

Simple structure detection through Bayesian Exploratory Multidimensional IRT models

Lara Fontanella¹, Sara Fontanella², Pasquale Valentini³, and Nickolay Trendafilov⁴

¹Department of Legal and Social Sciences, University of Chieti-Pescara

²Department of Medicine, Imperial College London

³Department of Economics, University of Chieti-Pescara

⁴Department of Mathematics & Statistics, The Open University

Supplementary materials organisation

This document is organised in four Sections. Section 1 presents the pseudo-code for SSVS-MIRT and CSTR-MIRT algorithms. In Section 2, we provide the details on the parameters adopted in the simulation study described in the paper along with complete results for $N = 250$ and $N = 1000$. In Section 3 we present a simulation study aimed at assessing the performance of the proposed methods for different numbers of items and latent traits and various levels of sparsity. Details on the simulation settings are provided in Section 3.1, while detailed results are presented in Sections 3.2 and 3.3. The proposed methods show good performances also in the presence of cross-loadings in the latent structure. Finally, Section 4 presents the detailed results for the application on the dataset related to the Humour Scale Questionnaire.

1 Estimation algorithms

In this Section we describe the MCMC algorithm for the Bayesian implementation of the MIRT model. In particular, we present the MCMC scheme for both the SSVS and the CSTR approaches. In the description of the MCMC

steps, we make use of the following notation: \mathbf{X} is the $(N \times K)$ matrix of the responses for N persons to K categorical items, rated on a C -points ordered scale; \mathbf{Z} is the $(N \times K)$ matrix containing the underlying variable scores; \mathbf{z}_i denotes the K -dimensional vector of scores for subject i , while \mathbf{z}_k denotes the N -dimensional vector of scores for item k ; \mathbf{A} is the $(K \times M)$ matrix of discrimination parameters and $\boldsymbol{\alpha}_k$ is the M -dimensional vector of loadings for item k ; $\boldsymbol{\Theta}$ is the $(N \times M)$ matrix containing the scores on the latent traits and $\boldsymbol{\theta}_i$ is the M -dimensional score vector for person i , while $\boldsymbol{\theta}_m$ is the N -dimensional vector for the latent variable m ; finally $\boldsymbol{\gamma}_k = (\gamma_{k,0} \gamma_{k,1} \dots \gamma_{k,C})'$, with $\gamma_{k,0} = -\infty$ and $\gamma_{k,C} = \infty$, denotes the threshold vector. The subsequent MCMC steps are shared by both methods, unless otherwise specified.

1. Sample $Z_{i,k} | \boldsymbol{\alpha}_k, \boldsymbol{\theta}_i, \boldsymbol{\gamma}_k, \mathbf{X}$

The underlying variable score, $Z_{i,k}$, is drawn from a truncated normal distribution implied by the threshold model

$$Z_{i,k} | \boldsymbol{\alpha}_k, \boldsymbol{\theta}_i, \boldsymbol{\gamma}_k, \mathbf{X} \sim \mathcal{N}(\boldsymbol{\alpha}'_k \boldsymbol{\theta}_i, 1) I(\gamma_{k,c-1} < Z_{i,k} < \gamma_{k,c-1}) \quad \text{if } X_{i,k} = c$$

2. Sample $\boldsymbol{\theta}_i | \mathbf{z}_i, \mathbf{A}, \boldsymbol{\Sigma}_{\theta}$

The full conditional is normal

$$\boldsymbol{\theta}_i | \mathbf{z}_i, \mathbf{A}, \boldsymbol{\Sigma}_{\theta} \sim \mathcal{N}_M((\mathbf{A}'\mathbf{A} + \boldsymbol{\Sigma}_{\theta}^{-1})^{-1}\mathbf{A}'\mathbf{z}_i, (\mathbf{A}'\mathbf{A} + \boldsymbol{\Sigma}_{\theta}^{-1})^{-1})$$

For the CSTR estimation procedure, we set $\boldsymbol{\Sigma}_{\theta} = \mathbf{I}_M$.

3. Sample $\boldsymbol{\alpha}_k = (\alpha_{k,1} \dots \alpha_{k,M})'$

- CSTR-MIRT: Sample $\boldsymbol{\alpha}_k | \mathbf{z}_k, \boldsymbol{\Theta}$

The full conditional is normal

$$\boldsymbol{\alpha}_k | \mathbf{z}_k, \boldsymbol{\Theta} \sim \mathcal{N}_M((\boldsymbol{\Theta}'\boldsymbol{\Theta} + \mathbf{I}_M)^{-1}(\boldsymbol{\Theta}'\mathbf{z}_k), (\boldsymbol{\Theta}'\boldsymbol{\Theta} + \mathbf{I}_M)^{-1})$$

- SSVS-MIRT

- Sample $\boldsymbol{\alpha}_k | \mathbf{z}_k, \boldsymbol{\Theta}, \{\zeta_{k,m}\}$

The normal mixture prior for $\boldsymbol{\alpha}_k$ can be written as (George & McCulloch, 1993):

$$\boldsymbol{\alpha}_k | \boldsymbol{\zeta}_k \sim \mathcal{N}_M(\mathbf{0}, \mathbf{D}_{\zeta_k} \mathbf{R} \mathbf{D}_{\zeta_k})$$

where $\boldsymbol{\zeta}_k = (\zeta_{k,1} \dots \zeta_{k,M})'$ and $\mathbf{D}_{\zeta_k} = \text{diag}(d_{k,1}\tau, \dots, d_{k,M}\tau)$ with $d_{k,m} = 1$ if $\zeta_{k,m} = 0$ and $d_{k,m} = g$ if $\zeta_{k,m} = 1$. Here as prior correlation we assume the identity matrix $\mathbf{R} = \mathbf{I}_M$. Under this prior, the full conditional for $\boldsymbol{\alpha}_k$ is normal, therefore we draw the discrimination parameters from

$$\boldsymbol{\alpha}_k | \mathbf{z}_k, \boldsymbol{\Theta}, \{\zeta_{k,m}\} \sim \mathcal{N}_M((\boldsymbol{\Theta}'\boldsymbol{\Theta} + \mathbf{D}_{\zeta_k}^{-2})^{-1}\boldsymbol{\Theta}'\mathbf{z}_k, (\boldsymbol{\Theta}'\boldsymbol{\Theta} + \mathbf{D}_{\zeta_k}^{-2})^{-1})$$

- Sample $\zeta_k | \alpha_k$

The vector ζ_k is obtained componentwise by sampling from the posterior distribution $\zeta_{k,m} \sim f(\zeta_{k,m} | \alpha_k, \zeta_{-k})$. If $\mathbf{R} = \mathbf{I}_M$ then the dependence of the posterior from ζ_{-k} can be removed, and each conditional distribution is Bernoulli, with probability

$$\begin{aligned} P(\zeta_{k,m} = 1 | \alpha_{k,m}) &= \frac{f(\alpha_{k,m} | \zeta_{k,m} = 1) p_k}{f(\alpha_{k,m} | \zeta_{k,m} = 1) p_k + f(\alpha_{k,m} | \zeta_{k,m} = 0) (1 - p_k)} \\ &= \frac{(g\tau)^{-1} \exp\left(-\frac{\alpha_{k,m}^2}{2g^2\tau^2}\right) p_k}{\tau^{-1} \exp\left(-\frac{\alpha_{k,m}^2}{2\tau^2}\right) (1 - p_k) + (g\tau)^{-1} \exp\left(-\frac{\alpha_{k,m}^2}{2g^2\tau^2}\right) p_k} \end{aligned}$$

4. Sample $\gamma_k | \{\theta_i\}, \{\alpha_k\}$

To draw the threshold parameters for a given item, we consider a Metropolis-Hastings step based on Cowles' algorithm (Cowles, 1996), and draw the candidate from

$$\gamma_{k,c}^* \sim \mathcal{N}(\gamma_{k,c}, \sigma_{MH}^2) I(\gamma_{k,c-1}^* < \gamma_{k,c} < \gamma_{k,c+1}) \text{ for } c = 1, \dots, C - 1$$

The tuning parameter σ_{MH}^2 has been set to obtain an acceptance rate of about 40%. The Metropolis-Hastings acceptance probability is then given by

$$\min \left[\prod_{i=1}^N \frac{Pr(X_{i,k} = c | \theta_i, \alpha_k, \gamma_k^*) f(\gamma_k | \gamma_k^*, \sigma_{MH}^2)}{Pr(X_{i,k} = c | \theta_i, \alpha_k, \gamma_k) f(\gamma_k^* | \gamma_k, \sigma_{MH}^2)}, 1 \right]$$

5. SSVS-MIRT: Sample $\Sigma_\theta | \{\theta_i\}$

To directly estimate the correlation matrix Σ_θ , we adapted the PX-RPMH algorithms proposed by Liu & Daniels (2006). First, an unrestricted covariance matrix is drawn from an Inverse-Wishart distribution, $\Sigma_\theta^* \sim (v, \mathbf{S})$, with degrees of freedom being the sample size, $v = N$, and scale matrix $\mathbf{S} = \sum_{i=1}^N \theta_i^* \theta_i^{*\prime}$, where $\theta_i^* = \mathbf{P} \theta_i$. The sampled Σ_θ^* is then translating back to Σ_θ through $\Sigma_\theta = \mathbf{P}^{-1} \Sigma_\theta^* \mathbf{P}^{-1}$. The provisional Σ_θ is accepted based on a Metropolis Hastings (MH) acceptance probability α , which at iteration r is given by

$$\alpha = \min \left[1, \exp \left(\frac{M+1}{2} (\log |\Sigma_\theta| - \log |\Sigma_\theta^{(r-1)}|) \right) \right]$$

In the CSTR approach, the output of the unconstrained MCMC sampler undergoes a Consensus Simple Target Rotation (Lorenzo-Seva et al., 2002) to obtain an oblique rotation of the MCMC draws with respect to a common sparse target matrix. The oblique rotation can be implemented using the routine promin available in the PsychologicalTestToolbox for matlab (Navarro et al., 2016). The semi-specified target matrix is derived from a consensus matrix obtained through the Generalised Procrustes Analysis (GPA; Gower, 1975).

Denoting with $\{\mathbf{A}_r\}_{r=1}^R$ the sequence of MCMC draws, the GPA yields the “consensus matrix” \mathbf{A}^* by minimising the quadratic loss function:

$$\arg \min_{\{\{\mathbf{U}_r\}_{r=1}^R, \mathbf{A}^*\}} \sum_{r=1}^R \text{tr} [(\mathbf{A}_r \mathbf{U}_r - \mathbf{A}^*)' (\mathbf{A}_r \mathbf{U}_r - \mathbf{A}^*)] \quad \text{s.t. } \mathbf{U}_r' \mathbf{U}_r = \mathbf{I}$$

This optimisation problem can be solved by an iterative procedure, based on an initial choice of the consensus matrix.

