PSYCHOMETRIKA—VOL. 85, NO. 3, 738–774
SEPTEMBER 2020
https://doi.org/10.1007/s11336-020-09724-3

A PARTIALLY CONFIRMATORY APPROACH TO THE MULTIDIMENSIONAL ITEM RESPONSE THEORY WITH THE BAYESIAN LASSO

JINSONG CHEN

THE UNIVERSITY OF HONG KONG

For test development in the setting of multidimensional item response theory, the exploratory and confirmatory approaches lie on two ends of a continuum in terms of the loading and residual structures. Inspired by the recent development of the Bayesian Lasso (least absolute shrinkage and selection operator), this research proposes a partially confirmatory approach to estimate both structures using Bayesian regression and a covariance Lasso within a unified framework. The Bayesian hierarchical formulation is implemented using Markov chain Monte Carlo estimation, and the shrinkage parameters are estimated simultaneously. The proposed approach with different model variants and constraints was found to be flexible in addressing loading selection and local dependence. Both simulated and real-life data were analyzed to evaluate the performance of the proposed model across different situations.

Key words: MIRT, Bayesian Lasso, partially confirmatory, Lasso loading, local dependence.

It is common to encounter multiple latent traits with categorical data when developing educational and psychological tests. Multidimensional item response theory (MIRT) provides an important analytical modeling framework for this scenario. MIRT models are constructed mainly on the basis of two relationships: (1) the relationship between the items and traits which will be referred to as loading; and (2) the relationship between the items after controlling the traits or the residual covariance matrix with nonzero off-diagonal elements which is often referred to as local dependence (e.g., Embretson & Reise, 2000; Reckase, 2009). Two typical approaches of test development can be differentiated. In the confirmatory approach, both relationships are prespecified based on constraints or substantive knowledge and the main purpose is to confirm the structure of the relationships. In the exploratory approach, both relationships are unspecified without any constraint or substantive knowledge and the purpose is to explore an appropriate structure of the relationships. The exploratory approach is often referred to as item factor analysis when assuming local independence (Bock, Gibbons, & Muraki, 1988; Wirth & Edwards, 2007). The two typical approaches can be connected with exploratory factor analysis (1966) or confirmatory factor analysis (CFA; Jöreskog, 1969), respectively, in the case of continuous data.

In general, however, different amounts of substantive input can be available, making the exploratory and confirmatory approaches two ends of a continuum. Recent development of regularization methods enables more flexibility to address the two relationships in both approaches. On the exploratory end, loading estimation is usually conceptualized as a regularized variable-selection problem with the assumption of local independence under the latent variable modeling context (e.g., Lu, Chow, & Loken, 2016; Sun, Chen, Liu, Ying, & Xin, 2016). Specifically, unspecified loadings with continuous data can be regularized via Bayesian structural equation modeling (BSEM) with the ridge regression (Muthén & Asparouhov, 2012) or spike-and-slab prior (Lu et al., 2016). The least absolute shrinkage and selection operator, or Lasso, regression based on the L_1-norm penalty (Tibshirani, 1996) has been extended to loading selection for dichotomous data under the MIRT setting (Sun et al., 2016). On the confirmatory end, the flexibility comes from

Correspondence should be made to Jinsong Chen, Faculty of Education, The University of Hong Kong, Room 420, 4/F, Meng Wah Complex, Pokfulam Road, Hong Kong, China. Email: jinsong.chen@live.com

the accommodation of local dependence (LD) when the loading structure is fully specified (e.g., Chen, Li, Liu, & Ying, 2018; Epskamp, Rhemtulla, & Borsboom, 2017; Pan, Ip, & Dubé, 2017). Specifically, the Bayesian covariance or graphical Lasso can be used to regularize the residual structure for CFA with continuous data (Pan et al., 2017). The network modeling with the Ising model and Lasso penalization implemented through the maximum likelihood estimation can also be conducted under similar CFA settings (Epskamp et al., 2017). Chen et al (2018) extended a similar network model with regularized pseudolikelihood estimation to dichotomous data under the MIRT setting.

Among different regularization methods, the Bayesian Lasso offers the opportunity to address both loading selection and LD simultaneously. Under the frequentist framework, the Lasso regression implements regularization by adding an L_1-norm penalty term to the usual likelihood function so that the unspecified parameters can be shrunk towards zero (Tibshirani, 1996). With hierarchical priors equivalent to the double-exponential (i.e., Laplace) distributions, a fully Bayesian Lasso regression with hierarchical formulation (Park & Casella, 2008) can be implemented with the Markov chain Monte Carlo (MCMC; Gilks, Richardson, & Spiegelhalter, 1996) algorithm. In the frequentist Lasso, cross-validation is commonly used to find the optimal value of the shrinkage parameter, and the standard errors of the estimates are difficult to obtain. In the Bayesian Lasso approach, no cross-validation is needed by assigning a prior to the Lasso shrinkage parameter, and reliable standard errors or interval estimates are available (Park & Casella, 2008). The Lasso has been extended to address the covariance structure with sparse off-diagonal terms bounded away from zero (Khondker, Zhu, Chu, Lin, & Ibrahim, 2013; Wang, 2012). Specifically, the L_1-norm penalty is applied to the inverse of the covariance matrix with the graphical model. Pan et al. (2017) further extended the Bayesian covariance Lasso to regularize a sparse residual covariance structure for CFA with continuous data.

Inspired by the development of the Bayesian Lasso, this research proposes a partially confirmatory approach to address loading selection and LD within a unified framework under the MIRT setting, which will be called a partially confirmatory item response model (PCIRM). Two variants are also provided for greater flexibility: the PCIRM with local independence (PCIRM-LI) and the confirmatory item response model with local dependence (CIRM-LD), which assumes a known loading structure (i.e., no loading selection). Although they are limited to dichotomous responses in their current forms, the models can be extended to address polytomous items in future research. The proposed approach can cover a wide range of the exploratory–confirmatory continuum and has been found to be flexible for test development with multiple latent traits. The issues of loading selection and LD are addressed with the hierarchical formulation using Bayesian regression and covariance Lasso, and the shrinkage parameters are simultaneously estimated. Based on derivations with different Lasso or regular priors on the loading and residual structures, the MCMC estimation is implemented with the Gibbs sampler (Casella & George, 1992; Geman & Geman, 1984) and Metropolis-Hastings (MH) algorithm (Hastings, 1970; Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953). Addressing both the loading and residual structures at the same time is challenging, and some constraint can be used. The PCIRM with the constraint of one specified loading per item and the CIRM-LD with a known loading structure tend to be confirmatory. In contrast, the PCIRM-LI with the constraint of one loading per trait is closer to the exploratory end. Both simulated and real data sets were analyzed to evaluate the validity, robustness, and practical usefulness of the proposed approach across different settings.

1. The Theoretical Framework

1.1. Model Formulation and Estimation

Consider a test with J items, K latent traits, and N examinees. Let $\mathbf{X} = (x_{ij})_{N \times J}$ denote the data matrix, with x_{ij} as the response of the examinee i to item j. For the present discussion, all item responses are dichotomous, i.e., $x_{ij} \in \{0, 1\}$. Associated with the observable item responses are the continuous latent traits in the K-dimensional Euclidean space, $\boldsymbol{\Theta} = (\theta_{ii})_{N \times K}$ with vector $\boldsymbol{\theta}_i = (\theta_{1i}, \ldots, \theta_{Ki})^{\mathrm{T}}$. One approach of model formulation is to introduce a continuous, latent response vector y_{ij} underlying each dichotomous, observable response vector x_{ij}, with:

$$x_{ij} = \begin{cases} 1 & \text{if } y_{ij} > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The latent response vector for examinee i is $\mathbf{y}_i = (y_{i1}, \ldots, y_{iJ})^{\mathrm{T}}$ with a corresponding latent response matrix of $\mathbf{Y} = (y_{ij})_{N \times J}$. The corresponding measurement model is:

$$\mathbf{y}_i = \boldsymbol{\Lambda}\boldsymbol{\theta}_i + \boldsymbol{\mu} + \boldsymbol{\varepsilon}_i, \tag{2}$$

where model parameters include the $J \times 1$ intercept vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_J)^{\mathrm{T}}$, $J \times K$ loading matrix $\boldsymbol{\Lambda} = (\lambda_{jk})_{J \times K}$, and $J \times 1$ random vector of measurement errors or residuals $\boldsymbol{\varepsilon_i} = (\varepsilon_{i1}, \ldots, \varepsilon_{iJ})^{\mathrm{T}}$, which is independent of $\boldsymbol{\theta}_i$ and follows $N_J[\mathbf{0}, \boldsymbol{\Psi}]$ with residual covariance matrix $\boldsymbol{\Psi} = (\psi_{jj'})_{J \times J}$. In addition, the latent traits are assumed to follow multivariate normal distributions, $\boldsymbol{\theta}_i \sim N_K[\mathbf{0}, \boldsymbol{\Phi}]$, with $\boldsymbol{\Phi} = (\phi_{kk'})_{K \times K}$ as a correlation matrix for scale determinacy. Alternatively, one can estimate a covariance matrix while fixing one item per trait to determine the trait scale.

