model contains, in addition to the parameters discussed above, six class proportion parameters.

Table 5. Parameter configuration for the latent class response model

Item type	Class I	Class II	Class III	Class IV	Class V
Simple balance	β_{11}	β_{11}	β_{11}	β_{11}	β_{11}
Simple weight	β_{21}	β_{21}	β_{21}	β_{21}	β_{21}
Simple distance	$-\beta_{31}$	β_{31}	β_{31}	β_{31}	β_{31}
Conflict balance A	$-\beta_{41}$	$-\beta_{41}$	0.41 *	β_{41}	β_{41}
Conflict balance B	$-\beta_{51}$	$-\beta_{51}$	0.41 *	$-\beta_{51}$	β_{51}
Conflict distance	$-\beta_{61}$	$-\beta_{61}$	0.41 *	β_{61}	β_{61}
Conflict weight	β_{71}	β_{71}	0.41 *	$-\beta_{71}$	β_{71}

Note. *This value is fixed to reflect guessing, i.e., $P(X_{pi} \mid \theta_p = 3) = \Phi(-\beta_{ic}) = \Phi(-0.41) \approx .34$.

Table 6. Class proportions, mean speed ($\mu_\tau \mid c$), variance of speed ($\sigma^2_\tau|c$), and probability correct P ($X_{pi} \mid \theta_p = c$) for each item type within each class

	Class I	Class II	Class III	Class IV	Class V
Class proportions	0.349	0.205	0.331	0.052	0.063
$\mu_\tau \mid c$	0*	0.406 (0.253)	1.004 (0.316)	1.372 (0.447)	1.426 (0.358)
$\sigma^2_\tau|c$	1*	0.914 (0.410)	1.386 (0.412)	1.033 (0.465)	0.341 (0.208)
$P(X_{pi}|\theta_p = c)$					
Simple balance	0.973	0.973	0.973	0.973	0.973
Simple weight	0.953	0.953	0.953	0.953	0.953
Simple distance	0.107	0.893	0.893	0.893	0.893
Conflict balance A	0.026	0.026	0.34*	0.974	0.974
Conflict balance B	0.035	0.035	0.34*	0.035	0.965
Conflict distance	0.099	0.099	0.34*	0.901	0.901
Conflict weight	0.939	0.939	0.34*	0.061	0.939

Note. *These parameters are fixed. The number in parentheses denotes standard errors.

5.1.2. Results

Table 6 displays the modelling results. For ease of interpretation, the estimated difficulties in the response time model, β_{ic}, are converted to proportions. As can be seen, the largest classes are the class with strategy I children (i.e., the children who use the simplest strategy) and the class with strategy III children (the children who guess on conflict item types). The class of children who use the most sophisticated and successful strategy V is relatively small. In Table 6, the mean and variance of the speed factor is shown within each class. As can be seen, the most sophisticated strategy V is associated with slower responses as all information available in the item needs to be evaluated. Strategies I and II are more fast as less information is taken into account (i.e., the items appear easier than they are).

By imposing a separate speed factor for the items of the same item type, it is also possible to estimate the mean speed for each item type separately (see Figure 3 for the results). Most notably, it can be seen that – as compared to strategy I – for the more sophisticated strategies III, IV, and V, the mean on the speed factor is larger for the conflict items but not for the simple items. Thus, the speed factor provides important information concerning the nature of the different strategies being used to solve the balance scale problems.

5.2. Modelling a non-linear relation between speed and ability

In the hierarchical model for responses and response times the relation between the latent ability and speed variables is modelled via a linear correlation coefficient. A linear relation between speed and ability might be a reasonable approximation for power tests (e.g., an arithmetic test). In such tests it can be expected that the higher a subject's position on the ability, the easier the items of the test are perceived, and therefore the faster the items can be solved. However, for binary personality and attitude items (bipolar items), this assumption is less reasonable as subjects with an extreme position on the personality trait, θ_p, are likely to respond faster than subjects whose position on the personality trait is more in the middle. For instance, the item 'I like to be the centre of attention', which purports to measure extraversion, is likely to be answered quickly by highly extroverted subjects (a fast 'yes') and by highly introverted people (a fast 'no'), but subjects with an average position on this dimension need more time to make this dichotomous decision.