Initialising \mathbf{A}^* equal to the last draw of the unconstrained MCMC algorithm, for $r = 1, \dots, R$

1. define $\mathbf{S}_r = \mathbf{A}'_r \mathbf{A}^*$
2. perform the singular value decomposition $\mathbf{S}_r = \mathbf{Q}_r \Lambda_r \mathbf{V}_r$
3. compute the orthogonal transformation matrix as $\mathbf{U}_r = \mathbf{Q}_r \mathbf{V}'_r$
4. rotate each draw of the discrimination parameter matrix as $\tilde{\mathbf{A}}_r = \mathbf{A}_r \mathbf{U}_r$
5. update the consensus matrix as $\mathbf{A}^* = \frac{1}{R} \sum_{r=1}^R \tilde{\mathbf{A}}_r$ and set
6. repeat steps 1 to 5 until $\|\mathbf{A}^{*(s+1)} - \mathbf{A}^{*(s)}\|_F^2 < \omega$, where ω is set close to 0, $\|\cdot\|_F$ indicates the Frobenius norm and s denotes the iteration.

2 Independent-cluster latent structure simulation

In this Section, we provide full details of the simulation settings and the estimation results for the study described in Section 5 of the paper. For the fitted models, the MCMC algorithms were run for 25000 iterations. Posterior inference was based on the last 20000 draws using every 5-th member of the chain to avoid autocorrelation within the sampled values. In the simulation, we considered 20 items rated on a four point Likert scale. We used the parameters represented in Table 1 and the following latent trait covariance

matrices for the weakly ($\Sigma_{(w)}$) and the strongly ($\Sigma_{(s)}$) correlated latent traits:

$$\Sigma_{(w)} = \begin{pmatrix} 1.000 & & & \\ -0.066 & 1.000 & & \\ 0.009 & -0.386 & 1.000 & \\ -0.338 & 0.244 & -0.193 & 1.000 \end{pmatrix}$$

$$\Sigma_{(s)} = \begin{pmatrix} 1.000 & & & \\ -0.235 & 1.000 & & \\ -0.308 & 0.556 & 1.000 & \\ 0.169 & -0.595 & -0.530 & 1.000 \end{pmatrix}$$

Main considerations on the performances of the SSVS and the CSTR approaches to infer simple factorial structures are discussed in the paper. Here, we provide more details on the estimation of the category thresholds and the correlation parameters.

The threshold parameter estimates for the between-item multidimensional IRT model simulation are provided in Table 2. Both SSVS and CSTR approaches seem to perform well in terms of bias and efficiency. As for the correlation parameters, Table 3 shows how the SSVS-MIRT approach leads to less biased and more efficient results than the CSTR-MIRT procedure for correlated latent traits. Tables 4 and 5 highlight the stability of the average absolute bias and the average mean squared error of the estimators of the correlation matrices for different choices of the hyperparameters of the SSVS prior.

Discrimination parameters				Thresholds		
$\alpha_{k,1}$	$\alpha_{k,2}$	$\alpha_{k,3}$	$\alpha_{k,4}$	$\gamma_{k,1}$	$\gamma_{k,2}$	$\gamma_{k,3}$
1.20				-2.052	-0.517	1.453
0.90				-1.452	0.584	1.373
0.50				-1.906	0.774	1.823
0.80				-2.070	-0.725	1.014
0.70				-1.721	-0.556	1.430
	1.00			-1.218	0.074	1.490
	0.70			-1.669	-0.163	1.373
	0.80			-1.408	0.102	1.441
	0.70			-1.454	-0.384	1.632
	0.90			-1.509	0.687	1.256
		1.50		-1.382	0.715	1.601
		0.60		-1.414	-0.117	1.538
		0.90		-2.012	-0.890	1.779
		0.60		-1.979	0.282	1.782
		0.70		-1.649	-0.655	1.007
			1.30	-1.463	0.786	1.257
			0.80	-1.052	-0.417	1.453
			1.00	-2.352	0.684	1.373
			0.80	-1.906	0.774	1.823
			0.50	-2.070	-0.525	1.314

Table 1: Item parameters for the between-item multidimensional IRT model (*to improve readability, only non-zero discrimination parameters are shown in the table*).

Sample size	Latent trait correlation	SSVS-MIRT			CSTR-MIRT			
		$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	
250	<i>uncorrelated</i>	AAB	0.058	0.014	0.048	0.079	0.020	0.065
		AMSE	0.043	0.014	0.032	0.049	0.015	0.035
	<i>weakly correlated</i>	AAB	0.063	0.016	0.045	0.087	0.024	0.065
		AMSE	0.045	0.014	0.033	0.053	0.016	0.037
	<i>strongly correlated</i>	AAB	0.057	0.019	0.051	0.084	0.027	0.074
		AMSE	0.041	0.017	0.035	0.051	0.017	0.039
500	<i>uncorrelated</i>	AAB	0.034	0.013	0.028	0.043	0.015	0.036
		AMSE	0.021	0.008	0.016	0.021	0.009	0.017
	<i>weakly correlated</i>	AAB	0.025	0.010	0.024	0.034	0.012	0.031
		AMSE	0.018	0.007	0.015	0.020	0.008	0.016
	<i>strongly correlated</i>	AAB	0.031	0.007	0.024	0.044	0.010	0.034
		AMSE	0.020	0.007	0.013	0.022	0.007	0.015
1000	<i>uncorrelated</i>	AAB	0.017	0.006	0.015	0.022	0.007	0.017
		AMSE	0.010	0.003	0.007	0.010	0.003	0.007
	<i>weakly correlated</i>	AAB	0.018	0.005	0.012	0.023	0.006	0.015
		AMSE	0.009	0.004	0.007	0.009	0.004	0.007
	<i>strongly correlated</i>	AAB	0.017	0.007	0.011	0.022	0.008	0.015
		AMSE	0.008	0.003	0.007	0.009	0.004	0.007

Table 2: Average absolute bias (AAB) and average mean squared error (AMSE) for the threshold parameters over the 100 simulated datasets (*SSVS*: $\tau = 0.01$, $g = 100$, $p_{k,m} = 0.5$)

Sample Size	Latent trait correlation	SSVS-MIRT		CSRT-MIRT	
		AAB	AMSE	AAB	AMSE
250	<i>uncorrelated</i>	0.011	0.009	0.007	0.004
	<i>weakly correlated</i>	0.009	0.007	0.068	0.010
	<i>strongly correlated</i>	0.020	0.008	0.154	0.032
500	<i>uncorrelated</i>	0.005	0.004	0.005	0.002
	<i>weakly correlated</i>	0.007	0.004	0.041	0.004
	<i>strongly correlated</i>	0.005	0.003	0.117	0.019
1000	<i>uncorrelated</i>	0.005	0.002	0.005	0.002
	<i>weakly correlated</i>	0.005	0.002	0.035	0.003
	<i>strongly correlated</i>	0.003	0.002	0.096	0.014

Table 3: Average absolute bias (AAB) and average mean squared error (AMSE) for the correlation parameters over the 100 simulated datasets ($\tau = 0.01$, $g = 100$, $p_{k,m} = 0.5$).

Latent trait correlation	$\tau = 0.001, g = 1000$		$\tau = 0.1, g = 10$	
	AAB	AMSE	AAB	AMSE
<i>uncorrelated</i>	0.007	0.004	0.006	0.004
<i>weakly correlated</i>	0.007	0.004	0.009	0.004
<i>strongly correlated</i>	0.005	0.003	0.003	0.003

Table 4: Average absolute bias (AAB) and average mean squared error (AMSE) for the correlation parameters over the 100 simulated datasets ($N = 500$) varying τ and g with $p_{k,m} = 0.50$.

Latent trait correlation	$p_{k,m} = 0.25$		$p_{k,m} = 0.75$	
	AAB	AMSE	AAB	AMSE
<i>uncorrelated</i>	0.015	0.019	0.006	0.004
<i>weakly correlated</i>	0.006	0.007	0.009	0.004
<i>strongly correlated</i>	0.015	0.009	0.005	0.004

Table 5: Average absolute bias (AAB) and average mean squared error (AMSE) for the correlation parameters over the 100 simulated datasets ($N = 500$) varying $p_{k,m}$ with $\tau = 0.01, g = 100$.

3 Further simulation studies: Between and Within-item multidimensional IRT models

In this Section, to further evaluate the ability of the proposed methods to investigate the underlying structure of a test, we perform additional simulation studies. In particular, we define several factorial structures by varying the dimensions of the discrimination parameter matrix and by considering four different levels of sparsity. The items are always measured on a four point rating scale. The structure corresponding to the first level of sparsity reflects the possibility that items included in the test fail to measure any latent constructs. For the second level of sparsity, we assume an independent cluster solution. In the remaining two settings, two within-item multidimensional IRT models (Adams et al., 1997), allowing for cross-loadings, are considered.

3.1 Simulation settings

In the first simulation study, we consider 12 items and 2 latent traits. We used the parameters represented in Tables 6 and 9 with uncorrelated, weakly correlated ($\sigma_{1,2} = 0.40$), and strongly correlated ($\sigma_{1,2} = 0.70$) latent traits.

Discrimination parameters					
Sparsity 58.3%		Sparsity 50.0%		Sparsity 37.5%	
$\alpha_{k,1}$	$\alpha_{k,2}$	$\alpha_{k,1}$	$\alpha_{k,2}$	$\alpha_{k,1}$	$\alpha_{k,2}$
1.20		1.20		1.20	
0.90		0.90		0.90	
0.70		0.70		0.70	0.40
0.80		0.80		0.80	
1.10		1.10		1.10	-0.35
		0.60		0.60	0.35
		0.90		0.90	
	1.00		1.00		1.00
	1.10		1.10	0.40	1.10
	0.80		0.80	0.80	
	0.70		0.70	0.70	-0.30
	0.80		0.80	0.80	0.70
					0.80

Table 6: Discrimination parameter matrices for 12 items and 2 latent traits, corresponding to 58.3%, 50.0%, 37.5% and 29.2% sparsity levels, respectively (to improve readability, only non-zero discrimination parameters are presented).

The second simulation setting follows the same specifications detailed in

Section 2. For a sparsity level equal to 75% the loading matrix in given in Table 1, while for the other levels of sparsity the discrimination parameter matrices are represented in Table 7. For both the first and the second simulation settings, the sample size is fixed to $N = 500$.