Let $\boldsymbol{\lambda}_j = (\lambda_{j1}, \ldots, \lambda_{jK})^{\mathrm{T}}$, with parameter vector $\boldsymbol{\omega}$ containing all unknown model parameters in $\boldsymbol{\Lambda}$, $\boldsymbol{\mu}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$. Given the assumption of local independence, the residual covariance matrix $\boldsymbol{\Psi}$ is diagonal, leading to the following conditional distributions:

$$y_{ij} | \boldsymbol{\theta}_i, \boldsymbol{\omega} \sim N(\boldsymbol{\lambda}_j^{\mathrm{T}}\boldsymbol{\theta}_i + \mu_j, \psi_{jj}) \tag{3}$$

or the following item response function:

$$P(x_{ij} = 1 | \boldsymbol{\theta}_i, \boldsymbol{\omega}) = P(y_{ij} > 0 | \boldsymbol{\theta}_i, \boldsymbol{\omega}) = \Phi^*[(\boldsymbol{\lambda}_j^T \boldsymbol{\theta}_i + \mu_j)/\psi_{jj}^{1/2}] \tag{4}$$

where Φ^* is the cumulative function of the standard normal distribution. This is the normal ogive model under the compensatory MIRT setting (Bock et al., 1988). Alternatively, the logistic function can be adopted, resulting in the multidimensional two-parameter logistic model (Reckase, 1985).

However, the above model is unidentified before determining the scale of the latent response vector for item j, $\mathbf{Y}_j = (y_{1j}, \ldots, y_{Nj})^{\mathrm{T}}$. One solution is to set $\mathrm{Var}(\mathbf{Y}_j) = 1$, resulting in the categorical CFA parameterization with slightly better performance in many situations (Forero & Maydeu-Olivares, 2009; Muthén & Asparouhov, 2002). Another solution is to set $\psi_{jj} = 1$, resulting in the common MIRT parameterization, as:

$$P(x_{ij} = 1 | \boldsymbol{\theta}_i) = \Phi^*(\mathbf{a}_j^{\mathrm{T}}\boldsymbol{\theta}_i + d_j) \tag{5}$$

where $\mathbf{a}_j = (a_{j1}, \ldots, a_{jK})^{\mathrm{T}}$ and d_j are the discrimination and location parameters, respectively. It is straightforward that the two solutions are transformable with:

$$a_{jk} = \lambda_{jk}/\psi_{jj}^{1/2} \quad \text{and} \quad d_j = \mu_j/\psi_{jj}^{1/2} \tag{6}$$

Loadings in $\mathbf{\Lambda}$ can be fixed as zero (based on substantive knowledge), specified as free to estimate (also based on substantive knowledge), or unspecified (learned through regularization). The unspecified loading is also called Lasso loading in this research. In a partially confirmatory approach, a different amount of substantive input can be transformed into the structure of $\mathbf{\Lambda}$ with a design matrix $\mathbf{Q} = (q_{jk})_{J\times K}$, where $q_{jk} = 1$, 0, and -1 for specified, zero, and unspecified loadings, respectively. The framework becomes essentially exploratory when all loadings are unspecified or all elements in $\mathbf{Q}$ are -1, and confirmatory when all loadings are either zero or specified. More generally, developers can designate some loadings as zero or specified while keeping others as unspecified in the framework.

The Bayesian estimation combines the idea of data augmentation (Tanner & Wong, 1987) with the MCMC resampling technique. Specifically, the observed data $\mathbf{X}$ are augmented with the latent responses $\mathbf{Y}$ and latent traits $\mathbf{\Theta}$, both of which can be considered as hypothetical missing data, to form a complete data set $\mathbf{Z} = (\mathbf{X}\ \mathbf{Y}, \mathbf{\Theta})$. With appropriate prior distributions, the conditional distribution for individual parameters in the unknown parameter vector $\boldsymbol{\omega}$ based on the complete data set is relatively easy to decompose and analyze. The log of the full conditional distribution for $\boldsymbol{\lambda}_j$ for instance, can be derived with:

$$\log p(\boldsymbol{\lambda}_j|\mathbf{Z}) \propto \log p(\mathbf{Z}|\boldsymbol{\omega}) + \log p(\boldsymbol{\lambda}_j) \tag{7}$$

where $p(\mathbf{Z}|\boldsymbol{\omega})$ is the likelihood function. Full conditional distributions for $\mathbf{Y}$, $p(\mathbf{Y}|\mathbf{X}, \mathbf{\Theta}, \boldsymbol{\omega})$, and $\mathbf{\Theta}$, $p(\mathbf{\Theta}|\mathbf{X}, \mathbf{Y}, \boldsymbol{\omega})$, can be similarly analyzed and obtained. Using the MCMC technique, one can iteratively sample from the conditional distributions to construct the posterior densities and obtain the posterior mean and variance for inference.

However, two thorny issues need to be addressed before the framework can be useful: the simultaneous estimation of both specified and unspecified loadings in $\mathbf{\Lambda}$, and the accommodation of LD with nonzero off-diagonal elements in $\mathbf{\Psi}$. Moreover, the issues are often morphed into identifying significant unspecified loadings and LD. We will turn to the Bayesian Lasso for both.

1.2. The Bayesian Lasso Approach

With LD, the conditional distribution in Eq. (3) or the item response function in Eq. (4) needs to be adjusted. Denote $\mathbf{Y}_{-j}$ as the submatrix of $\mathbf{Y}$ with the jth column deleted, $\mathbf{\Lambda}_{-j}$ as the submatrix of $\mathbf{\Lambda}$ with the jth row deleted, and $\boldsymbol{\mu}_{-j}$ as the subvector of $\boldsymbol{\mu}$ with the jth element deleted. Also denote $\boldsymbol{\psi}_j = (\psi_{j1}, \ldots, \psi_{j,j-1}, \psi_{j,j+1}, \ldots, \psi_{jJ})^{\mathrm{T}}$ as the vector of all off-diagonal elements of the jth column in $\mathbf{\Psi}$, and $\mathbf{\Psi}_{-jj}$ as the $(J-1)\times(J-1)$ matrix resulting from deleting the jth row and jth column from $\mathbf{\Psi}$. The conditional distribution in Eq. (3) can be rewritten as:

$$\mathbf{Y}_j|\mathbf{Y}_{-j}, \mathbf{\Theta}, \boldsymbol{\omega} \sim N[\mathbf{\Theta}\boldsymbol{\lambda}_j + \mu_j + \boldsymbol{\psi}_j^{\mathrm{T}}\mathbf{\Psi}_{-jj}^{-1}(\mathbf{Y}_{-j} - \boldsymbol{\mu}_{-j} - \mathbf{\Theta}\mathbf{\Lambda}_{-j}^{\mathrm{T}}), \psi_{jj}^*] \tag{8}$$

where $\psi_{jj}^* = \psi_{jj} - \boldsymbol{\psi}_j^T\mathbf{\Psi}_{-jj}^{-1}\boldsymbol{\psi}_j$ (Eaton, 1983, pp. 116–117). It can be shown that $\mathbf{Y}_j^* = (y_{1j}^*, \ldots, y_{Nj}^*)^{\mathrm{T}} = \mathbf{Y}_j - \boldsymbol{\psi}_j^{\mathrm{T}}\mathbf{\Psi}_{-jj}^{-1}\mathbf{Y}_{-j}$ and $\mathbf{Y}_{-j}$ are locally independent (Eaton, 1983, p. 88). Moreover, one has:

$$\mathbf{Y}_j^*|\mathbf{\Theta}, \boldsymbol{\omega} \sim N(\mathbf{\Theta}\boldsymbol{\lambda}_j + \boldsymbol{\mu}_j^*, \psi_{jj}^*) \tag{9}$$

where $\boldsymbol{\mu}_j^* = \mu_j - \boldsymbol{\psi}_j^{\mathrm{T}} \boldsymbol{\Psi}_{-jj}^{-1} (\boldsymbol{\mu}_{-j} + \boldsymbol{\Theta}\boldsymbol{\Lambda}_{-j}^{\mathrm{T}})$.

For a specific item, one can rearrange the latent traits and loading vector in Eq. (2) so that the zero or specified loadings and unspecified loadings can be partitioned. Denote the $K \times 1$ loading vector for item j as $\boldsymbol{\lambda}_j = \begin{pmatrix} \boldsymbol{\lambda}'_j \\ \boldsymbol{\lambda}''_j \end{pmatrix}$, where the $K' \times 1$ vector $\boldsymbol{\lambda}'_j = (\lambda'_{j1}, \ldots, \lambda'_{jK'})^{\mathrm{T}}$ and $K'' \times 1$ row vector $\boldsymbol{\lambda}''_j = (\lambda''_{j1}, \ldots, \lambda''_{jK''})^{\mathrm{T}}$ are the unspecified and zero/specified parts, respectively. With the complete data, Eq. (2) becomes a common regression model, and the unspecified loadings can be shrunk with the following objective function:

$$PL(\boldsymbol{\lambda}'_j) = \log p(\mathbf{Z}|\boldsymbol{\omega}) + \delta_j \sum\nolimits_{k=1}^{K'} |\lambda'_{jk}| \tag{10}$$

where $PL(\boldsymbol{\lambda}'_j)$ is the penalized log-likelihood and $\delta_j \geq 0$ is the tuning parameter that determines the amounts of shrinkage (Tibshirani, 1996). By comparing Eqs. (7) and (10), it is straightforward that the Lasso approach is connected with the Bayesian framework if the penalty term in Eq. (10) can be cast into the log prior term in Eq. (7). As Tibshirani (1996) suggested, the connection is established when the unspecified parameters can be assigned independent and identical double-exponential priors, as $\frac{\delta_j}{2} \exp(-\delta_j |\lambda'_{jk}|)$.