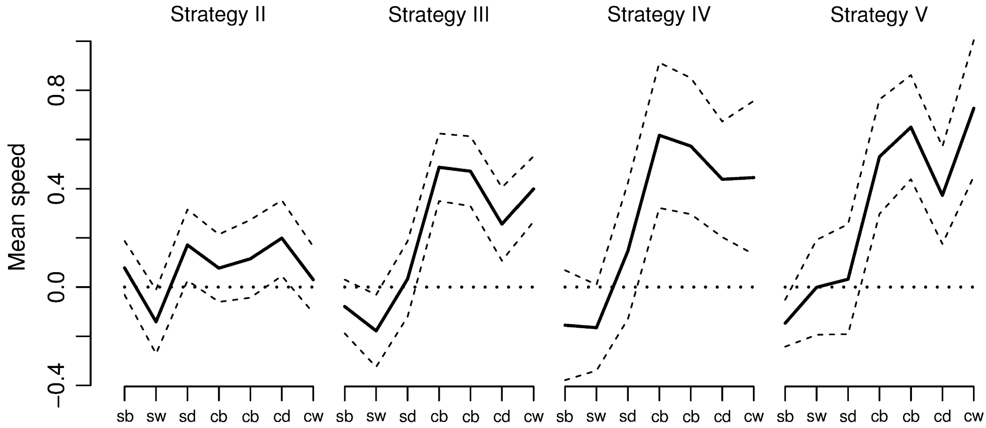

Figure 3. Estimated means of the speed factor for each item type and each strategy (solid lines) and 99% confidence bounds (dashed lines). The dotted lines depict the zero baseline means in the reference class, strategy I. In addition, sb = simple balance; sw = simple weight; sd = simple distance; cb = conflict balance; cd = conflict distance; and cw = conflict weight.

This postulation is related to the distance-difficulty hypothesis (Ferrando & Lorenzo-Seva, 2007; see also Ranger & Kuhn, 2012) which predicts that when $\theta_p - \beta_i$ is small (i.e., the item difficulty matches the ability of a given person), response times will be longer than when $\theta_p - \beta_i$ is larger. This effect is supported in various empirical studies (e.g., Holden, Fekken, & Cotton, 1991; Kuncel, 1973; Tuerlinckx *et al.*, 2012). Given that the item difficulties are more or less normally distributed across the ability continuum, the distance-difficulty hypothesis will result in an inverted U-shape relation between the latent speed and ability variables in personality data.

Within the generalized linear factor framework it is straightforward to model non-linear relations among latent variables (see Bauer, 2005; Klein & Moosbrugger, 2000; Klein & Muthén, 2007; Marsh, Zhonglin, & Kit-Tai, 2004). Using this methodology we can test for non-linear relations between the latent speed and ability variable in the hierarchical model for responses and response times. We will illustrate this using a data set from Ferrando and Lorenzo-Seva (2007). The data comprise the scores of 750 subjects on 11 extraversion items and 11 neuroticism items (for details, see Ferrando & Lorenzo-Seva, 2007). We fitted two models to the data of the extraversion and neuroticism scales: the generalized linear factor model with random person effects and a linear correlation between the speed and ability variables; and the generalized linear factor model in which we regressed the latent speed factor curvilinear on the ability factor, that is,

$$\tau_p = \zeta_1 \theta_p + \zeta_2 \theta_p^2 + \delta_p. \tag{8}$$

In equation (8), ζ_1 reflects the linear effect of θ_p on τ_p, ζ_2 reflects the quadratic effect, and δ_p is a residual term. Note that we do not have an intercept as this term is not identified in a single group application. In addition, by fixing ζ_2 to 0, the model is equivalent to the frequentist hierarchical model without random item effects as studied in Section 4, as both τ_p and θ_p are standardized ζ_1 will be equal to the linear correlation between ability and speed, $\rho_{\theta\tau}$. We fitted both the linear and the curvilinear models to the extraversion and neuroticism data separately.

5.2.1. Results

We fitted both the linear baseline model and the curvilinear model in equation (8). In Table 7, the AIC, BIC, and sBIC fit indices, together with the parameter estimates for $\rho_{\theta\tau}$ (linear baseline model) and ζ_1 and ζ_2 (curvilinear model), are depicted for both models. As can be seen, for both extraversion and neuroticism the AIC, BIC, and sBIC favour the curvilinear model over the linear model, indicating that the relation between speed and ability is not strictly linear. This is also evident from the significant parameter estimate of the quadratic effect, ζ_2 for both personality scales. In Figure 4 the predicted relation between the speed, τ_p, and ability, θ_p, variables is depicted using the parameter estimates from the curvilinear model in Table 7 for the extraversion and neuroticism scales. In addition, the figure contains a plot of the estimated factor scores for τ_p and θ_p. As can be seen, the inverted U-shape is clearly found, with faster responses at the extremes of the extraversion and neuroticism dimensions.