Discrimination parameters											
Sparsity 77.5%				Sparsity 65.0%				Sparsity 48.8%			
$\alpha_{k,1}$	$\alpha_{k,2}$	$\alpha_{k,3}$	$\alpha_{k,4}$	$\alpha_{k,1}$	$\alpha_{k,2}$	$\alpha_{k,3}$	$\alpha_{k,4}$	$\alpha_{k,1}$	$\alpha_{k,2}$	$\alpha_{k,3}$	$\alpha_{k,4}$
1.20				1.20			-0.50	1.20			-0.50
0.90				0.90				0.90		0.60	
0.50				0.50	-0.60	0.40		0.50	-0.60	0.40	
0.80				0.80				0.80			-0.40
				0.70				0.70			
1.00					1.00				1.00		0.35
0.70				0.50	0.70	0.50		0.50	0.70	0.50	
0.80					0.80				0.80		0.45
0.70					0.70			0.40	0.70		
0.90					0.90	0.70			0.90		0.70
				1.50		1.50			-0.45	1.50	
				0.60	-0.50	0.60		-0.50		0.60	
				0.90		0.90				0.90	0.70
				0.60		0.60			0.50	0.60	
				0.70		0.70		0.35		0.70	
				1.30			1.30				1.30
				0.80			0.80	0.60	-0.35		0.80
				1.00			1.00			0.40	1.00
				0.80			0.80			0.50	0.80
							0.50	0.35			0.50

Table 7: Discrimination parameter matrices for 20 items and 4 latent traits, corresponding to 77.5%, 65.0% and 48.8% sparsity levels, respectively (*to improve readability, only non-zero discrimination parameters are presented*).

The third simulation study investigates the factorial structure of a test comprising 42 items related to 6 latent traits, and a sample size equal to 1000. The simulation parameters for the discrimination coefficients and the thresholds are described in Tables 8 and 9, respectively. The covariance matrices for weakly ($\Sigma_{(w)}$) and strongly ($\Sigma_{(s)}$) correlated latent traits are set as follows:

$$\Sigma_{(w)} = \begin{pmatrix} 1.000 & & & & & \\ 0.104 & 1.000 & & & & \\ 0.269 & 0.281 & 1.000 & & & \\ 0.208 & 0.024 & 0.201 & 1.000 & & \\ 0.187 & 0.042 & 0.148 & 0.324 & 1.000 & \\ -0.338 & 0.125 & 0.032 & -0.148 & 0.039 & 1.000 \end{pmatrix}$$

$$\Sigma_{(s)} = \begin{pmatrix} 1.000 & & & & & \\ -0.365 & 1.000 & & & & \\ 0.124 & 0.072 & 1.000 & & & \\ -0.134 & 0.124 & 0.548 & 1.000 & & \\ 0.280 & -0.340 & -0.411 & -0.788 & 1.000 & \\ -0.261 & 0.149 & 0.132 & -0.229 & 0.119 & 1.000 \end{pmatrix}$$

Discrimination parameters																	
Sparsity 85.7%						Sparsity 83.3%			Sparsity 72.6%			Sparsity 61.9%					
$\alpha_{k,1}$	$\alpha_{k,2}$	$\alpha_{k,3}$	$\alpha_{k,4}$	$\alpha_{k,5}$	$\alpha_{k,6}$	$\alpha_{k,1}$	$\alpha_{k,2}$	$\alpha_{k,3}$	$\alpha_{k,4}$	$\alpha_{k,5}$	$\alpha_{k,6}$	$\alpha_{k,1}$	$\alpha_{k,2}$	$\alpha_{k,3}$	$\alpha_{k,4}$	$\alpha_{k,5}$	$\alpha_{k,6}$
1.00						1.00				0.35		1.00				0.35	-0.40
0.60						0.60				0.40		0.60				0.40	
0.60						0.60				0.40		0.60				0.40	
0.40						0.40				0.40		0.40				0.40	
0.70						0.70				0.70		0.70				0.40	
0.90						0.90				0.90		0.90				0.50	-0.35
0.80						0.80				0.80		0.80				0.35	
1.00						1.00				1.00		1.00				0.40	0.70
1.00						1.00				1.00		0.60				1.00	0.40
0.70						0.70				0.70		0.70				0.70	-0.40
0.60						0.60				0.60		0.60				0.60	0.35
0.60						0.60				0.60		0.40				0.60	
1.00						1.00				1.00		0.40				1.00	0.70
1.00						1.00				1.00		0.60				1.00	0.60
1.20						1.20				1.20		0.70				1.20	-0.40
0.70						0.70				0.70		0.70				0.70	0.40
0.90						0.90				0.90		0.90				0.90	0.60
0.90						0.90				0.90		0.90				0.90	0.60
0.70						0.70				0.70		0.70				0.70	-0.35
0.50						0.50				-0.50		0.50				-0.50	0.50
0.80						0.80				0.80		0.80				0.80	-0.35
1.00						0.40				0.40		0.40				0.40	
0.60						1.00				1.00		0.50				-0.35	1.00
0.50						0.60				0.60		0.60				0.60	0.35
0.70						0.50				0.50		0.50				0.50	0.35
0.80						0.70				0.70		0.70				0.70	-0.35
1.10						0.80				-0.50		0.80				-0.50	0.80
1.10						1.10				0.50		1.10				1.10	0.60
1.10						1.10				1.10		1.10				1.10	-0.80
0.80						1.10				1.10		0.60				1.10	0.80
0.50						0.80				0.80		0.50				0.50	-0.35
1.00						0.50				0.50		1.00				1.00	
1.00						1.00				1.00		0.35				0.50	1.00
1.10						0.40				-0.40		0.40				-0.40	0.40
1.10						1.10				0.35		1.10				1.10	0.50
0.70						0.70				0.70		0.70				0.50	0.70
1.00						1.00				0.60		1.00				0.35	1.00
0.70						0.70				-0.35		0.70				-0.35	0.70
0.80						0.80				0.60		0.80				0.60	0.80
0.80						0.80				0.80		0.80				0.35	0.80
1.10						1.10				0.35		1.10				0.35	1.10
0.70						0.70				0.40		0.40				0.40	0.70

Table 8: Discrimination parameter matrices for 42 items and 6 latent traits, corresponding to 85.7%, 83.3%, 72.6% and 61.9% sparsity levels, respectively (*to improve readability, only non-zero discrimination parameters are presented*).

			Threshold parameters					
$K = 12$ and $M = 2$			$K = 20$ and $M = 4$			$K = 42$ and $M = 6$		
$\gamma_{k,1}$	$\gamma_{k,2}$	$\gamma_{k,3}$	$\gamma_{k,1}$	$\gamma_{k,2}$	$\gamma_{k,3}$	$\gamma_{k,1}$	$\gamma_{k,2}$	$\gamma_{k,3}$
-2.052	-0.517	1.453	-2.052	-0.517	1.453	-1.051	0.246	1.306
-1.452	0.584	1.373	-1.452	0.584	1.373	-1.194	0.761	1.495
-1.382	0.715	1.601	-1.906	0.774	1.823	-1.208	0.002	1.089
-1.414	-0.117	1.538	-2.070	-0.725	1.014	-2.410	-0.178	1.205
-2.012	-0.890	1.779	-1.721	-0.556	1.430	-1.861	-0.855	2.174
-1.979	0.282	1.782	-1.218	0.074	1.490	-1.160	-0.006	1.618
-1.649	-0.655	1.007	-1.669	-0.163	1.373	-1.238	-0.262	2.086
-1.463	0.786	1.257	-1.408	0.102	1.441	-1.549	-0.482	1.139
-1.052	-0.417	1.453	-1.454	-0.384	1.632	-1.825	-0.082	2.349
-2.352	0.684	1.373	-1.509	0.687	1.256	-1.312	0.027	1.327
-1.906	0.774	1.823	-1.382	0.715	1.601	-1.205	0.245	2.255
-2.070	-0.525	1.314	-1.414	-0.117	1.538	-1.108	0.034	1.862
			-2.012	-0.890	1.779	-1.064	-0.148	2.275
			-1.979	0.282	1.782	-2.080	-0.556	1.263
			-1.649	-0.655	1.007	-2.357	0.442	1.183
			-1.463	0.786	1.257	-2.200	-0.942	1.405
			-1.052	-0.417	1.453	-1.820	-0.470	1.432
			-2.352	0.684	1.373	-2.116	0.265	1.292
			-1.906	0.774	1.823	-2.027	0.889	1.627
			-2.070	-0.525	1.314	-2.480	0.255	2.472
						-1.752	0.312	1.685
						-2.227	0.776	2.117
						-2.210	0.486	2.182
						-1.899	0.059	2.060
						-1.162	-0.054	2.399
						-1.629	0.142	1.816
						-1.577	-0.282	2.266
						-1.411	-0.221	2.291
						-1.304	0.085	1.351
						-1.895	-0.126	1.980
						-2.138	0.844	1.715
						-2.421	-0.102	1.089
						-1.160	0.214	1.185
						-2.430	0.600	2.335
						-1.531	0.067	1.863
						-2.435	-0.629	1.899
						-1.786	0.709	1.966
						-2.160	0.992	1.674
						-1.741	-0.171	2.307
						-2.354	-0.510	1.122
						-1.591	0.692	1.020
						-1.067	-0.452	1.143

Table 9: Threshold parameters.

3.2 Bias and Mean Squared Error in model parameter estimation

Tables 10 and 11 show the average absolute bias (AAB) and the average mean squared error (AMSE) over all the discrimination parameter posterior estimates while Tables 12 and 13 provide AAB and AMSE for the correlation matrix estimators. The posterior estimates are obtained as the means of the corresponding posterior distributions. Results confirm how both procedures yield comparable results, though the SSVS estimates are less biased and more efficient especially in presence of a stronger correlation and a lower level of sparsity.

Number of items (K) and latent traits (M)	level of sparsity	uncorrelated		weakly correlated		strongly correlated	
		AAB	AMSE	AAB	AMSE	AAB	AMSE
K=12, M=2	0.583	0.008	0.005	0.009	0.005	0.007	0.008
	0.500	0.009	0.006	0.006	0.005	0.008	0.007
	0.375	0.007	0.006	0.010	0.008	0.017	0.017
	0.292	0.012	0.007	0.010	0.009	0.031	0.026
K=20, M=4	0.775	0.007	0.004	0.006	0.004	0.006	0.004
	0.750	0.008	0.004	0.007	0.004	0.008	0.005
	0.650	0.008	0.007	0.007	0.006	0.012	0.009
	0.488	0.015	0.014	0.021	0.016	0.051	0.020
K=42, M=6	0.857	0.002	0.001	0.002	0.001	0.002	0.001
	0.833	0.002	0.001	0.002	0.001	0.003	0.002
	0.726	0.003	0.001	0.003	0.001	0.004	0.002
	0.619	0.009	0.006	0.006	0.003	0.008	0.004

Table 10: SSVS-MIRT: Average absolute bias and average mean squared error for the discrimination parameters over the 100 simulated datasets for different numbers of items and latent traits and different levels of sparsity ($\tau = 0.01$, $g = 100$, $p_{k,m} = 0.5$).