Similarly, the LD or off-diagonal elements in the residual covariance matrix $\boldsymbol{\Psi}$ can be penalized using the graphical or covariance Lasso (Yuan & Lin, 2007). In general, one works on the inverse of the covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Psi}^{-1} = (\sigma_{jj'})_{J \times J}$, as:

$$PL(\boldsymbol{\Sigma}) = \log(\boldsymbol{\Sigma}|\text{rest}) + \delta_s \|\boldsymbol{\Sigma}\|_1 \tag{11}$$

where $\|\boldsymbol{\Sigma}\|_1$ is the L_1 norm of Σ and δ_s is the shrinkage parameter. More details will be given below.

Motivated by the fact that the double-exponential distribution can be expressed as a scale mixture of normal distributions, Park and Casella (2008) developed a fully hierarchical model and an efficient Gibbs sampler for the Bayesian Lasso, as:

$$\begin{aligned} \boldsymbol{\lambda}'_j &\sim N_{K'}(\mathbf{0}, \mathbf{D}_{\tau_j}), \\ \mathbf{D}_{\tau_j} &= \mathrm{diag}(\tau_{j1}^2, \ldots, \tau_{jK'}^2), \\ \tau_{jk}^2 &\sim Gamma\left(1, \frac{\delta_j^2}{2}\right) \end{aligned} \tag{12}$$

where $k = 1, 2, \ldots, K'$ and δ_j is the shrinkage parameter on item j. The explicit form of the full conditional distribution can be derived as shown in "Appendix A". A hyperprior can be assigned to δ_j as $\delta_j^2 \sim Gamma(\alpha_j, \beta_j)$, and the common choices for hyperparameters are $\alpha_j = 1$ and β_j to be small. For specified loadings, conjugate prior distribution $\boldsymbol{\lambda}''_j \sim N_{K''}(\boldsymbol{\lambda}_{0j}, \mathbf{H}_{0j})$ can be assigned with λ_{0j} and $\mathbf{H}_{0j}$ as hyperparameters, which can also lead to explicit full conditional distribution (Appendix A). A conjugate prior for the intercept vector can be assigned as $\boldsymbol{\mu} \sim N_j(\boldsymbol{\mu}_0, \mathbf{H}_{\mu 0})$ with $\boldsymbol{\mu}_0$ and $\mathbf{H}_{\mu 0}$ as hyperparameters; the explicit full conditional distribution for sampling (Lee, 2007) is also given in "Appendix A".

For accommodation of LD, the entire $\boldsymbol{\Sigma} = \boldsymbol{\Psi}^{-1}$ can be modeled as a sparse structure on off-diagonal elements with the Bayesian graphical or covariance Lasso (Khondker et al., 2013; Wang, 2012). Double exponential priors can be assigned to the off-diagonal elements, which has the form $\frac{\delta_s}{2} \exp(-\delta_s |\sigma_{jj'}|)$ with $j < j'$, whereas independent exponential priors can be assigned to the diagonal elements with the form $\frac{\delta_s}{2}\exp(-\frac{\delta_s}{2}\sigma_{jj})$ (Wang, 2012). Hyperpriors can be assigned to the shrinkage parameter δ_s as $\delta_s \sim Gamma(\alpha_s, \beta_s)$, and the common choices for hyperparameters are $\alpha_s = 1$ and β_s to be small. The full conditional distribution can be obtained and directly sampled from using block Gibbs sampler. The related distributions and procedures are presented in "Appendix B".

For estimation of the correlation matrix $\mathbf{\Phi}$, one can first directly sample from the original covariance matrix $\mathbf{\Phi}^*$ with conjugate prior $\mathbf{\Phi}^* \sim$ Inv-Wishart$(\mathbf{S}_0^{-1}, v_0)$. It is suggested that the hyperparameters of the inverse-Wishart distribution are set as $v_0 = K \quad + 2$ with the diagonal and off-diagonal elements of $\mathbf{S}$ as 1 s and 0.5 s, respectively. The sampled $\mathbf{\Phi}^*$ is then transformed into corresponding $\mathbf{\Phi}$ by adapting the relevant MCMC algorithms (Liu, 2008; Liu & Daniels, 2006). More details are provided in the section below.

For model identification, it is difficult to obtain necessary and sufficient conditions under the Bayesian Lasso context; the main interest is to find a reasonable and convenient way to solve the problem. Specifically, it is found that parameter estimates will be unstable or difficult to converge if there is no specified loading for many items. The reason might be that the L_1-penalties on the loadings and off-diagonal residual elements can interfere across items and shift the estimates in different directions. Two sets of constraints will be evaluated for usefulness under the PCIRM. In Constraint I, there is a specified loading per item while all other loadings can be unspecified. The constraint is meaningful substantively since there is usually a target trait for each item during test development. The specified loading can be the major loading under the BSEM context (Muthén & Asparouhov, 2012) or any other loading that the developer is confident to specify In Constraint II, there is a specified loading for about half of the items, while all other loadings can be unspecified. Constraint II is more relaxed than Constraint I.

Two variants of the PCIRM with different constraints can be useful across different situations. First, a diagonal residual covariance matrix is assumed in the PCIRM with local independence (PCIRM-LI). This is similar to the regularized MIRT model with the frequentist Lasso proposed by Sun et al. (2016) and can be identified with the same constraint of one loading per trait. However, PCIRM-LI enjoys the benefit of the Bayesian Lasso approach (i.e., the availability of the standard error and no need for cross-validation) and simultaneous estimation of the trait correlations. Compared to PCIRM, PCIRM-LI sacrifices the regularization of the residual covariance matrix for a more relaxed constraint when regularizing the loading matrix. The prior and full conditional distributions of PCIRM-LI are the same as those in PCIRM except for $\mathbf{\Psi}$. For the diagonal elements in $\mathbf{\Psi}$, the priors are $\psi_{jj}^{-1} \sim Gamma(\alpha_{0j}, \beta_{0j})$ and the common choices for hyperparameters are $\alpha_{0j} = 1$ and β_{0j} to be small. The full conditional distributions can be found in "Appendix A".

Second, when the loading structure is known (specified or fixed as zero), it will be called the confirmatory item response model with local dependence (CIRM-LD). It is similar to Pan et al. (2017)'s Bayesian covariance Lasso CFA model, but with a dichotomous response instead of continuous data. Accordingly, both models can be similarly identified under regular CFA or MIRT conditions. The prior and full conditional distributions of CIRM-LD are the same as those in PCIRM without regularization of the loading matrix.

1.3. Markov Chain Monte Carlo Estimation

A graphical representation of the model structure of different parameters is presented in Fig. 1. The figure shows how the item responses x_{ij} can be modeled using latent traits, loadings, and residuals, which are characterized by their mean, covariance structure, and shrinkage parameters in a hierarchical formulation. The Bayesian hierarchical structure can be estimated using MCMC, which is a simulation-based algorithm that iteratively resamples from the probability distributions based on a stochastic process of Markov chains (Gill, 2002). Under mild regularity conditions, the Markov chains will converge to stationary posterior distribution after a sufficiently large number

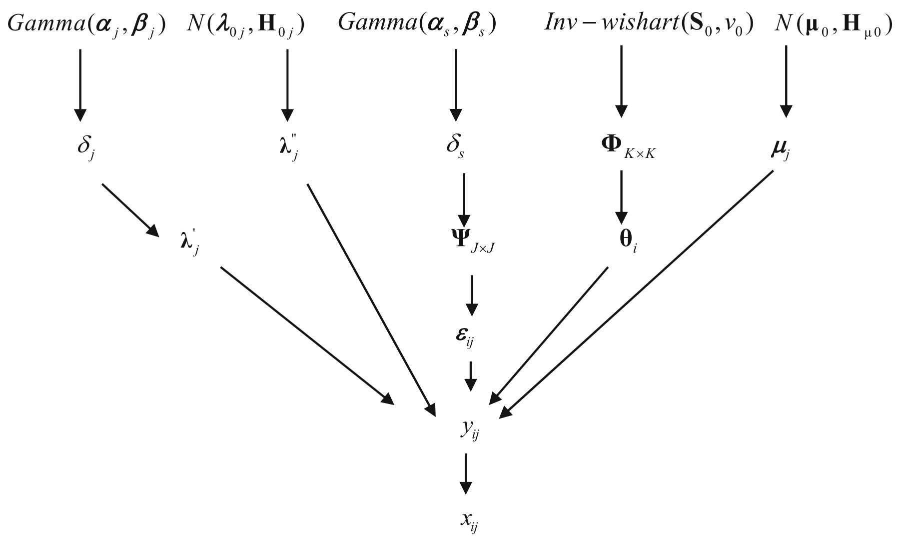

FIGURE 1.
A directed acyclic graph for the PCIRM

of iterations called the burn-in period. Two types of MCMC will be used: the Gibbs sampler (Casella & George, 1992) which is easier to implement but requires the explicit full conditional distributions, and the MH algorithm (Chib & Greenberg, 1995) which relies on the joint posterior distribution with proposal distribution. Rather than directly sample from the posterior, the MH algorithm samples candidates from the proposal distribution first and then accepts or rejects the candidates based on comparisons with the target distribution. This research adopted a mix of the Gibbs sampler and MH algorithm, or more specifically, a MH-algorithm-within-Gibbs sampler.