This result is of interest from both a statistical and a substantive perspective. First, it illustrates that the relation between speed and ability is not necessarily linear, as commonly modelled in the hierarchical model framework by means of a linear correlation coefficient. Neglecting departures from linearity may bias parameter estimates and goodness-of-fit measures. It is therefore advisable to test the underlying relation between

Table 7. Model fit indices and parameter estimates for the linear are curvilinear models in Section 5.2

	Baseline model (linear)				Curvilinear model				
Scale	AIC	BIC	sBIC	$\rho_{\theta\tau}$	AIC	BIC	sBIC	ζ_1	ζ_2
Extraversion	19,442	19,700	19,522	0.11 (0.05)	19,411	19,674	19,493	0.10 (0.05)	0.23 (0.04)
Neuroticism	19,024	19,283	19,105	0.23 (0.05)	18,995	19,258	19,077	0.29 (0.06)	0.23 (0.06)

Note. Standard errors are in parentheses.

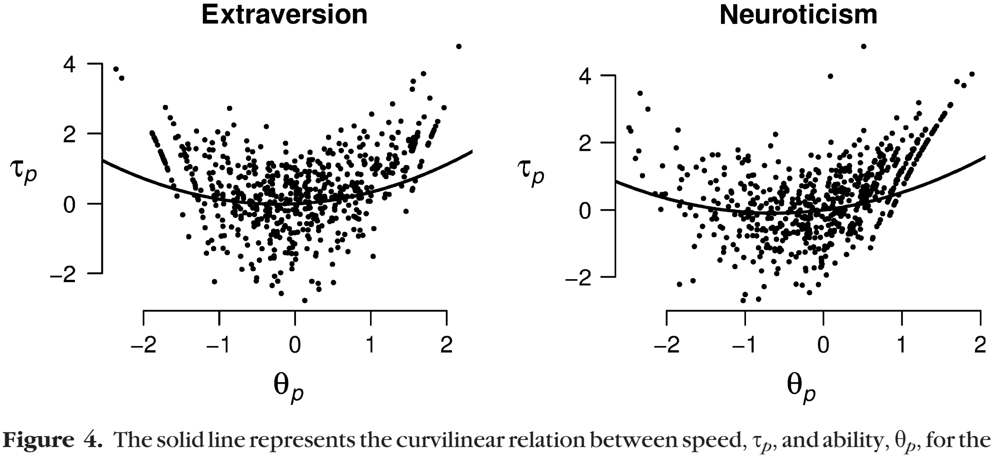

Figure 4. The solid line represents the curvilinear relation between speed, τ_p, and ability, θ_p, for the extraversion scale (left) and the neuroticism scale (right). The dots represent the estimated factor scores for τ_p and θ_p. We have omitted one outlying case with τ_p estimated to be 6.58 and θ_p estimated to be 3.90 for the neuroticism scale.

speed and ability on non-linearity. Second, from a substantive perspective, the result suggests that the distance-difficulty hypothesis – which is commonly formulated at item level – can also occur at test level. Of course the occurrence of this effect depends on the distribution of the item difficulties across the ability scale. For instance, in the (unlikely) case that item difficulties are only located at one extreme of the ability scale, a linear relation would be found between speed and ability.

6. Discussion

The goal of the present study was to advance the use of the hierarchical framework developed by van der Linden (2007), Fox *et al.* (2007), Klein Entink, Fox, *et al.* (2009), and Glas and van der Linden (2010). This was achieved by making explicit that the hierarchical model can be seen as a generalized linear factor model with two factors for the responses and log response times. The major asset of our approach is that by showing how the hierarchical model can be formulated within a generalized linear factor model framework, one can employ a variety of tools and extensions that are available in this context.

For pragmatic reasons, we omitted the correlated random item effects from the hierarchical model to enable the use of flexible latent variable modelling software to estimate the model. It turned out that this omission did not affect parameter estimates substantially. However, one should be carefully consider whether it is justified to omit these effects, as there may be substantive reasons for the inclusion of (possibly correlated) random item effects. De Boeck (2008) distinguishes three different reasons for random item effects to arise in a given data set. First, when items are sampled from a larger set of items (e.g., words from a larger word list), sampling variance rises in the data that should be taken into account (see Raaijmakers, Schrijnemakers, & Gremmen, 1999). Second, in some paradigms researchers work with item families. Item within a family are considered highly similar (e.g., basic addition items), as compared to items from a different family (e.g., basic multiplication items). As families are relatively homogenous, one can focus on the item family parameters instead of the individual item parameters, which introduces a notion of random items (see Janssen, Tuerlinckx, Meulders, & De Boeck, 2000). Third, as noted by De Boeck (2008), in longitudinal designs one may rely on randomly drawn items from a larger pool of items to prevent the same items being administered on the different time points (see Albers, Does, Ombos, & Janssen, 1989). Clearly, the above examples require the inclusion of random item effects for substantive reasons.