Number of items (K) and latent traits (M)	level of sparsity	uncorrelated		weakly correlated		strongly correlated	
		AAB	AMSE	AAB	AMSE	AAB	AMSE
K=12, M=2	0.583	0.011	0.010	0.019	0.008	0.063	0.019
	0.500	0.010	0.007	0.019	0.008	0.079	0.022
	0.375	0.030	0.017	0.068	0.014	0.091	0.025
	0.292	0.054	0.023	0.053	0.012	0.070	0.019
K=20, M=4	0.775	0.008	0.010	0.016	0.008	0.047	0.013
	0.750	0.010	0.010	0.018	0.007	0.046	0.012
	0.650	0.032	0.020	0.042	0.010	0.077	0.025
	0.488	0.116	0.047	0.092	0.030	0.119	0.033
K=42, M=6	0.857	0.022	0.010	0.013	0.003	0.058	0.018
	0.833	0.015	0.008	0.014	0.004	0.055	0.015
	0.726	0.087	0.031	0.090	0.032	0.121	0.041
	0.619	0.078	0.014	0.082	0.015	0.101	0.026

Table 11: CSTR-MIRT: Average absolute bias and average mean squared error for the discrimination parameters over the 100 simulated datasets for different numbers of items and latent traits and different levels of sparsity ($|\alpha| \geq 0.3$).

Number of items (K) and latent traits (M)	level of sparsity	uncorrelated		weakly correlated		strongly correlated	
		AAB	AMSE	AAB	AMSE	AAB	AMSE
K=12, M=2	0.583	0.007	0.003	0.005	0.003	0.018	0.002
	0.500	0.002	0.003	0.006	0.002	0.011	0.002
	0.375	0.003	0.005	0.001	0.003	0.015	0.006
	0.292	0.019	0.005	0.003	0.005	0.032	0.008
K=20, M=4	0.775	0.003	0.004	0.002	0.003	0.015	0.003
	0.750	0.005	0.004	0.007	0.004	0.005	0.003
	0.650	0.006	0.004	0.002	0.004	0.019	0.006
	0.488	0.017	0.008	0.046	0.017	0.112	0.020
K=42, M=6	0.857	0.001	0.002	0.003	0.002	0.003	0.001
	0.833	0.002	0.002	0.004	0.002	0.001	0.002
	0.726	0.006	0.002	0.002	0.002	0.002	0.001
	0.619	0.002	0.005	0.013	0.004	0.004	0.002

Table 12: SSVS-MIRT: Average absolute bias (AAB) and average mean squared error (AMSE) for the correlation parameters over the 100 simulated datasets for different numbers of items and latent traits and different levels of sparsity ($\tau = 0.01$, $g = 100$, $p_{k,m} = 0.5$).

Number of items (K) and latent traits (M)	level of sparsity	uncorrelated		weakly correlated		strongly correlated	
		AAB	AMSE	AAB	AMSE	AAB	AMSE
K=12, M=2	0.583	0.006	0.002	0.047	0.004	0.134	0.019
	0.500	0.002	0.003	0.050	0.004	0.123	0.016
	0.375	0.076	0.014	0.015	0.002	0.089	0.009
	0.292	0.092	0.012	0.002	0.002	0.097	0.010
K=20, M=4	0.775	0.005	0.003	0.051	0.005	0.122	0.022
	0.750	0.005	0.002	0.041	0.004	0.117	0.019
	0.650	0.050	0.006	0.059	0.009	0.205	0.052
	0.488	0.097	0.016	0.098	0.017	0.286	0.089
K=42, M=6	0.857	0.002	0.001	0.041	0.004	0.118	0.030
	0.833	0.001	0.001	0.042	0.004	0.097	0.020
	0.726	0.062	0.007	0.084	0.011	0.111	0.025
	0.619	0.055	0.006	0.091	0.016	0.132	0.032

Table 13: CSTR-MIRT: Average absolute bias (AAB) and average mean squared error (AMSE) for the correlation parameters over the 100 simulated datasets for different numbers of items and latent traits and different levels of sparsity ($|\alpha| \geq 0.3$).

3.3 Factorial structure retrieval

For each simulation setting, Tables 14 and 15 show the mean and standard deviation for the sparsity level, and the sensitivity, specificity and accuracy indexes, considering the median probability model for the SSVS method ($ppi \geq 0.5$) and a the traditional cut-off value $|\alpha_{k,m}| \leq 0.3$ for the CSTR approach. The accuracy of the estimated sparse structure is high for both approaches. In case of greater sparsity in the factorial structure, the performances of both methods are satisfactory in terms of correct identification of relevant and non relevant coefficients. The specificity index decreases in presence of a larger number of cross-loadings. Figures from 1 to 6 show how the sensitivity, specificity and accuracy indexes vary for different cut-off values. Results confirm how more complex structure in terms of an increasing number of cross-loadings yield lower levels of accuracy. However, in correspondence to the lowest level of sparsity of each simulation study, the accuracy index assumes values around 0.8 for $ppi \geq 0.5$ in the SSVS approach and for $|\alpha| = 0.3$ in the CSTR approach.

	level of sparsity		uncorrelated				weakly correlated				strongly correlated					
			retrieved sparsity level		sensitivity specificity accuracy		retrieved sparsity level		sensitivity specificity accuracy		retrieved sparsity level		sensitivity specificity accuracy			
			mean	std	sensitivity	specificity	accuracy	mean	std	sensitivity	specificity	accuracy	mean	std	sensitivity	specificity
K=12, M=2	0.583	mean	0.572	1.000	0.980	0.988		0.578	1.000	0.990	0.994		0.566	1.000	0.970	0.983
		std	0.020	0.000	0.034	0.020		0.015	0.000	0.025	0.015		0.039	0.000	0.066	0.039
	0.500	mean	0.487	1.000	0.974	0.987		0.493	1.000	0.985	0.993		0.486	1.000	0.972	0.986
		std	0.023	0.000	0.045	0.023		0.016	0.000	0.032	0.016		0.027	0.000	0.053	0.027
	0.375	mean	0.370	1.000	0.887	0.953		0.368	0.999	0.881	0.950		0.380	0.980	0.884	0.940
		std	0.017	0.000	0.042	0.017		0.020	0.010	0.046	0.021		0.042	0.035	0.088	0.042
	0.292	mean	0.280	1.000	0.748	0.905		0.288	0.998	0.764	0.910		0.315	0.964	0.780	0.895
		std	0.028	0.000	0.074	0.028		0.022	0.011	0.057	0.023		0.063	0.042	0.160	0.068
K=20, M=4	0.775	mean	0.762	1.000	0.983	0.987		0.761	1.000	0.982	0.986		0.761	1.000	0.982	0.986
		std	0.014	0.000	0.017	0.014		0.013	0.000	0.017	0.013		0.015	0.000	0.019	0.015
	0.750	mean	0.739	1.000	0.985	0.989		0.737	1.000	0.982	0.987		0.735	1.000	0.980	0.985
		std	0.010	0.000	0.013	0.010		0.012	0.000	0.016	0.012		0.012	0.000	0.017	0.012
	0.650	mean	0.635	0.995	0.922	0.945		0.638	0.998	0.928	0.950		0.634	0.991	0.918	0.941
		std	0.019	0.019	0.034	0.029		0.016	0.012	0.027	0.021		0.019	0.033	0.029	0.026
	0.488	mean	0.469	0.977	0.815	0.886		0.472	0.971	0.817	0.885		0.470	0.951	0.797	0.865
		std	0.024	0.047	0.065	0.054		0.029	0.047	0.072	0.057		0.025	0.042	0.056	0.045
K=42, M=6	0.857	mean	0.847	1.000	0.989	0.990		0.849	1.000	0.991	0.992		0.848	1.000	0.989	0.991
		std	0.006	0.000	0.007	0.006		0.005	0.000	0.006	0.005		0.008	0.000	0.009	0.008
	0.833	mean	0.823	1.000	0.988	0.990		0.825	1.000	0.990	0.992		0.824	0.999	0.988	0.990
		std	0.007	0.000	0.008	0.007		0.006	0.002	0.008	0.006		0.010	0.004	0.012	0.010
	0.726	mean	0.718	1.000	0.947	0.960		0.719	1.000	0.949	0.961		0.719	0.997	0.947	0.959
		std	0.006	0.000	0.008	0.006		0.005	0.000	0.007	0.005		0.006	0.007	0.008	0.007
	0.619	mean	0.604	0.989	0.874	0.910		0.610	0.997	0.888	0.922		0.610	0.994	0.886	0.920
		std	0.025	0.035	0.049	0.043		0.011	0.012	0.020	0.017		0.008	0.014	0.015	0.013

Table 14: SSVS-MIRT: Mean and standard deviation over the 100 simulated datasets of the sparsity level, and the sensitivity, specificity and accuracy indexes. Cut-off values: $ppi \geq 0.5$ ($K = 12, 20, 42; M = 2, 4, 6$; SSVS: $\tau = 0.01$, $g = 100$, $p_{k,m} = 0.5$)