The procedure for sampling the parameters of interest from their full conditional distributions with $\mathbf{Z} = (\mathbf{X}, \mathbf{Y}, \boldsymbol{\Theta})$, has seven steps:

1. Draw θ_i from $p(\boldsymbol{\theta}_i|\mathbf{X}, \mathbf{Y}, \boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Psi})$ for $i = 1$ to N
2. For PCIRM and PCIRM-LI, draw unspecified loadings $\boldsymbol{\lambda}'_j$ from $p(\boldsymbol{\lambda}'_j \big| \mathbf{Z}, \boldsymbol{\Lambda}_{-j}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \delta_j)$ for $j = 1$ to J
3. Draw specified loadings $\boldsymbol{\lambda}''_j$ from $p(\boldsymbol{\lambda}''_j \big| \mathbf{Z}, \boldsymbol{\Lambda}_{-j}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Psi})$ for $j = 1$ to J
4. Draw $\boldsymbol{\mu}$ from $p(\boldsymbol{\mu}| \mathbf{Z}, \boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi})$
5. First draw covariance matrix $\boldsymbol{\Phi}^*$ from Inv-Wishart$(\mathbf{S}_N^{-1}, \upsilon_N)$ (i.e., Gibbs sampling), where $\upsilon_N = \upsilon_0 + N$ and $\mathbf{S}_N = \mathbf{S}_0 + \sum_{i=1}^{N} \theta_i \theta'_i$. Then obtain a provisional $\boldsymbol{\Phi}$ as $\boldsymbol{\Phi} = \mathbf{D}^{-1}\boldsymbol{\Phi}^*\mathbf{D}^{-1}$, where $\mathbf{D} = \mathrm{diag}(\phi^*_{11}, \ldots, \phi^*_{KK})^{1/2}$. Finally $\boldsymbol{\Phi}$ is accepted with a MH probability of $\min\left((\frac{|\boldsymbol{\Phi}^{(t)}|}{|\boldsymbol{\Phi}^{(t-1)}|})^{(K+1)/2}, 1\right)$, where t is the current draw and $(t-1)$ is the previous draw.
6. For PCIRM and CIRM-LD, draw $\boldsymbol{\Sigma}$ from $p(\boldsymbol{\Sigma}| \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Phi}, \delta_s)$ with an efficient block Gibbs sampler and compute $\boldsymbol{\Psi} = \boldsymbol{\Sigma}^{-1}$ based on "Appendix B"; for PCIRM-LI, draw ψ_{jj} from $p(\psi_{jj} \big| \mathbf{Z}, \boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\Phi})$ based on "Appendix A".
7. Draw $\mathbf{Y}^*_j = (y^*_{1j}, \ldots, y^*_{Nj})^T$ with

$$p(y^*_{ij}|\boldsymbol{\Theta}, \boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Psi}) \sim \begin{cases} N[(\boldsymbol{\lambda}_j^T \theta_i + \boldsymbol{\mu}^*_j)/(\psi^*_{jj})^{1/2}, 1] I_{(-\infty,0)}(y^*_{ij}), & \text{if } x_{ij} = 0, \\ N[(\boldsymbol{\lambda}_j^T \theta_i + \boldsymbol{\mu}^*_j)/(\psi^*_{jj})^{1/2}, 1] I_{(0,+\infty)}(y^*_{ij}), & \text{if } x_{ij} = 1, \end{cases}$$

for $i = 1, \ldots, N$ where $I_A(y_{ij}^*)$ is an indicator function that takes the value 1 if y_{ij}^* is in A, and 0 otherwise; then compute $\mathbf{Y}_j = \mathbf{Y}_j^* + \boldsymbol{\psi}_j^{\mathrm{T}} \boldsymbol{\Psi}_{-jj}^{-1} \mathbf{Y}_{-j}$ and standardize Var($\mathbf{Y}_j$) as 1 for $j = 1$ to J

Multiple chains with different initial values can be run to monitor the convergence of the algorithm. After the burn-in period, the convergence for the parameters of interest can be determined using the estimated potential scale reduction (EPSR) value (Gelman, 1996). The EPSR value compares the ratio of the weighted average of the within-chain variance and between-chain variance to the within-chain variance. The chains are said to converge to the stationary distribution if this ratio is less than 1.1 (Gelman, Carlin, Stern, & Rubin, 2004). The uncertainty of estimates will be characterized with the concept of highest posterior density (HPD) intervals, or more specifically the $100(1-\alpha)\%$ HPD interval (Box & Tiao, 1973). Both the EPSR value and HPD interval can be obtained with the R package coda (Plummer, Best, Cowles, & Vines, 2006). All programming was conducted on the R platform (R Development Core Team, 2010).

Table 1 summarizes different parameters' initial values and prior distributions with related hyperparameters and prior values $\mathbf{Y}$ and $\boldsymbol{\Theta}$ were randomly initialized with $MVN(\mathbf{0}, \mathbf{I})$ where $\mathbf{I}$ is the identity matrix. To determine the burn-in period, two Markov chains were initialized and run until EPSR < 1.1. In a preliminary study, two sets of initial values and prior values had been evaluated for sensitivity analysis as shown in Table 1, and the differences turned out to be trivial. For subsequent analyses, the first set of prior values was used, which was less informative. The fit of the model can be assessed via the posterior predictive (PP) p value (Gelman, Meng, & Stern, 1996; Meng, 1994). Given that the Bayesian Lasso was already used as a model selection tool, the PP p value was only used as a complementary statistic. A model is considered plausible if the PP p value estimate is not far from 0.5, which all proposed models in this research satisfied.

2. Empirical Studies

2.1. Study 1: Performance of PCIRM Under Constraints I and II

Study 1 evaluated the performance of the proposed model under Constraints I and II with three traits and 24 items, namely, $J = 24$ and $K = 3$, through a simulation study. The trait correlations were set as $\phi_{kk'} = .5$ for k and $k' = 1$ to 3 and $k \neq k'$. $\mu_j = -.5$ and .5 for odd and even numbers of j, respectively. The true loading matrix was

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_{11}\ \lambda_{21}\ \lambda_{31}\ \lambda_{41}\ \lambda_{51}\ \lambda_{61}\ \lambda_{71}\ \lambda_{81}\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_{17,1}\ \lambda_{18,1}\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_0 \\ \lambda_{12}\ \lambda_{22}\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_{92}\ \lambda_{10,2}\ \lambda_{11,2}\ \lambda_{12,2}\ \lambda_{13,2}\ \lambda_{14,2}\ \lambda_{15,2}\ \lambda_{16,2}\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_0 \\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_{93}\ \lambda_{10,3}\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_0\ \lambda_{17,3}\ \lambda_{18,3}\ \lambda_{19,3}\ \lambda_{20,3}\ \lambda_{21,3}\ \lambda_{22,3}\ \lambda_{23,3}\ \lambda_{24,3} \end{bmatrix}^{\mathrm{T}}$$

where some loadings were set as .7 (λ_{31} to λ_{81}, $\lambda_{11,2}$ to $\lambda_{16,2}$, and $\lambda_{19,3}$ to $\lambda_{24,3}$), some as .5 (λ_{11}, λ_{21}, $\lambda_{17,1}$, $\lambda_{18,1}$, λ_{12}, λ_{22}, $\lambda_{9,2}$, $\lambda_{10,2}$, λ_{93}, $\lambda_{10,3}$, $\lambda_{17,3}$, and $\lambda_{18,3}$), and all others as $\lambda_0 = 0$. For the residual structure in $\boldsymbol{\Psi}$, the diagonal elements were $\psi_{jj} = .51$ for $j = 3$ to 8, 11 to 16, and 19 to 24 and $\psi_{jj} = .25$ for $j = 1, 2, 9, 10, 17$, and 18. The nonzero off-diagonal elements for the lower triangle were $\psi_{43} = \psi_{11,12} = \psi_{19,20} = \psi_{5,13} = \psi_{14,11} = \psi_{6,22} = .4$ with a symmetric upper triangle. The first three terms represent LD within one trait, whereas the last three were between trait. All other off-diagonal elements were set as $\psi_0 = 0$. This was a balanced design where each latent trait was measured by 10 items, four with cross-loadings, two with within-trait LD, and two with between-trait LD. Two levels of sample sizes were simulated ($N = 250$ and 500), and the number of replications was 200 for each sample size.