References

Albers, W., Does, R. J. M. M., Ombos, T., & Janssen, M. P. E. (1989). A stochastic growth model applied to tests of academic knowledge. *Psychometrika*, *54*, 451–466. doi:10.1007/BF02294628

Arbuckle, J. L. (1997). *Amos (version 3.61) [computer software]*. Chicago: Small Waters.

Asparouhov, T., & Muthén, B. (2012, May). General random effect latent variable modeling: Random subjects, items, contexts, and parameters. Paper presented at the third UK Mplus Users' Meeting, London, UK.

Bauer, D. J. (2005). A semiparametric approach to modeling nonlinear relations among latent variables. *Structural Equation Modeling*, *12*, 513–535. doi:10.1207/s15328007sem1204_1

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In E. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (Chapters 17–20). Reading, MA: Addison Wesley.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*, 443–459. doi:10.1007/BF02293801

Boker, S., Neale, M. C., Maes, H. H., *et al.* (2010). OpenMx: An open source extended structural equation modeling framework. *Psychometrika*, *76*, 306–317. doi:10.1007/s11336-010-9200-6

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York, NY: Wiley.

Burnham, K. P., & Anderson, D.R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York: Springer-Verlag.

Cox, D. R., & Miller, H. D. (1970). *The theory of stochastic processes*. London: Methuen.

De Boeck, P. (2008). Random item IRT models. *Psychometrika*, *73*, 533–559. doi:10.1007/s11336-008-9092-x

Dolan, C. V., van der Maas, H. L. J., & Molenaar, P. C. M. (2002). A framework for ML estimation of parameters of (mixtures of) common reaction time distributions given optional truncation or censoring. *Behavior Research Methods*, *34*, 304–323. doi:10.3758/BF03195458

Ferrando, P. J., & Lorenzo-Seva, U. (2007). An item response theory model for incorporating response time data in binary personality items. *Applied Psychological Measurement*, *31*, 525–543.

Fox, J.-P. (2010). *Bayesian item response modeling: Theory and applications*. New York: Springer.

Fox, J.-P., Klein Entink, R. H., & van der Linden, W. J. (2007). Modeling of responses and response times with the package cirt. *Journal of Statistical Software*, *20*, 1–14.

Glas, C. A., & van der Linden, W. J. (2010). Marginal likelihood inference for a model for item responses and response times. *British Journal of Mathematical and Statistical Psychology*, *63* (3), 603–626. doi:10.1348/000711009X481360

Holden, R. R., Fekken, C. G., & Cotton, D. H. G. (1991). Assessing psychopathology using structured test-item response latencies. *Psychological Assessment*, *3*, 111–118. doi:10.1037/1040-3590.3.1.111

Janssen, R., Tuerlinckx, F., Meulders, M., & De Boeck, P. (2000). A hierarchical IRT model for criterion-referenced measurement. *Journal of Educational and Behavioral Statistics*, *25*, 285–306. doi:10.3102/10769986025003285

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modelling with the SIMPLIS command language*. Chicago: Scientific Software International.

Klein Entink, R. H., Fox, J.-P., & van der Linden, W. J. (2009). A multivariate multilevel approach to the modeling of accuracy and speed of test takers. *Psychometrika*, *74*, 21–48. doi:10.1007/s11336-008-9075-y

Klein Entink, R. H., Kuhn, J.-T., Hornke, L. F., & Fox, J.-P. (2009). Evaluating cognitive theory: A joint modeling approach using responses and response times. *Psychological Methods*, *14*, 54–75. doi:10.1037/a0014877

Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, *65*, 457–474. doi:10.1007/BF02296338

Klein, A. G., & Muthén, B.O. (2007). Quasi-maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research*, *42*, 647–673. doi:10.1080/00273170701710205

Kuncel, R. B. (1973). Response processes and relative location of subject and item. *Educational and Psychological Measurement*, *33*, 545–563. doi:10.1177/001316447303300302

Lord, F.M. (1952). *A theory of test scores* (Psychometric Monograph No. 7). Iowa City, IA: Psychometric Society.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000) WinBUGS – a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, *10*, 325–337.