	level of sparsity		uncorrelated				weakly correlated				strongly correlated			
			retrieved sparsity level	sensitivity	specificity	accuracy	retrieved sparsity level	sensitivity	specificity	accuracy	retrieved sparsity level	sensitivity	specificity	accuracy
K=12, M=2	0.583	mean	0.578	1.000	0.990	0.994	0.583	1.000	1.000	1.000	0.569	1.000	0.976	0.986
		std	0.032	0.000	0.055	0.032	0.000	0.000	0.000	0.000	0.026	0.000	0.044	0.026
	0.500	mean	0.500	1.000	1.000	1.000	0.500	1.000	1.000	1.000	0.478	1.000	0.956	0.978
		std	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.031	0.000	0.063	0.031
	0.375	mean	0.422	0.941	0.929	0.936	0.437	0.944	0.970	0.955	0.411	0.974	0.950	0.964
		std	0.054	0.029	0.119	0.051	0.023	0.029	0.046	0.028	0.038	0.035	0.084	0.043
	0.292	mean	0.363	0.917	0.831	0.885	0.356	0.966	0.892	0.938	0.364	0.983	0.942	0.968
		std	0.053	0.053	0.140	0.070	0.035	0.038	0.068	0.035	0.036	0.032	0.084	0.039
K=20, M=4	0.775	mean	0.773	0.991	0.995	0.994	0.775	0.998	1.000	1.000	0.769	0.998	0.992	0.993
		std	0.009	0.041	0.017	0.020	0.003	0.010	0.002	0.002	0.014	0.010	0.019	0.016
	0.750	mean	0.749	0.988	0.994	0.993	0.750	1.000	1.000	1.000	0.748	0.999	0.997	0.997
		std	0.008	0.055	0.021	0.029	0.001	0.005	0.000	0.001	0.006	0.009	0.007	0.006
	0.650	mean	0.652	0.951	0.926	0.934	0.653	0.992	0.947	0.961	0.655	0.939	0.926	0.930
		std	0.032	0.105	0.054	0.062	0.007	0.017	0.007	0.007	0.025	0.075	0.041	0.046
	0.488	mean	0.538	0.838	0.829	0.833	0.540	0.857	0.850	0.853	0.592	0.779	0.881	0.837
		std	0.042	0.081	0.077	0.067	0.030	0.065	0.052	0.050	0.025	0.045	0.041	0.035
K=42, M=6	0.857	mean	0.855	0.952	0.990	0.984	0.856	0.998	0.999	0.999	0.850	0.934	0.981	0.974
		std	0.010	0.075	0.014	0.020	0.004	0.010	0.006	0.007	0.014	0.082	0.010	0.015
	0.833	mean	0.836	0.953	0.993	0.987	0.834	0.991	0.999	0.997	0.826	0.947	0.980	0.975
		std	0.012	0.103	0.012	0.026	0.002	0.044	0.010	0.016	0.010	0.040	0.009	0.010
	0.726	mean	0.745	0.941	0.965	0.959	0.748	0.937	0.966	0.959	0.762	0.857	0.960	0.935
		std	0.007	0.031	0.007	0.011	0.007	0.033	0.008	0.012	0.010	0.036	0.008	0.011
	0.619	mean	0.687	0.851	0.933	0.907	0.690	0.855	0.939	0.913	0.706	0.759	0.918	0.868
		std	0.013	0.047	0.013	0.020	0.011	0.025	0.012	0.012	0.023	0.062	0.015	0.021

Table 15: CSTR-MIRT: Mean and standard deviation over the 100 simulated datasets of the sparsity level, and the sensitivity, specificity and accuracy indexes. Cut-off values: $|\alpha| \geq 0.3$ ($K = 12, 20, 42; M = 2, 4, 6$)

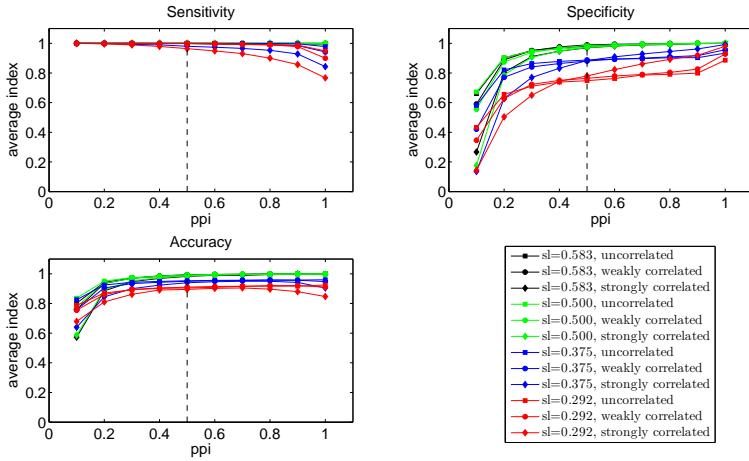


Figure 1: Sensitivity, specificity and accuracy indexes according to different cut-off values for the posterior probability of inclusion for $K = 12, M = 2$ (SSVS-MIRT procedure: $p_{k,m} = 0.5, \tau = 0.01, g = 100$)

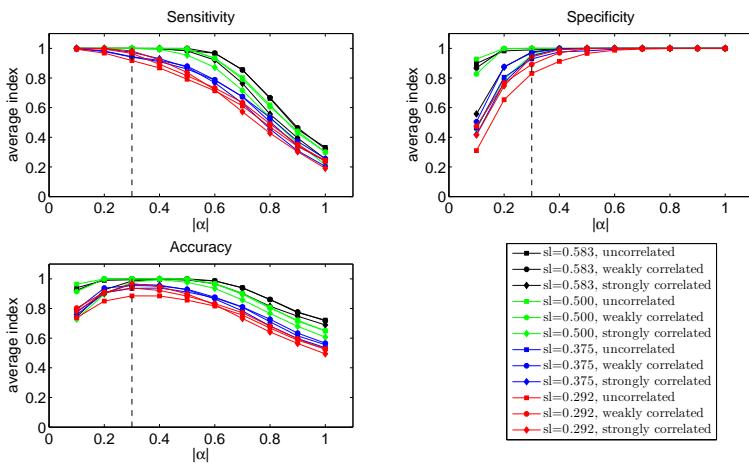


Figure 2: Sensitivity, specificity and accuracy indexes according to different cut-off values for the discrimination parameter absolute value for $K = 12, M = 2$ (CSTR-MIRT procedure)

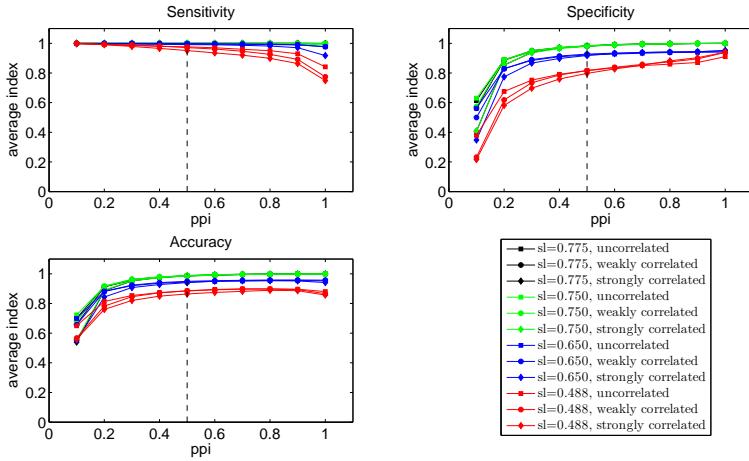


Figure 3: Sensitivity, specificity and accuracy indexes according to different cut-off values for the posterior probability of inclusion for $K = 20, M = 4$ (SSVS-MIRT procedure: $p_{k,m} = 0.5, \tau = 0.01, g = 100$)

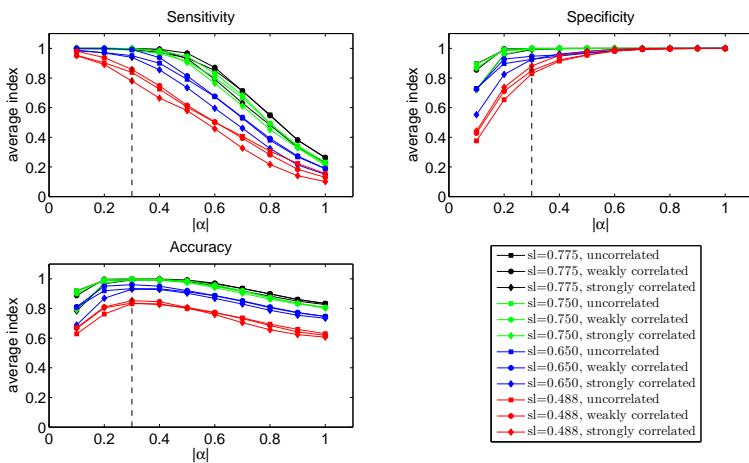


Figure 4: Sensitivity, specificity and accuracy indexes according to different cut-off values for the discrimination parameter absolute value for $K = 20, M = 4$ (CSTR-MIRT procedure)

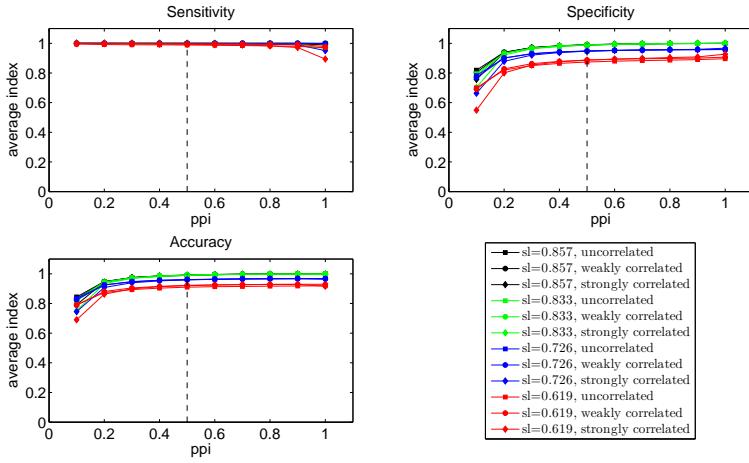


Figure 5: Sensitivity, specificity and accuracy indexes according to different cut-off values for the posterior probability of inclusion for $K = 42$, $M = 6$ (SSVS-MIRT procedure: $p_{k,m} = 0.5, \tau = 0.01, g = 100$)

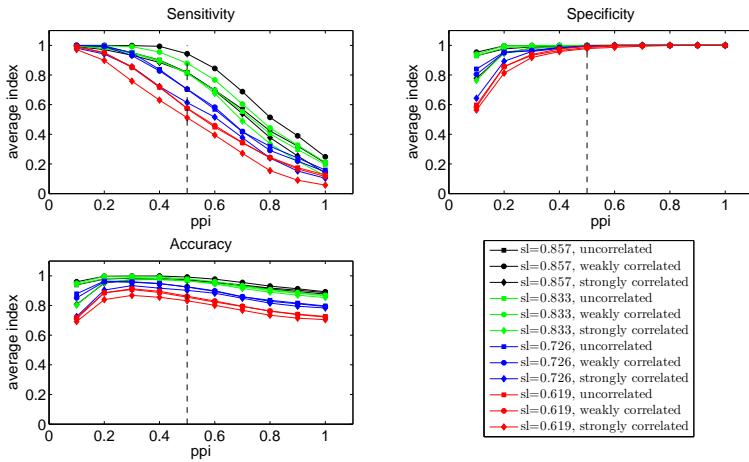


Figure 6: Sensitivity, specificity and accuracy indexes according to different cut-off values for the discrimination parameter absolute value for $K = 42$, $M = 6$ (CSTR-MIRT procedure)

4 Detailed results for the Human Styles Questionnaire

In this Section we provide detailed results for the application on the data responses to the Human Styles Questionnaire (HSQ). For the item formulation see Martin et al. (2003).