TABLE 1.
Parameters' initial values, prior distributions, hyperparameters, and prior values

Parameter	$\boldsymbol{\lambda}'_j$		$\boldsymbol{\Psi}^{-1}$		$\boldsymbol{\mu}$		$\boldsymbol{\lambda}''_j$		$\boldsymbol{\Phi}$		ψ_{jj} (PCIRM-LI)	
IV	1		$5\mathbf{I}$		$\mathbf{0}$		.3		$\mathbf{I}+.1_{\mathrm{od}}$		2	
PD	DE(δ_j)		DE(δ_s)Exp(δ_s)		Normal		Normal		Inv-Wishart		Inv-Gamma	
HP	α_j	β_j	α_s	β_s	$\boldsymbol{\mu}_0$	$\mathbf{H}_{\boldsymbol{\mu}0}$	λ_{0j}	$\mathbf{H}_{0j}$	$\boldsymbol{S}_0$	$\boldsymbol{v}_0$	α_{0j}	β_{0j}
PV 1	1	.01	1	.01	$\mathbf{0}$	$4\mathbf{I}$	0	$4\mathbf{I}$	$\mathbf{I}+.1_{\mathrm{od}}$	K+2	1	.01
PV 2	1	.1	1	.1	$\mathbf{0}$	$\mathbf{I}$	0	$\mathbf{I}$	$\mathbf{I}+.5_{\mathrm{od}}$	K+8	1	.1

Note. IV, initial value; PD, prior distribution; HP, hyperparameter; PV, prior value; DE, double exponential; Exp, exponential; $\mathbf{I}$ = identity matrix; $\mathbf{I}+.1_{\mathrm{od}}$, diagonal elements as 1 and off-diagonal elements as .1; $\mathbf{I}+.5_{\mathrm{od}}$, diagonal elements as 1 and off-diagonal elements as .5.

For estimation, the design matrix under Constraint I was:

$$\mathbf{Q} = \begin{bmatrix} 1&1&1&1&1&1&1&1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1 \\ -1&-1&-1&-1&-1&-1&-1&-1&1&1&1&1&1&1&1&1&-1&-1&-1&-1&-1&-1&-1&-1 \\ -1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&1&1&1&1&1&1&1&1 \end{bmatrix}^{\mathrm{T}}$$

which implied 8 specified loadings and 16 unspecified or Lasso loadings (2 nonzero and 14 zero, i.e., λ_0) per trait. The design matrix under Constraint II was

$$\mathbf{Q} = \begin{bmatrix} 1&1&1&1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1 \\ -1&-1&-1&-1&-1&-1&-1&-1&1&1&1&1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1 \\ -1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&1&1&1&1&-1&-1&-1&-1 \end{bmatrix}^{\mathrm{T}}$$

which implied 4 specified loadings and 20 unspecified loadings (6 nonzero and 14 zero, i.e., λ_0) per trait. For performance assessment over replications, the mean bias of the parameter estimates (BIAS), the mean of the standard error estimates (SE), the root mean squares error (RMSE) between the estimates and the true values, and the percentage of estimates significantly different from zero at $\alpha = .05$ (SIG%) based on the HPD interval were computed. Through a preliminary study, it was found that all chains had reached stationary status (i.e., EPSR < 1.1) within 10,000 iterations, and the burn-in was set at 20,000 iterations. After the burn-in phase, parameters were estimated based on an additional 40,000 iterations

Table 2 presents simulation results under Constraint I. Due to the symmetry of the design, the estimates of loadings, intercepts, and residual covariance were similar across different traits and are only shown for the first trait to save space. When $N = 500$, parameter recovery was satisfactory in general except for the off-diagonal residuals. Specifically, most biases of estimates were trivial, but those for specified loadings associated with between-trait LD and small diagonal residuals were slightly larger. Moreover, the nonzero unspecified loadings were slightly underestimated, while the nonzero off-diagonal residuals were highly underestimated. The RMSE and SE showed that the estimates were stable either across replications or within one replication. (Note that the relatively large RMSEs for the off-diagonal residuals were mainly caused by the corresponding biases.) The SIG% showed that the Type I error rates for zero Lasso loadings and off-diagonal residuals were close to zero, while the statistical power for most nonzero parameters was close to one. The only exception was the power for the within-trait off-diagonal residuals, which was moderate ($\sim .7$). When $N = 250$, recoveries of most parameters were only slightly worse, with the exception of statistical power for the nonzero off-diagonal residuals, which dropped largely.

Table 3 presents simulation results under Constraint II. When $N = 500$, most estimates tended to be worse than, but not far away from, those in Constraint I under similar conditions, as measured by BIAS, RMSE, SE, or SIG%. But statistical power for the within-trait off-diagonal residuals was substantially smaller. When $N = 250$, the performance of most estimates was similarly worse, and the power for significant off-diagonal residuals was rather low. In brief, the more relaxed Constraint II was less preferred, and the only concern under Constraint I was the underestimation of LD, which can be alleviated if the main purpose is to detect significant effects and the sample size is large enough.

2.2. Study 2: Model Comparisons Under Local Independence

Study 2 compared the performance of different proposed and existing models with different constraints under the condition of local independence. The models for comparison covered the PCIRM with the constraint of one specified loading per item (Constraint I), the PCIRM-LI with the constraint of one specified loading per trait (Constraint III), and the BSEMs for dichotomous

TABLE 2.
Parameter estimates under constraint I in study 1

Par	TRUE	$N = 500$				$N = 250$			
		BIAS	RMSE	SE	SIG%	BIAS	RMSE	SE	SIG%
λ_{11}	.5	.003	.064	.094	1.000	.013	.091	.127	.990
λ_{21}	.5	.007	.063	.094	1.000	.007	.090	.128	.990
λ_{31}	.7	.040	.075	.107	1.000	.026	.100	.134	.990
λ_{41}	.7	.043	.078	.106	1.000	.027	.100	.133	.990
λ_{51}	.7	− .075	.099	.104	1.000	− .086	.124	.135	.990
λ_{61}	.7	− .082	.103	.105	1.000	− .096	.135	.137	.985
λ_{71}	.7	− .028	.072	.104	1.000	− .043	.102	.134	.990
λ_{81}	.7	− .032	.071	.103	1.000	− .048	.107	.134	.985
$\lambda_{17,1}$	.5	− .086	.106	.098	1.000	− .136	.163	.132	.825
$\lambda_{18,1}$	.5	− .078	.103	.098	1.000	− .124	.155	.132	.885
λ_0	.0	.012	.057	.092	.001	.016	.064	.113	.004
μ_{p}	.5	− .003	.058	.058	1.000	.015	.096	.086	.986
μ_{n}	− .5	− .002	.056	.058	1.000	− .015	.099	.086	.986
$\phi_{kk'}$	.5	.016	.049	.110	1.000	− .005	.076	.134	.983
ψ_{jj}	.3	.082	.098	.055	1.000	.114	.124	.081	1.000
ψ_{ii}	.5	.030	.074	.082	1.000	.028	.084	.114	1.000
ψ_w	.4	− .269	.277	.063	.693	− .276	.291	.078	.122
ψ_b	.4	− .230	.239	.053	.976	− .247	.261	.074	.568
ψ_0	.0	.003	.019	.040	.000	.007	.065	.042	.006

Note. For $\phi_{kk'}$, k and $k' = 1$ to 3 and $k \neq k'$; for ψ_{jj}, $j = 2$ to 4, 8 to12, and 14 to 17; for ψ_{ii}, $i = 1, 6, 7$, 11, 12 and 18; ψ_w averaged across ψ_{43}, $\psi_{16,15}$, and $\psi_{10,9}$ (within trait); ψ_b averaged across ψ_{85}, $\psi_{14,11}$, and $\psi_{17,2}$ (between traits); μ_{p} for mean positive intercepts, μ_{n} for mean negative intercepts.
TRUE, true values; BIAS, bias of the parameter estimates; SE, mean of the standard error estimates; RMSE, root mean squares error between the estimates and the true values; SIG%, percentage of estimates significantly different from zero at $\alpha = .05$.

responses with Constraint I (BSEM-DRt) and Constraint III (BSEM-DRi). The BSEM approach proposed by Muthén and Asparouhov (2012) relies on the ridge regression priors to penalize the unspecified loadings, but its performance for dichotomous responses hasn't been addressed in the existing literature. The generalized latent factor model (Chen, Li, & Zhang, 2019) for dichotomous responses (GLFM-DR) with Constraint IV was also included. The GLFM-DR proposed by Chen et al. (2019) is a categorical CFA using the joint maximum likelihood estimator with a structured Q-matrix to specify the loadings as free to estimate or fixed as zero. In Constraint IV, all but one loading in one item per trait were fixed as zero, while all other loadings were specified as free to estimate. This is similar to, but slightly more restricted than, Constraint III. However, the SEs of estimates are not available in the GLFM-DR, which is a challenging problem under a double asymptotic regime, as both the sample size and test length tend to infinity[1].