Marsh, H. W., Zhonglin, W., & Kit-Tai, H. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological methods*, *9*, 275–300. doi:10.1037/1082-989X.9.3.275

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149–174. doi:10.1007/BF02296272

Mehta, P. D., Neale, M. C., & Flay, B. R. (2004). Squeezing interval change from ordinal panel data: Latent growth curves with ordinal outcomes. *Psychological Methods*, *9*, 301–333. doi:10.1037/1082-989X.9.3.301

Mellenbergh, G. J. (1994a). Generalized linear item response theory. *Psychological Bulletin*, *115*, 300–307. doi:10.1037/0033-2909.115.2.300

Mellenbergh, G. J. (1994b). A unidimensional latent trait model for continuous item responses. *Multivariate Behavioral Research*, *29*, 223–236. doi:10.1207/s15327906mbr2903_2

Molenaar, D., Dolan, C. V., & De Boeck, P. (2012). The heteroscedastic graded response model with a skewed latent trait: Testing statistical and substantive hypotheses related to skewed item category functions. *Psychometrika*, *77*, 455–478. doi:10.1007/s11336-012-9273-5

Muthén, L. K., & Muthén, B. O. (2007). *Mplus user's guide* (5th ed.). Los Angeles: Muthén & Muthén.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2006). *Mx: Statistical modeling* (7th ed.). Richmond, VA: VCU, Department of Psychiatry.

R Development Core Team (2007). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org.

Raaijmakers, J., Schrijnemakers, J., & Gremmen, F. (1999). How to deal with 'the language-as fixed-effect-fallacy': Common misconceptions and alternative solutions. *Journal of Memory and Language*, *41*, 416–426. doi:10.1006/jmla.1999.2650

Ranger, J., & Kuhn, J. T. (2012). Improving item response theory model calibration by considering response times in psychological tests. *Applied Psychological Measurement*, *36*, 214–231. doi:10.1177/0146621612439796

Ranger, J., & Ortner, T. (2012). The case of dependency of responses and response times: A modeling approach based on standard latent trait models. *Psychological Test and Assessment Modeling*, *54*, 128–148.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Paedogogiske Institut.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108. doi:10.1037/0033-295X.85.2.59

Samejima, F. (1969). Estimation of ability using a response pattern of graded scores. *Psychometrika Monograph No. 17*. doi:10.1007/BF02291114

Samejima, F. (1973). Homogeneous case of the continuous response level. *Psychometrika*, *38*, 203–219.

SAS Institute Inc. (2011). *SAS/STAT software: Release 9.3*. Cary, NC: SAS Institute, Inc.

Savalei, V. (2006). Logistic approximation to the normal: The KL rationale. *Psychometrika*, *71*, 763–767. doi:10.1007/s11336-004-1237-y

Sclove, L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*, 333–343. doi:10.1007/BF02294360

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Chapman and Hall/CRC.

Spearman, C. E. (1904). 'General intelligence' objectively determined and measured. *American Journal of Psychology*, *15*, 201–293. doi:10.2307/1412107

Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, *52*, 393–408. doi:10.1007/BF02294363

Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.

Tuerlinckx, F., & De Boeck, P. (2005). Two interpretations of the discrimination parameter. *Psychometrika*, *70*, 629–650. doi:10.1007/s11336-000-0810-3

Tuerlinckx, F., Molenaar, D., & van der Maas, H. L. (2012). Diffusion-based item response theory. Manuscript submitted for publication.

van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika*, *72*(3), 287–308. doi:10.1007/s11336-006-1478-z

van der Linden, W. J. (2008). Using response times for item selection in adaptive testing. *Journal of Educational and Behavioral Statistics*, *33*, 5–20. doi:10.3102/1076998607302626 van der Linden, W. J. (2009). A bivariate lognormal response-time model for the detection of collusion between two test takers. *Journal of Educational and Behavioral Statistics*, *34*, 378–394. doi:10.3102/1076998609332107

van der Linden, W. J., Breithaupt, K., Chuah, S. C., & Zhang, Y. (2007). Detecting differential speededness in multistage testing. *Journal of Educational Measurement*, *44*, 117–130. doi:10.1111/j.1745-3984.2007.00030.x

van der Linden, W. J., & Guo, F. (2008). Bayesian procedures for identifying aberrant response-time patterns in adaptive testing. *Psychometrika*, *73*, 365–384. doi:10.1007/s11336-007-9046-8

van der Maas, H. L., & Jansen, B. R. (2003). What response times tell of children's behavior on the balance scale task. *Journal of Experimental Child Psychology*, *85*, 141–177. doi:10.1016/S0022-0965(03)00058-4

van der Maas, H. L., Molenaar, D., Maris, G., Kievit, R. A., & Borsboom, D. (2011). Cognitive psychology meets psychometric theory: On the relation between process models for decision making and latent variable models for individual differences. *Psychological Review*, *118*(2), 339–177. doi:10.1037/a0022749