As for the estimation results, Tables 16 and 18 contain the estimated discrimination parameters obtained considering the SSVS-MIRT and the CSTR-MIRT approaches. The threshold parameter estimates are shown in Tables 17 and 19.

HSQ items	Affiliative humour	Self-Enhancing humour	Aggressive humour	Self-defeating humour
X1	1.06 (1.00); [0.93;1.20] -0.02	0.00 0.87 (0.03); [-0.11;0.10]	0.00 0.03 0.71 (0.04); [-0.11;0.11]	0.04 -0.04 0.14 (0.04); [-0.07;0.14]
X2	(0.05); [-0.19;0.13] 0.11	(1.00); [0.75;1.01] 0.02	(0.04); [-0.07;0.12] 0.71 (0.04); [-0.07;0.11]	(0.04); [-0.14;0.06] 0.14 (0.04); [-0.14;0.06]
X3	(0.09); [0.00;0.23] 0.00	(0.04); [-0.07;0.13] 0.01	(1.00); [0.6;0.81] 0.00 0.87 (0.04); [-0.10;0.10]	(0.11); [0.04;0.25] 0.06 (0.05); [-0.04;0.15]
X4	(0.04); [-0.12;0.11] 0.89	(0.03); [-0.09;0.11] 0.03	(0.04); [-0.10;0.10] 0.00	(1.00); [0.76;0.98] 0.06 (0.05); [-0.04;0.15]
X5	(1.00); [0.77;1.02] 0.29	(0.04); [-0.07;0.13] 0.55	(0.04); [-0.10;0.10] 0.05	(0.05); [-0.04;0.15] -0.07
X6	(0.62); [0.13;0.44] -0.25	(1.00); [0.42;0.68] 0.00	(0.04); [-0.05;0.14] 0.85	(0.05); [-0.17;0.02] 0.00
X7	(0.45); [-0.41;-0.10] 0.01	(0.04); [-0.11;0.11] -0.01	(1.00); [0.73;0.97] -0.05	(0.04); [-0.11;0.10] 1.47
X8	(0.05); [-0.14;0.16] 0.66	(0.04); [-0.14;0.11] -0.02	(0.05); [-0.18;0.08] 0.01	(1.00); [1.31;1.64] 0.13
X9	(1.00); [0.56;0.77] -0.27	(0.04); [-0.11;0.08] 1.45	(0.04); [-0.08;0.10] -0.01	(0.10); [0.04;0.22] 0.10
X10	(0.49); [-0.60;-0.02] -0.15	(1.00); [1.24;1.72] 0.08	(0.04); [-0.13;0.11] 0.71	(0.08); [-0.03;0.23] -0.05
X11	(0.15); [-0.29;-0.04] 0.11	(0.05); [-0.02;0.18] 0.01	(1.00); [0.61;0.82] -0.09	(0.05); [-0.16;0.04] 0.93
X12	(0.08); [-0.01;0.22] 1.10	(0.03); [-0.09;0.12] 0.08	(0.06); [-0.19;0.01] 0.03	(1.00); [0.82;1.05] -0.05
X13	(1.00); [0.93;1.27] 0.15	(0.06); [-0.04;0.20] 1.08	(0.04); [-0.08;0.15] -0.02	(0.04); [-0.17;0.06] -0.05
X14	(0.22); [-0.06;0.35] -0.01	(1.00); [0.92;1.25] -0.07	(0.04); [-0.12;0.09] 1.03	(0.05); [-0.16;0.05] 0.04
X15	(0.05); [-0.15;0.13] 0.06	(0.06); [-0.20;0.04] -0.05	(1.00); [0.91;1.17] 0.13	(0.04); [-0.08;0.15] 0.77
X16	(0.04); [-0.05;0.17] 1.43	(0.05); [-0.15;0.04] -0.04	(0.09); [0.03;0.22] 0.02	(1.00); [0.67;0.87] 0.04
X17	(1.00); [1.26;1.61] -0.32	(0.05); [-0.17;0.08] 1.43	(0.05); [-0.11;0.15] -0.03	(0.05); [-0.08;0.17] 0.00
X18	(0.61); [-0.64;-0.07] 0.20	(1.00); [1.23;1.67] 0.15	(0.04); [-0.16;0.09] 0.57	(0.04); [-0.13;0.12] 0.11
X19	(0.26); [0.08;0.33] -0.19	(0.13); [0.05;0.26] -0.06	(1.00); [0.47;0.67] 0.11	(0.07); [0.02;0.21] 1.66
X20	(0.29); [-0.42;-0.02] 1.28	(0.06); [-0.21;0.07] 0.05	(0.10); [-0.03;0.26] -0.25	(1.00); [1.45;1.89] 0.12
X21	(1.00); [1.10;1.47] 0.06	(0.05); [-0.08;0.17] 0.44	(0.45); [-0.43;-0.10] 0.08	(0.10); [0.00;0.26] -0.01
X22	(0.05); [-0.05;0.16] 0.06	(0.99); [0.34;0.54] 0.02	(0.05); [0.00;0.16] 0.63	(0.04); [-0.09;0.07] -0.05
X23	(0.05); [-0.04;0.17] -0.24	(0.04); [-0.08;0.12] 0.04	(1.00); [0.53;0.74] -0.02	(0.04); [-0.14;0.04] 0.61
X24	(0.41); [-0.37;-0.13] 1.40	(0.04); [-0.05;0.13] 0.01	(0.03); [-0.11;0.07] 0.00	(1.00); [0.52;0.71] -0.12
X25	(1.00); [1.21;1.61] 0.00	(0.04); [-0.11;0.14] 1.11	(0.04); [-0.13;0.13] -0.02	(0.10); [-0.26;0.01] 0.03
X26	(0.07); [-0.19;0.18] -0.11	(1.00); [0.96;1.28] 0.02	(0.04); [-0.13;0.09] 0.71	(0.04); [-0.08;0.13] 0.03
X27	(0.09); [-0.24;0.00] 0.07	(0.04); [-0.08;0.12] 0.23	(1.00); [0.61;0.83] 0.18	(0.04); [-0.07;0.12] 0.28
X28	(0.05); [-0.03;0.17] 0.81	(0.36); [0.13;0.33] -0.06	(0.17); [0.10;0.27] 0.17	(0.58); [0.19;0.37] -0.22
X29	(1.00); [0.70;0.93] 0.01	(0.04); [-0.16;0.04] 0.60	(0.16); [0.06;0.28] 0.00	(0.33); [-0.34;-0.11] -0.17
X30	(0.04); [-0.12;0.13] 0.06	(1.00); [0.49;0.72] -0.10	(0.04); [-0.09;0.09] 1.09	(0.17); [-0.27;-0.08] -0.11
X31	(0.07); [-0.09;0.21] 0.10	(0.08); [-0.23;0.03] 0.06	(1.00); [0.95;1.24] -0.04	(0.09); [-0.24;0.01] 1.03
X32	(0.07); [-0.03;0.22] 0.07	(0.04); [-0.05;0.16] 0.05	(0.05); [-0.14;0.06] 0.04	(1.00); [0.92;1.16]

Table 16: SSVS-MIRT solution for the HSQ data: discrimination parameter estimates, posterior probabilities of inclusion (*in round brackets*) and 95% credible intervals (*in square brackets*). The discrimination parameters with posterior probabilities of inclusion greater than 0.5 are represented in bold.

HSQ items	γ_1	γ_2	γ_3	γ_4
X1	-2.67 [-2.89;-2.46]	-1.82 [-1.98;-1.68]	-0.92 [-1.04;-0.82]	0.45 [0.36;0.55]
X2	-2.07 [-2.23;-1.91]	-1.03 [-1.13;-0.92]	0.06 [-0.03;0.15]	1.37 [1.25;1.49]
X3	-1.61 [-1.74;-1.48]	-0.62 [-0.71;-0.52]	0.37 [0.28;0.46]	1.54 [1.41;1.67]
X4	-1.43 [-1.56;-1.31]	-0.27 [-0.36;-0.18]	0.70 [0.61;0.80]	1.84 [1.70;1.98]
X5	-2.43 [-2.64;-2.23]	-1.47 [-1.60;-1.34]	-0.30 [-0.39;-0.21]	1.11 [1.01;1.22]
X6	-2.71 [-2.98;-2.48]	-1.87 [-2.02;-1.72]	-1.05 [-1.16;-0.95]	0.14 [0.05;0.22]
X7	-1.48 [-1.62;-1.35]	-0.14 [-0.22;-0.05]	0.97 [0.87;1.08]	1.92 [1.77;2.08]
X8	-1.29 [-1.42;-1.15]	0.21 [0.11;0.31]	1.21 [1.08;1.35]	2.69 [2.46;2.93]
X9	-1.62 [-1.75;-1.49]	-0.84 [-0.94;-0.75]	-0.17 [-0.25;-0.09]	1.08 [0.98;1.19]
X10	-1.76 [-1.94;-1.59]	-0.48 [-0.59;-0.38]	0.76 [0.65;0.87]	2.24 [2.05;2.46]
X11	-1.14 [-1.24;-1.03]	0.00 [-0.09;0.08]	0.59 [0.50;0.68]	1.62 [1.49;1.75]
X12	-1.45 [-1.58;-1.32]	-0.44 [-0.53;-0.35]	0.50 [0.41;0.59]	1.73 [1.59;1.87]
X13	-3.39 [-3.75;-3.06]	-2.66 [-2.91;-2.42]	-1.82 [-1.99;-1.66]	-0.48 [-0.58;-0.37]
X14	-2.00 [-2.17;-1.83]	-0.87 [-0.98;-0.76]	0.12 [0.03;0.21]	1.24 [1.12;1.36]
X15	-0.76 [-0.86;-0.66]	0.03 [-0.06;0.12]	0.80 [0.70;0.91]	1.89 [1.73;2.05]
X16	-1.38 [-1.50;-1.26]	-0.24 [-0.32;-0.15]	0.58 [0.49;0.67]	1.71 [1.58;1.84]
X17	-3.05 [-3.33;-2.79]	-2.13 [-2.33;-1.94]	-1.19 [-1.33;-1.06]	0.20 [0.09;0.31]
X18	-1.62 [-1.77;-1.47]	-0.26 [-0.36;-0.17]	0.91 [0.80;1.03]	2.33 [2.14;2.54]
X19	-1.58 [-1.71;-1.45]	-0.68 [-0.77;-0.58]	0.10 [0.02;0.18]	1.19 [1.08;1.30]
X20	-0.58 [-0.69;-0.46]	1.05 [0.91;1.20]	2.12 [1.91;2.35]	3.56 [3.21;3.97]
X21	-3.72 [-4.10;-3.36]	-2.86 [-3.13;-2.61]	-1.85 [-2.02;-1.68]	-0.33 [-0.43;-0.23]
X22	-1.31 [-1.42;-1.21]	-0.33 [-0.41;-0.25]	0.43 [0.35;0.52]	1.35 [1.24;1.46]
X23	-1.61 [-1.75;-1.49]	-0.65 [-0.74;-0.56]	0.21 [0.13;0.30]	1.23 [1.12;1.34]
X24	-0.92 [-1.02;-0.83]	0.24 [0.15;0.32]	1.10 [1.00;1.21]	1.84 [1.70;1.99]
X25	-3.61 [-3.99;-3.27]	-2.93 [-3.21;-2.66]	-2.14 [-2.36;-1.94]	-0.49 [-0.60;-0.38]
X26	-2.39 [-2.59;-2.19]	-1.40 [-1.53;-1.27]	-0.25 [-0.34;-0.16]	1.11 [1.00;1.23]
X27	-0.43 [-0.51;-0.34]	0.44 [0.36;0.52]	1.00 [0.90;1.10]	1.83 [1.68;1.98]
X28	-1.29 [-1.41;-1.18]	-0.55 [-0.64;-0.46]	0.09 [0.01;0.18]	0.99 [0.89;1.09]
X29	-1.95 [-2.12;-1.81]	-1.23 [-1.34;-1.12]	-0.43 [-0.52;-0.34]	0.74 [0.64;0.84]
X30	-2.29 [-2.47;-2.11]	-1.42 [-1.54;-1.31]	-0.72 [-0.81;-0.63]	0.32 [0.24;0.40]
X31	-1.83 [-2.00;-1.67]	-0.67 [-0.77;-0.57]	0.24 [0.15;0.33]	1.20 [1.08;1.32]
X32	-1.32 [-1.45;-1.20]	-0.34 [-0.43;-0.25]	0.69 [0.60;0.80]	1.91 [1.75;2.06]