The simulation settings were the same as those in Study 1, except for the following changes: λ_{93}, $\lambda_{17,3}$ and $\lambda_{18,3}$ were set as .3; λ_{92}, $\lambda_{17,2}$ and $\lambda_{18,2}$ were set as .7; all off-diagonal elements in $\boldsymbol{\Psi}$ were set as zero; $N = 1000$. For Constraint III, λ_{81}, $\lambda_{16,2}$ and $\lambda_{24,3}$ were set as specified; for Constraint IV, λ_{82}, λ_{83}, $\lambda_{16,1}$, $\lambda_{16,3}$, $\lambda_{24,1}$ and $\lambda_{24,2}$ were fixed as zero. For the PCIRM and PCIRM-LI, all chains were found stationary (i.e., EPSR < 1.1) within 10,000 iterations, which was set as the burn-in. Parameters were estimated based on another 10,000 iterations. Both the

[1] Personal communication with the corresponding author.

TABLE 3.
Parameter estimates under constraint II in study 1

Par	TRUE	N = 500				N = 250			
		BIAS	RMSE	SE	SIG%	BIAS	RMSE	SE	SIG%
λ_{11}	.5	.021	.083	.132	.985	−.017	.106	.177	.855
λ_{21}	.5	.011	.078	.132	.980	−.089	.135	.181	.640
λ_{31}	.7	.069	.091	.137	.990	.004	.111	.202	.910
λ_{41}	.7	.066	.089	.137	.990	−.057	.132	.216	.835
λ_{51}	.7	−.096	.098	.141	.990	−.101	.147	.201	.875
λ_{61}	.7	−.082	.091	.140	.990	−.196	.229	.209	.690
λ_{71}	.7	−.021	.075	.138	.990	−.061	.123	.196	.925
λ_{81}	.7	−.026	.074	.137	.990	−.160	.198	.208	.790
$\lambda_{17,1}$	.5	−.140	.164	.139	.985	−.152	.179	.180	.420
$\lambda_{18,1}$	.5	−.125	.154	.138	.980	−.112	.150	.180	.575
λ_0	.0	.007	.066	.121	.000	.025	.076	.174	.000
μ_{p}	.5	−.003	.057	.058	1.000	−.001	.083	.082	.986
μ_{n}	−.5	−.001	.056	.058	1.000	.001	.086	.082	.986
$\phi_{kk'}$	.5	.031	.073	.145	.990	.027	.123	.185	.918
ψ_{jj}	.3	.086	.106	.064	1.000	.155	.197	.097	1.000
ψ_{ii}	.5	.038	.082	.099	1.000	.076	.123	.131	1.000
ψ_w	.4	−.278	.282	.070	.383	−.279	.287	.093	.078
ψ_b	.4	−.235	.239	.059	.962	−.267	.276	.078	.257
ψ_0	.0	.005	.035	.039	.005	.022	.077	.060	.025

Note. TRUE, true values; BIAS, bias of the parameter estimates; SE, mean of the standard error estimates; RMSE, root mean squares error between the estimates and the true values; SIG%, percentage of estimates significantly different from zero at $\alpha = .05$.

BSEM-DRt and BSEM-DRi were implemented through Mplus (Muthén & Muthén, 1998–2015), while the GLFM-DR was implemented through the R package mirtjml (Zhang, Chen, & Li, 2019).

Simulation results for the PCIRM and PCIRM-LI are summarized in Table 4, with details for nonzero loading estimates in “Appendix C”. The performance of the two models was satisfactory and similar, although generally the PCIRM-LI was slightly better. This is understandable since PCIRM-LI is a simpler model with correct assumptions. In both models, the power for nonzero unspecified loadings could be lower with a smaller loading value, but still good enough even when the value was as low as .3. The Type I error rates for zero loadings in both models or that for off-diagonal elements of $\boldsymbol{\Psi}$ in the PCIRM were either zero or close to zero, implying the conservative nature of the regularization.

Simulation results for the BSEMs and GLFM-DR are summarized in Table 5, with details for nonzero loading estimates in “Appendix C”. All three models performed worse than the PCIRM or PCIRM-LI. In comparison, the biases from the GLFM-DR were acceptable, but the RMSEs for loading estimates were large. In the BSEM-DRi, the loading estimates were stable with small SEs, but could be biased. The Type I error rates for zero loadings were also high ($\sim .16$). Results from the BSEM-DRt were the worst among the three models.

2.3. *Study 3: Model Comparisons Under Local Dependence*

The above models with the same constraints were further compared under the condition of LD. The CIRM-LD with a known loading structure (i.e., all loadings are either specified or fixed as zero correctly) was also included. All simulation settings were the same as those in Study 2, except that the following off-diagonal elements in $\boldsymbol{\Psi}$ were set as .3: ψ_{43}, $\psi_{11,12}$, $\psi_{19,20}$, $\psi_{5,13}$, $\psi_{14,11}$, $\psi_{6,22}$ in the lower triangle, with a symmetric upper triangle.

Simulation results for the PCIRM, PCIRM-LI, and CIRM-LD are summarized in Table 6, with details for nonzero loading estimates in “Appendix C”. The average computational time for each replication of the PCIRM, PCIRM-LI, and CIRM-LD was about 460, 163, and 388 seconds, respectively, on a HP Envy laptop with an Intel Core i7 CPU.

The performance of the three models was satisfactory in general. For the PCIRM-LI, the loading estimates associated with within-trait LD were slightly biased, suggesting the limited impact of the LD. For both PCIRM-LI and PCIRM, the power and Type I error for loading estimates were similar to those in Study 2. For both PCIRM and CIRM-LD, the estimates for LD were similarly and highly biased, suggesting that the biases should be less associated with the regularization of the loadings. Note that the related biases were small for the Bayesian covariance Lasso CFA model proposed by Pan et al. (2017), which is the counterpart of CIRM-LD for continuous data. Accordingly, it seemed that the high biases of LD were more likely associated with the measurement level of the dichotomous data. Compared to Study 1, the magnitude of the biases was about the same (i.e., $\sim$ two-thirds of the true effect), but the power to detect LD increased as the sample size was larger. A reasonable explanation is the smaller SEs of the estimates with a larger sample size. Results for the BSEMs and GLFM-DR (Table 7 and “Appendix C”) were basically the same as those in Study 2 and are not repeated for discussion. The average computational time for each replication of the BSEM and GLFM-DR was about 16 and 3 seconds, respectively, on the same computer described above.

2.4. *Study 4: Practicability of PCIRM to Identify Minor Trait*

Here minor traits refer to traits measured only by minor loadings (three or more). Different from major loadings which measure the traits by design, minor loadings are cross-loadings that load on the traits unintentionally. Minor traits can be easily neglected during the modeling process. But it can offer a simpler interpretation to the local dependence among a cluster of items (i.e.,

TABLE 4.
(Summary) Parameter estimates for PCIRM-LI and PCIRM in study 2

Par	TRUE	PCIRM-LI				PCIRM			
		BIAS	RMSE	SE	SIG%	BIAS	RMSE	SE	SIG%
λ_{s5}	.5	–	–	–	–	.005	.044	.093	1.000
λ_{s7}	.7	.014	.049	.075	1.000	−.006	.044	.096	1.000
λ_{u3}	.3	−.031	.059	.090	.910	−.025	.052	.106	.863
λ_{u5}	.5	.000	.049	.078	1.000	−.026	.053	.097	1.000
λ_{u7}	.7	.007	.047	.074	1.000	–	–	–	–
λ_0	.0	−.006	.044	.081	.001	.008	.046	.106	.000
μ_p	.5	.000	.040	.041	1.000	−.001	.040	.041	1.000
μ_n	−.5	−.002	.042	.041	1.000	−.002	.041	.041	1.000
ψ_0	.0	–	–	–	–	.002	.010	.024	.000

Note TRUE, true values; BIAS, bias of the parameter estimates; SE, mean of the standard error estimates; RMSE, root mean squares error between the estimates and the true values; SIG%, percentage of estimates significantly different from zero at $\alpha = .05$; λ_{s5}, λ_{s7}, mean specified loadings with values of .5, .7; λ_{u3}, λ_{u5}, λ_{u7}, mean unspecified loadings with values of .3, .5, .7; μ_p, mean positive intercepts; μ_n,: mean negative intercepts.

TABLE 5.
(Summary) Parameter estimates for BSEMs and GLFM-DR in study 2

Par	TRUE	BSEM-DRt				BSEM-DRi				GLFM-DR	
		BIAS	RMSE	SE	SIG%	BIAS	RMSE	SE	SIG%	BIAS	RMSE
λ_{s5}	.5	–	–	–	–	.182	.188	.054	1.000	–	–
λ_{s7}	.7	−.531	.531	.073	.928	.054	.070	.054	1.000	.105	.244
λ_{u3}	.3	−.221	.221	.062	.333	−.211	.212	.039	.712	.031	.168
λ_{u5}	.5	−.212	.304	.064	.833	−.296	.297	.045	1.000	.038	.222
λ_{u7}	.7	−.455	.474	.074	.954	–	–	–	–	−.002	.325
λ_0	.0	.157	.158	.070	.862	−.035	.058	.049	.161	.064	.173
μ_p	.5	.025	.050	.042	1.000	.001	.042	.041	1.000	−.048	.081
μ_n	−.5	−.026	.050	.042	1.000	−.005	.041	.041	1.000	.079	.101

Note. TRUE, true values; BIAS, bias of the parameter estimates; SE, mean of the standard error estimates; RMSE, root mean squares error between the estimates and the true values; SIG%, percentage of estimates significantly different from zero at $\alpha = .05$.