Wagenmakers, E.-J., Lee, M. D., Lodewyckx, T., & Iverson, G. (2008). Bayesian versus frequentist inference. In H. Hoijtink, I. Klugkist & P. A. Boelen (Eds.), *Bayesian evaluation of informative hypotheses* (pp. 181–207). New York: Springer.

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, *12*, 58–79. doi:10.1037/1082-989X.12.1.58

Wood, R., Wilson, D. T., Gibbons, R. D., Schilling, S. G., Muraki, E., & Bock, R. D. (2002). *TESTFACT: Test scoring, item statistics, and item factor analysis*. Chicago: Scientific Software International.

Received 28 February 2014; revised version received 21 May 2014

Appendix A: OpenMx syntax to fit the generalized linear factor model for responses and response times

```
# General settings
nit=20          #number of items
nfac=2          #number of factors equals 2 (latent speed and ability factor)

# Setting things up
library(OpenMx)
nms=c(paste(''y'',1:(nit),sep=''''),paste(''t'',1:(nit),sep=''''))
dimnames(data)[[2]]<-nms

# Preparing data
data=data.frame(data)
data[,1:nit] <- mxFactor(data[,1:nit], levels=c(0, 1))

# Bulding up the OpenMx model
GFM = mxModel(''General'',
mxMatrix(type=''Full'', nrow=2*nit, ncol=nfac, values=c(
   1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1, 0,1,0,1,0,1,0,1,0,
   0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1, 0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1),byrow=T, free=c(
   T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,
   F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T,F,T),name=''lamb''),          # matrix of time discrimination
mxMatrix(type=''Diag'', nrow=2*nit, ncol=2*nit, free=c(
   F,F,F,F,F,F,F,F,F,F,F,F,F,F,F,F,F,F,F,F,T,T,T,T,T,T,T,T,T,T,T,T,T,T,T,T,T,T,T,T),
values=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1), name=''resVar''), # matrix of residual variances
mxMatrix(type=''Full'', nrow=1, ncol=2*nit, values=0, free=c(
   F,F,F,F,F,F,F,F,F,F,F,F,F,F,F,F,F,F,F,F,T,T,T,T,T,T,T,T,T,T,T,T,T,T,T,T,T,T,T,T), name=''upsil''),        # time intensity parameters
mxMatrix(type=''Full'', nrow=1, ncol=nit, values=1, free=T, name=''thres''),                                # item difficulty parameters
mxMatrix(type=''Symm'', nrow=nfac, ncol=nfac, values=c(1,0,0,1), free=c(F,T,T,F), name=''varFact''),         # covariance matrix of the
                                                                                                              latent
```

```
                                                                                                   # speed and ability variables
mxMatrix(type=''Full'', nrow=1, ncol=nfac, values=0, free=F, name=''meanFact''),                 # mean vector of latent speed
                                                                                                   #and ability variables
mxAlgebra(expression= upsil + t(lamb%*%t(meanFact)), name=''Mean''),                             # predicted mean vector
mxAlgebra(expression= lamb%*%varFact%*%t(lamb)+resVar, name=''Cov''),                           # predicted covariance matrix
mxFIMLObjective(covariance=Cov, means=Mean,dimnames=nms,threshnames=nms[1:nit],
 thresholds=thres),                                                                                # setting up the likelihood
mxData(data,type=''raw'')
)
results -->
summary(results)
```

Appendix B: MPlus syntax to fit the generalized linear factor model for responses and response times

```
TITLE:        generalized linear factor model for responses and response times
DATA:         FILE IS data.dat;
VARIABLE:     NAMES ARE y1-y20 t1-t20;
              CATEGORICAL ARE y1-y20;
ANALYSIS:     ALGORITHM=INTEGRATION;
MODEL:        abil BY y1-y20*;
              abil@1;
              speed BY t1-t20*;
              speed@1;
OUTPUT:       MODINDICES;
```