Table 17: SSVS-MIRT solution for the HSQ data: threshold parameter estimates and 95% credible intervals (*in square brackets*).

HSQ items	Affiliative humour	Self-Enhancing humour	Aggressive humour	Self-defeating humour
X1	1.02 [0.90;1.15]	0.06 [-0.03;0.14]	0.05 [-0.03;0.13]	0.08 [0.00;0.16]
X2	0.05 [-0.03;0.13]	0.84 [0.73;0.95]	0.04 [-0.04;0.12]	-0.02 [-0.10;0.05]
X3	0.15 [0.07;0.23]	0.03 [-0.05;0.11]	0.69 [0.60;0.80]	0.18 [0.11;0.26]
X4	0.02 [-0.07;0.10]	0.02 [-0.06;0.10]	0.00 [-0.07;0.08]	0.86 [0.76;0.97]
X5	0.85 [0.74;0.96]	0.09 [0.01;0.18]	0.04 [-0.04;0.12]	0.09 [0.02;0.17]
X6	0.34 [0.25;0.44]	0.54 [0.44;0.64]	0.08 [-0.01;0.16]	-0.07 [-0.15;0.01]
X7	-0.24 [-0.32;0.16]	0.00 [-0.08;0.07]	0.84 [0.72;0.95]	0.02 [-0.05;0.09]
X8	0.05 [-0.03;0.13]	0.01 [-0.07;0.09]	-0.05 [-0.12;0.03]	1.46 [1.30;1.64]
X9	0.64 [0.55;0.74]	0.02 [-0.06;0.10]	0.04 [-0.04;0.12]	0.17 [0.09;0.25]
X10	-0.15 [-0.23;-0.07]	1.37 [1.21;1.54]	-0.01 [-0.08;0.07]	0.14 [0.07;0.21]
X11	-0.16 [-0.24;-0.08]	0.10 [0.02;0.18]	0.71 [0.61;0.81]	-0.05 [-0.12;0.03]
X12	0.14 [0.06;0.22]	0.02 [-0.06;0.10]	-0.10 [-0.18;-0.02]	0.93 [0.82;1.04]
X13	1.04 [0.90;1.20]	0.17 [0.07;0.27]	0.09 [0.00;0.19]	-0.04 [-0.12;0.06]
X14	0.25 [0.17;0.33]	1.03 [0.90;1.16]	-0.01 [-0.08;0.07]	-0.03 [-0.10;0.05]
X15	0.02 [-0.06;0.10]	-0.08 [-0.16;0.00]	1.02 [0.89;1.16]	0.08 [0.00;0.15]
X16	0.09 [0.01;0.17]	-0.05 [-0.13;0.03]	0.15 [0.07;0.23]	0.76 [0.67;0.86]
X17	1.38 [1.22;1.55]	0.03 [-0.05;0.12]	0.09 [0.00;0.17]	0.10 [0.02;0.19]
X18	-0.21 [-0.29;-0.13]	1.36 [1.20;1.53]	-0.04 [-0.11;0.04]	0.02 [-0.05;0.09]
X19	0.23 [0.15;0.31]	0.18 [0.10;0.26]	0.57 [0.48;0.67]	0.15 [0.07;0.23]
X20	-0.17 [-0.25;-0.09]	-0.06 [-0.13;0.02]	0.13 [0.05;0.21]	1.63 [1.45;1.85]
X21	1.23 [1.07;1.41]	0.13 [0.04;0.23]	-0.22 [-0.31;-0.13]	0.19 [0.10;0.27]
X22	0.10 [0.02;0.19]	0.41 [0.33;0.50]	0.10 [0.02;0.18]	0.00 [-0.07;0.07]
X23	0.09 [0.01;0.17]	0.03 [-0.06;0.11]	0.62 [0.53;0.72]	-0.04 [-0.11;0.04]
X24	-0.24 [-0.32;-0.15]	0.05 [-0.03;0.14]	-0.03 [-0.11;0.05]	0.60 [0.52;0.70]
X25	1.36 [1.18;1.56]	0.11 [0.01;0.20]	0.07 [-0.03;0.16]	-0.10 [-0.20;-0.02]
X26	0.10 [0.02;0.18]	1.07 [0.94;1.21]	-0.01 [-0.09;0.06]	0.06 [-0.02;0.13]
X27	-0.11 [-0.19;-0.03]	0.03 [-0.05;0.12]	0.70 [0.60;0.81]	0.05 [-0.03;0.12]
X28	0.10 [0.01;0.18]	0.25 [0.16;0.33]	0.20 [0.12;0.29]	0.30 [0.22;0.38]
X29	0.78 [0.67;0.89]	-0.02 [-0.10;0.06]	0.23 [0.15;0.31]	-0.21 [-0.29;-0.13]
X30	0.06 [-0.03;0.14]	0.58 [0.48;0.68]	0.02 [-0.06;0.10]	-0.18 [-0.26;-0.10]
X31	0.10 [0.02;0.17]	0.10 [-0.18;-0.03]	1.09 [0.96;1.23]	-0.09 [-0.16;-0.02]
X32	0.14 [0.06;0.22]	0.07 [-0.01;0.15]	-0.04 [-0.12;0.04]	1.03 [0.92;1.15]

Table 18: CSTR-MIRT solution for the HSQ data: discrimination parameter estimates, and 95% credible intervals (*in square brackets*). The discrimination parameters whose absolute value is greater than 0.3 are represented in bold.

HSQ items	γ_1	γ_2	γ_3	γ_4
X1	-2.66 [-2.87;-2.45]	-1.82 [-1.97;-1.68]	-0.92 [-1.03;-0.82]	0.45 [0.36;0.54]
X2	-2.07 [-2.25;-1.90]	-1.03 [-1.14;-0.92]	0.05 [-0.03;0.14]	1.37 [1.25;1.48]
X3	-1.61 [-1.74;-1.49]	-0.62 [-0.71;-0.53]	0.37 [0.28;0.46]	1.54 [1.42;1.66]
X4	-1.43 [-1.56;-1.31]	-0.27 [-0.36;-0.18]	0.71 [0.61;0.80]	1.84 [1.70;1.99]
X5	-2.44 [-2.64;-2.24]	-1.47 [-1.60;-1.34]	-0.30 [-0.39;-0.21]	1.11 [1.01;1.22]
X6	-2.75 [-3.01;-2.51]	-1.89 [-2.04;-1.74]	-1.06 [-1.17;-0.96]	0.14 [0.05;0.22]
X7	-1.49 [-1.63;-1.36]	-0.14 [-0.22;-0.05]	0.98 [0.88;1.08]	1.93 [1.77;2.09]
X8	-1.29 [-1.44;-1.15]	0.21 [0.11;0.31]	1.22 [1.08;1.36]	2.70 [2.46;2.96]
X9	-1.63 [-1.76;-1.49]	-0.85 [-0.94;-0.75]	-0.17 [-0.26;-0.09]	1.09 [0.99;1.19]
X10	-1.77 [-1.93;-1.61]	-0.48 [-0.58;-0.37]	0.76 [0.66;0.87]	2.25 [2.05;2.45]
X11	-1.14 [-1.26;-1.04]	-0.01 [-0.09;0.08]	0.59 [0.50;0.68]	1.63 [1.50;1.76]
X12	-1.45 [-1.58;-1.33]	-0.44 [-0.53;-0.34]	0.50 [0.41;0.60]	1.73 [1.6;1.88]
X13	-3.39 [-3.72;-3.08]	-2.65 [-2.89;-2.43]	-1.82 [-1.99;-1.66]	-0.47 [-0.57;-0.37]
X14	-1.99 [-2.17;-1.83]	-0.86 [-0.97;-0.76]	0.13 [0.04;0.22]	1.24 [1.12;1.36]
X15	-0.76 [-0.87;-0.66]	0.03 [-0.06;0.12]	0.81 [0.70;0.92]	1.89 [1.73;2.07]
X16	-1.38 [-1.50;-1.26]	-0.24 [-0.33;-0.15]	0.58 [0.49;0.67]	1.71 [1.57;1.85]
X17	-3.06 [-3.33;-2.81]	-2.14 [-2.33;-1.96]	-1.20 [-1.34;-1.06]	0.19 [0.09;0.29]
X18	-1.62 [-1.78;-1.48]	-0.26 [-0.36;-0.16]	0.92 [0.80;1.03]	2.35 [2.13;2.57]
X19	-1.59 [-1.73;-1.46]	-0.68 [-0.78;-0.60]	0.10 [0.02;0.19]	1.20 [1.09;1.31]
X20	-0.58 [-0.69;-0.47]	1.05 [0.92;1.19]	2.13 [1.93;2.33]	3.55 [3.23;3.92]
X21	-3.73 [-4.13;-3.38]	-2.87 [-3.15;-2.62]	-1.86 [-2.04;-1.69]	-0.33 [-0.44;-0.23]
X22	-1.31 [-1.42;-1.20]	-0.33 [-0.41;-0.25]	0.43 [0.35;0.51]	1.35 [1.24;1.46]
X23	-1.61 [-1.74;-1.48]	-0.65 [-0.74;-0.56]	0.21 [0.13;0.30]	1.23 [1.12;1.34]
X24	-0.93 [-1.02;-0.83]	0.24 [0.16;0.33]	1.11 [0.01;1.21]	1.85 [1.71;2.00]
X25	-3.65 [-4.01;-3.30]	-2.96 [-3.25;-2.70]	-2.18 [-2.40;-1.97]	-0.50 [-0.61;-0.39]
X26	-2.39 [-2.60;-2.20]	-1.40 [-1.53;-1.27]	-0.25 [-0.34;-0.16]	1.11 [1.00;1.23]
X27	-0.43 [-0.51;-0.34]	0.45 [0.36;0.53]	1.00 [0.90;1.11]	1.83 [1.68;1.98]
X28	-1.30 [-1.41;-1.19]	-0.55 [-0.64;-0.47]	0.10 [0.02;0.17]	1.00 [0.90;1.09]
X29	-1.96 [-2.13;-1.81]	-1.23 [-1.35;-1.12]	-0.43 [-0.53;-0.34]	0.75 [0.65;0.85]
X30	-2.30 [-2.51;-2.12]	-1.43 [-1.55;-1.32]	-0.72 [-0.81;-0.63]	0.32 [0.24;0.41]
X31	-1.84 [-2.01;-1.69]	-0.68 [-0.78;-0.57]	0.24 [0.15;0.34]	1.20 [1.08;1.33]
X32	-1.32 [-1.46;-1.20]	-0.34 [-0.43;-0.25]	0.70 [0.60;0.80]	1.91 [1.76;2.07]