TABLE 6.
(Summary) Parameter estimates for PCIRM-LI, PCIRM and CIRM-LD in study 3

Par	T	PCIRM-LI				PCIRM				CIRM-LD			
		B	R	SE	SIG%	B	R	SE	SIG%	B	R	SE	SIG%
λ_{s3}	.3	–	–	–	–	–	–	–	–	.000	.042	.057	1.000
λ_{s5}	.5	–	–	–	–	.004	.044	.095	1.000	−.010	.042	.057	1.000
λ_{s7}	.7	−.008	.071	.074	.995	−.001	.055	.098	1.000	−.003	.034	.046	1.000
λ_{u3}	.3	−.034	.063	.089	.928	−.035	.057	.107	.837	–	–	–	–
λ_{u5}	.5	−.011	.057	.079	.998	−.034	.057	.101	1.000	–	–	–	–
λ_{u7}	.7	.020	.088	.075	.996	–	–	–	–	–	–	–	–
λ_0	.0	−.004	.066	.083	.007	.007	.054	.109	.000	–	–	–	–
μ_p	.5	.001	.040	.041	1.000	−.001	.040	.041	1.000	−.001	.040	.041	1.000
μ_n	−.5	−.002	.042	.041	1.000	−.003	.042	.041	1.000	−.003	.042	.041	1.000
ψ_0	.0	–	–	–	–	.002	.010	.024	.000	.001	.011	.024	.000
ψ_w	.3	–	–	–	–	−.203	.204	.043	.782	−.192	.193	.038	.975
ψ_b	.3	–	–	–	–	−.195	.196	.040	.957	−.185	.187	.036	.998

Note. T, True; B, BIAS; R, RMSE; λ_{s3}, λ_{s5}, λ_{s7}, mean specified loadings with values of .3, .5, .7; λ_{u3}, λ_{u5}, λ_{u7}, mean unspecified loadings with values of .3, .5, .7; μ_p, mean positive intercepts, μ_n, mean negative intercepts; ψ_w averaged across ψ_{43}, $\psi_{16,15}$, and $\psi_{10,9}$ (within trait); ψ_b averaged across ψ_{85}, $\psi_{14,11}$, and $\psi_{17,2}$ (between traits).

TABLE 7.
(Summary) Parameter estimates for BSEMs and GLFM-DR in study 3

Par.	TRUE	BSEM-DRt				BSEM-DRi				GLFM-DR	
		BIAS	RMSE	SE	SIG%	BIAS	RMSE	SE	SIG%	BIAS	RMSE
λ_{s5}	.5	–	–	–	–	.150	.157	.054	1.000	–	–
λ_{s7}	.7	−.532	.532	.073	.928	.059	.079	.053	1.000	.065	.223
λ_{u3}	.3	−.221	.221	.063	.333	−.200	.201	.040	.838	.023	.175
λ_{u5}	.5	−.213	.303	.065	.833	−.278	.280	.045	1.000	.019	.215
λ_{u7}	.7	−.453	.471	.074	.969	–	–	–	–	.007	.333
λ_0	.0	.159	.160	.070	.877	−.033	.059	.048	.165	.065	.183
μ_p	.5	.025	.051	.042	1.000	.003	.042	.041	1.000	−.055	.088
μ_n	−.5	−.026	.050	.042	1.000	.006	.041	.041	1.000	.079	.103

Note. TRUE, true values; BIAS, bias of the parameter estimates; SE, mean of the standard error estimates; RMSE, root mean squares error between the estimates and the true values; SIG%, percentage of estimates significantly different from zero at $\alpha = .05$.

clustered LD). In the following true loading matrix, there were three ordinary traits each with eight major loadings and one minor trait with four minor loadings:

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_{11} & \lambda_{21} & \lambda_{31} & \lambda_{41} & \lambda_{51} & \lambda_{61} & \lambda_{71} & \lambda_{81} & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 \\ \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_{92} & \lambda_{10,2} & \lambda_{11,2} & \lambda_{12,2} & \lambda_{13,2} & \lambda_{14,2} & \lambda_{15,2} & \lambda_{16,2} & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 \\ \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_{17,3} & \lambda_{18,3} & \lambda_{19,3} & \lambda_{20,3} & \lambda_{21,3} & \lambda_{22,3} & \lambda_{23,3} & \lambda_{24,3} \\ \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_{15,4} & \lambda_{16,4} & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_0 & \lambda_{23,4} & \lambda_{24,4} \end{bmatrix}^{\mathrm{T}}$$

where the major loadings (λ_{11} to λ_{81}, λ_{92} to $\lambda_{16,2}$, and $\lambda_{17,3}$ to $\lambda_{24,3}$) were set as .7, the minor loadings ($\lambda_{15,4}$, $\lambda_{16,4}$, $\lambda_{23,4}$, and $\lambda_{24,4}$) as .55, and all others as $\lambda_0 = 0$. Statistically, the four minor loadings were equivalent to clustered LD, and, when ignored, should be transformed as $\psi_{16,15} = \psi_{24,23} = \psi_{24,16} = \psi_{24,15} = \psi_{23,16} = \psi_{23,15} = .3$ with a symmetric upper triangle in $\mathbf{\Psi}$. The correlations between ordinary traits were set as .5, whereas the minor trait was independent with others. For the residual structure in $\mathbf{\Psi}$, the only nonzero off-diagonal elements were $\psi_{78} = \psi_{87} = .3$. The sample sizes were 1,000, and 200 datasets were randomly generated. All other conditions were similar above.

A three-step approach was adopted to assess the performance of the proposed models. First, the PCIRM with all major loadings specified (i.e., $K = 3$) was fitted, which represented the case ignoring the minor trait. Except for the minor loadings and LD, parameter recovery was satisfactory with small bias, stable estimates, and high power or low Type I error (Table 8). As expected, the ignored minor loadings were transformed as clustered LD, which saw underestimation similar to the ordinary LD term ψ_{78}. The statistical power for the clustered LD was lower than that for the ordinary LD, though, which implied that only partial information about the significance of minor loadings would be mostly likely available in practice.

Second, the PCIRM-LI with all major loadings and the first two minor loadings specified (i.e., $K = 4$) was fitted, representing the case where there was partial information about the minor trait. As shown in Table 9, λ_{71} and λ_{81} were somewhat overestimated due to ignoring the LD term ψ_{78}. All other loadings including the minor ones were recovered satisfactorily. Third, the PCIRM with all major and minor loadings specified (i.e., $K = 4$) was fitted, representing the case with full information about the minor trait and accommodation of LD. The LD term was underestimated, but still possessed a high power (Table 9). Other than that, the parameter recovery was satisfactory. It is noteworthy all four minor loadings were accurately recovered in both Steps Two and Three, whereas the corresponding clustered LD was substantially underestimated in Step One. Accordingly, the minor trait can not only provide a simpler interpretation, but also reduce the bias when estimating the residual covariance, and the three-step approach provides a viable means to identify the minor trait among clustered LD.

2.5. *Study 5: Real-Life Example*

The short scale of the revised Eysenck's Personality Questionnaire (EPQ-R; Eysenck & Barrett, 2013; Eysenck, Eysenck, & Barrett, 1985) was used to demonstrate the usefulness of the PCIRM in a real-life setting. The dataset contains responses to 36 dichotomous items from 824 women in the UK.[2] The dataset had been preprocessed so that the negatively worded items had already been reversely scored. There were three traits underlying the items, each with 12 major loadings ("Appendix D"). Accordingly, Constraint I with one specified loading per item can be adopted, while all other loadings can be treated as unspecified and estimated with the Bayesian Lasso approach.

[2] The normative data were not open access due to "publisher sensitivity issues," but was obtained from Dr. Paul Barrett through personal communication.