Table 19: CSTR-MIRT solution for the HSQ data: threshold parameter estimates and 95% credible intervals (*in square brackets*).

The estimated covariances are:

$$\hat{\Sigma} = \begin{pmatrix} 1.00 & & & \\ 0.49 & 1.00 & & \\ [0.34;0.64] & & 1.00 & \\ 0.27 & 0.18 & 1.00 & \\ [0.12;0.42] & [0.03;0.32] & & \\ 0.20 & 0.26 & 0.24 & 1.00 \\ [0.06;0.34] & [0.13;0.38] & [0.10;0.37] & \end{pmatrix}$$

HSQ: Estimated covariance parameters for the 4-factor SVSS-MIRT solution and 95% credible intervals (*in square brackets*)

$$\hat{\Sigma}_C = \begin{pmatrix} 1.00 & & & \\ 0.36 & 1.00 & & \\ [0.31;0.40] & & 1.00 & \\ 0.19 & 0.13 & 1.00 & \\ [0.14;0.24] & [0.07;0.18] & & \\ 0.13 & 0.21 & 0.2 & 1.00 \\ [0.07;0.18] & [0.16;0.26] & [0.14;0.25] & \end{pmatrix}$$

HSQ: Estimated covariance parameters for the 4-factor CSTR-MIRT solution and 95% credible intervals (*in square brackets*) It is worth mentioning that, in general, the credible intervals are shorter for the CSTR-MIRT procedure and that it does not require the tuning of any hyperparameter.

4.1 Model fit

To assess the accuracy of the model, we use the posterior predictive model checking (PPMC) method (Gelman et al., 2013), which, given its simplicity and close relation to classic goodness-of-fit tests (Gelman et al., 1996), is a popular Bayesian checking tool. The PPMC technique depends on the choice of discrepancy measure. Various discrepancy measures have been proposed in the IRT literature, but they are mainly targeted on models for dichotomous responses. The PPMC was extended to polytomous IRT models, and in particular to the graded response model, by Zhu & Stone (2011, 2012). One simple measure at test level is the “observed test score distribution”, representing the number of subjects with each total test score. The overall model fit can be examined through comparing the observed and posterior predictive test score distributions. The credible interval for the posterior predictive score distribution across multiple predicted response data sets and observed score distribution can be shown in a same graph. If the observed score distribution falls within the credible interval, there is no evidence of model misfit at the test level. In Figures 7 and 8 we show the observed total

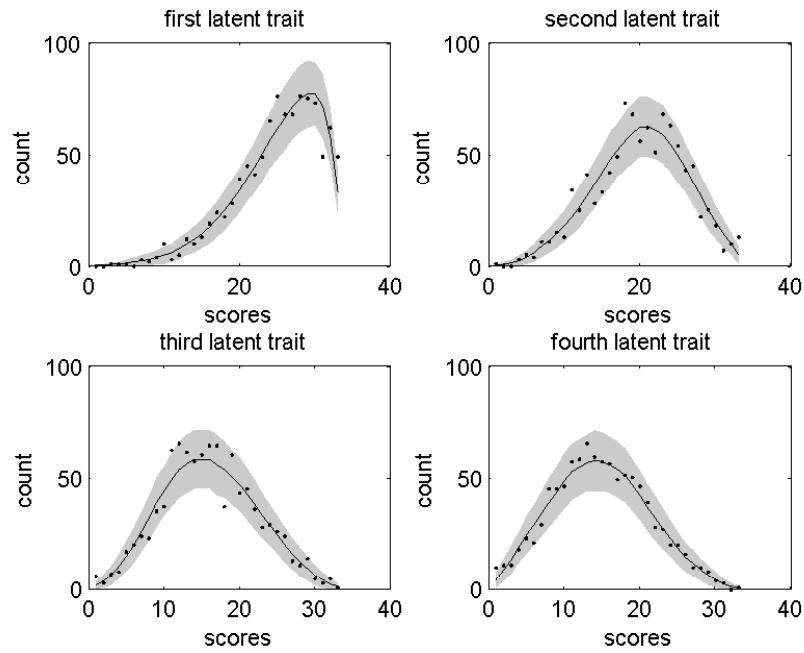


Figure 7: Observed total scores (\cdot), mean (*solid line*) and 95% interval of the posterior predictive test score distributions for the SSVS-MIRT procedure

scores for each dimension, and the mean and 95% interval derived from the posterior predictive test score distributions. It can be highlighted how the model fit with respect to the dimension total scores is adequate for both approaches.

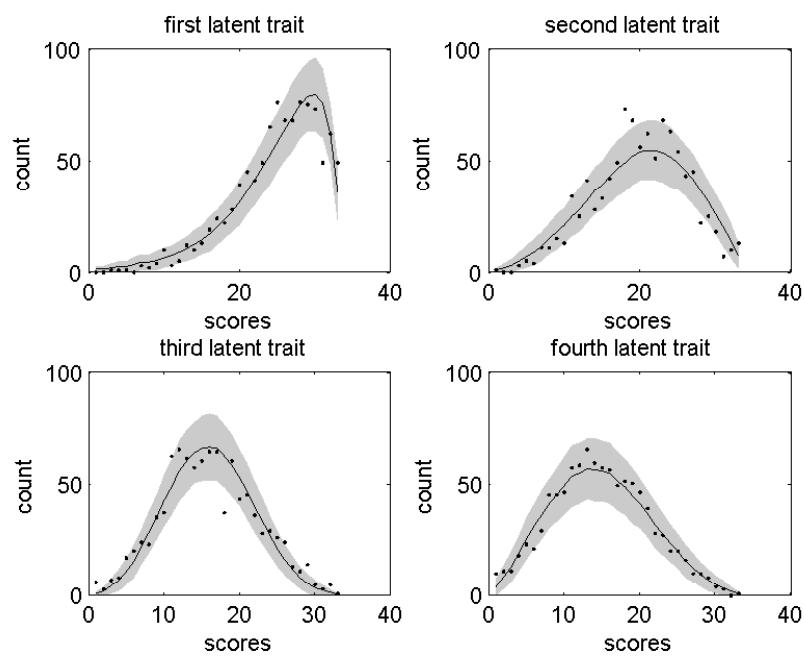


Figure 8: Observed total scores (\cdot), mean (*solid line*) and 95% interval of the posterior predictive test score distributions for the CSTR-MIRT procedure

References

- Adams, R. J., Wilson, M., & Wang, W. (1997). The Multidimensional Random Coefficients Multinomial Logit Model. *Applied Psychological Measurement*, 21(1), 1–23. doi: 10.1177/0146621697211001
- Cowles, M. (1996). Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing*, 6(2), 101–111. doi: 10.1007/BF00162520
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis, Third Edition*. Chapman and Hall/CRC .
- Gelman, A., Meng, X. L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6, 733–807. doi: 10.1.1.142.9951
- George, E. I., & McCulloch, R. E. (1993). Variable Selection Via Gibbs Sampling. *Journal of the American Statistical Association*, 88(423), 881–889. doi: 10.1080/01621459.1993.10476353
- Gower, J. C. (1975). Generalised Procrustes Analysis. *Psychometrika*, 1(40), 33–51. doi: 10.1007/BF02291478
- Liu, X., & Daniels, M. J. (2006). A New Algorithm for Simulating a Correlation Matrix Based on Parameter Expansion and Reparameterization. *Journal of Computational and Graphical Statistics*, 15(4), 897–914. doi: 10.1198/106186006X160681
- Lorenzo-Seva, U., Kiers, H. A. L., & ter Berge, J. M. F. (2002). Techniques for oblique factor rotation of two or more loading matrices to a mixture of simple structure and optimal agreement. *British Journal of Mathematical and Statistical Psychology*, 55, 337–360. doi: 10.1348/000711002760554624
- Martin, R. A., Puhlik-Doris, P., Larsen, G., Gray, J., & Weir, K. (2003). Individual differences in uses of humor and their relation to psychological well-being: Development of the Humor Styles Questionnaire. *Journal of Research in Personality*, 37(1), 48–75. doi: 10.1016/S0092-6566(02)00534-2
- Navarro, D., Vigil-Colet, A., Ferrando, P. J., & Lorenzo-Seva, U. (2016). *Psychological Test Toolbox*. Retrieved from <http://psico.fcep.urv.es/utilitats/PsychologicalTestToolbox/index.html>

Zhu, X., & Stone, C. A. (2011). Assessing fit of unidimensional graded response models using Bayesian methods. *Journal of Educational Measurement*, 48, 81–97. doi: 10.1111/j.1745-3984.2011.00132.x

Zhu, X., & Stone, C. A. (2012). Bayesian comparison of alternative graded response models for performance assessment applications. *Educational and Psychological Measurement*, 72, 774–799. doi: 10.1177/0013164411434638