TABLE 8.
Step one results in study 4

Par	T	BIAS	RMSE	SE	SIG%	Par	T	BIAS	RMSE	SE	SIG%
λ_{11}	.7	−.022	.045	.085	1.000	$\lambda_{17,3}$	.7	−.040	.058	.088	1.000
λ_{21}	.7	−.029	.049	.087	1.000	$\lambda_{18,3}$	.7	−.047	.064	.087	1.000
λ_{31}	.7	−.022	.046	.082	1.000	$\lambda_{19,3}$	.7	−.034	.052	.088	1.000
λ_{41}	.7	−.026	.045	.083	1.000	$\lambda_{20,3}$	.7	−.038	.057	.087	1.000
λ_{51}	.7	−.026	.051	.083	1.000	$\lambda_{21,3}$	.7	−.040	.056	.089	1.000
λ_{61}	.7	−.028	.047	.083	1.000	$\lambda_{22,3}$	.7	−.042	.061	.090	1.000
λ_{71}	.7	.033	.053	.090	1.000	$\lambda_{23,3}$	.7	−.018	.051	.102	1.000
λ_{81}	.7	.036	.053	.091	1.000	$\lambda_{24,3}$	.7	−.013	.044	.103	1.000
λ_{92}	.7	−.029	.051	.088	1.000	ψ_{78}	.3	−.164	.166	.050	.990
$\lambda_{10,2}$	.7	−.031	.052	.085	1.000	$\psi_{16,15}$	.3	−.156	.160	.063	.800
$\lambda_{11,2}$	.7	−.033	.056	.086	1.000	$\psi_{24,23}$	.3	−.151	.155	.065	.790
$\lambda_{12,2}$	.7	−.031	.054	.085	1.000	$\psi_{24,15}$	.3	−.181	.184	.059	.600
$\lambda_{13,2}$	.7	−.029	.049	.086	1.000	$\psi_{24,16}$	.3	−.182	.185	.059	.620
$\lambda_{14,2}$	.7	−.033	.052	.087	1.000	$\psi_{23,15}$	.3	−.183	.186	.059	.540
$\lambda_{15,2}$	.7	.006	.045	.101	1.000	$\psi_{23,16}$	.3	−.183	.186	.059	.580
$\lambda_{16,2}$	.7	.007	.049	.100	1.000	ψ_0	.0	.003	.003	.032	.000
λ_0	.0	.022	.052	.097	.002						

Note. T, True; $\psi_{16,15}$, $\psi_{24,23}$, $\psi_{24,16}$, $\psi_{24,15}$, $\psi_{23,16}$, and $\psi_{23,15}$ are clustered LD equivalent to the minor loadings.

TABLE 9.
Step Two and Three results in study 4

Par	True	Step Two (PCIRM-LI)				Step Three (PCIRM)			
		BIAS	RMSE	SE	SIG%	BIAS	RMSE	SE	SIG%
λ_{11}	.7	−.037	.056	.056	1.000	−.025	.045	.085	1.000
λ_{21}	.7	−.043	.063	.057	1.000	−.031	.052	.087	1.000
λ_{31}	.7	−.039	.061	.057	1.000	−.028	.048	.086	1.000
λ_{41}	.7	−.039	.059	.057	1.000	−.033	.050	.088	1.000
λ_{51}	.7	−.035	.059	.057	1.000	−.025	.049	.086	1.000
λ_{61}	.7	−.042	.060	.057	1.000	−.036	.052	.085	1.000
λ_{71}	.7	.147	.153	.060	1.000	.029	.048	.089	1.000
λ_{81}	.7	.144	.150	.059	1.000	.029	.046	.090	1.000
λ_{92}	.7	−.005	.041	.060	1.000	−.023	.046	.094	1.000
$\lambda_{10,2}$	.7	−.006	.043	.060	1.000	−.029	.051	.095	1.000
$\lambda_{11,2}$	.7	.000	.048	.061	1.000	−.021	.049	.093	1.000
$\lambda_{12,2}$	.7	−.005	.047	.060	1.000	−.031	.055	.092	1.000
$\lambda_{13,2}$	.7	−.005	.040	.060	1.000	−.032	.049	.094	1.000
$\lambda_{14,2}$	.7	−.009	.043	.060	1.000	−.029	.046	.096	1.000
$\lambda_{15,2}$	.7	−.017	.049	.123	1.000	−.092	.103	.171	.990
$\lambda_{16,2}$	.7	−.015	.053	.121	1.000	−.094	.107	.169	1.000
$\lambda_{17,3}$	.7	.003	.042	.063	1.000	−.026	.047	.089	1.000
$\lambda_{18,3}$	.7	−.011	.046	.063	1.000	−.036	.055	.089	1.000
$\lambda_{19,3}$	.7	−.007	.041	.063	1.000	−.026	.046	.089	1.000
$\lambda_{20,3}$	.7	−.004	.042	.062	1.000	−.033	.049	.089	1.000
$\lambda_{21,3}$	.7	−.007	.037	.062	1.000	−.034	.049	.085	1.000
$\lambda_{22,3}$	.7	−.008	.045	.063	1.000	−.035	.053	.087	1.000

TABLE 9.
continued

Par	True	Step Two (PCIRM-LI)				Step Three (PCIRM)			
		BIAS	RMSE	SE	SIG%	BIAS	RMSE	SE	SIG%
$\lambda_{23,3}$	.7	.013	.054	.113	1.000	−.022	.055	.150	1.000
$\lambda_{24,3}$	.7	.007	.042	.114	1.000	−.027	.051	.150	1.000
$\lambda_{15,4}$	.55	.030	.053	.081	1.000	−.032	.089	.135	.970
$\lambda_{16,4}$	.55	.028	.049	.079	1.000	−.038	.092	.132	.980
$\lambda_{23,4}$	.55	−.014	.053	.095	.970	−.020	.085	.130	.980
$\lambda_{24,4}$	.55	−.012	.048	.096	.980	−.019	.081	.133	.990
λ_0	.0	.008	.045	.074	.003	.018	.047	.110	.000
ψ_{78}	.3					−.164	.166	.051	.990
ψ_0	.0					.003	.003	.028	.000

The above three-step approach was employed to identify possible minor trait. The Markov chain reached stationary status within 20,000 iterations, which was set as the burn-in. After that, an additional 40,000 iterations were drawn for estimation, which took about 10 and 26 minutes for the PCIRM-LI and PCIRM, respectively, on an HP Envy laptop with an Intel Core i7 CPU. In Step One, the PCIRM with three traits and all major loadings specified resulted in three significant minor loadings and 11 significant residual covariance terms, suggesting some degree of within-item multidimensionality and local dependence (Table 10). Moreover, items 25, 26, and 29 were involved in several residual terms, implying the possibility of a minor trait. In Step Two, the PCIRM-LI with an additional trait by the three items was fitted and the results can be found in the same table. Ignoring the LD, a minor trait measured by eight items was identified. Other than that, the two steps gave remarkably similar results. Note that although some minor loadings that were insignificant in Step One became marginally significant in Step Two, their point estimates were still close.

In Step Three, the PCIRM with the three ordinary traits by major loadings and minor trait by the eight items was fitted and the results can be found in Table 11. In addition to the major loadings, the model identified ten significant minor loadings, six of which loaded on the minor trait, and four significant residual terms. Except for the residual terms, estimates between Steps Two and Three were close, although some minor loadings marginally significant in Step Two became insignificant in Step Three. It is noteworthy most minor loadings in the minor trait were from items targeted at the third ordinary trait, which implied that the operational definition of the trait might be of concern.

To triangulate the analysis, a regular MIRT model was fitted using Mplus (Muthén & Muthén, 1998–2015) with the default categorical CFA parameterization and robust weighted least squares estimation. All significant loadings and residual terms in Step Three were free to estimate, while other loadings and residual terms were fixed at zero. As shown in Table 11, the results were close enough except for the residual terms which, consistent with the simulation findings, were substantially higher than those in Step Three. The goodness of fit for the MIRT model was satisfactory (root mean squares error of approximation = .034; 90% confidence interval = (.031, .037); comparative fit index = .946; Tucker-Lewis index = .941).

3. Discussion

For test development under the MIRT setting, within-item dimensionality and local dependence are two critical issues to address. The issues can be transformed into specification of the loading and residual structures under the exploratory or confirmatory approaches. In practice, however, a different amount of substantive knowledge can be available, making the exploratory and confirmatory approaches two ends of a continuum. With the capacity of addressing both issues, the partially confirmatory approach with Bayesian Lasso is flexible to accommodate a wide range of the substantive continuum that would be challenging otherwise. Specifically, the proposed approach can be derived using the Bayesian hierarchical formulation with the Bayesian regression and covariance Lasso. By combining the Gibbs sampler and adapted MH algorithm, the PCIRM with two variants can be implemented with different constraints, depending on the availability of substantive knowledge.

When Constraint I of one specified loading per item can be satisfied, the PCIRM is suggested so that both issues can be addressed simultaneously. The constraint is practical for test developers, since items are often designed to target at a specific trait. Although there will be substantial underestimation of local dependence, the power to detect significant effect is still large. When the substantive knowledge cannot satisfy Constraint I, the PCIRM-LI with Constraint III with at least one specified loading per trait can be adopted to recover the loading structure. It can be useful if test

TABLE 10.
Steps One and Two for the EPQ-R scale

Item	Step One				Step Two					COV & COR	
	P	E	N	Int	P	E	N	MT	Int	Step One	
1	.609			−1.116	.616			**.236**	−1.121	$\psi_{2,12}$	.172
2	.544			.235	.622				.250	$\psi_{18,20}$	.186
3	.557			−1.137	.588				−1.141	$\psi_{14,22}$	.088
4	.394			−.488	.394				−.489	$\psi_{13,24}$	.095
5	.506			−1.513	.528				−1.474	$\psi_{25,26}$	.219
6	.429		−**.317**	−.643	.477		−**.327**		−.626	$\psi_{25,27}$	.160
7	.415		−***.177***	−1.193	.417		−**.188**		−1.213	$\psi_{25,29}$	.211
8	.551			−1.111	.562				−1.092	$\psi_{26,29}$	.216
9	.515	−**.222**		−1.321	.543	−**.235**			−1.315	$\psi_{31,35}$	.117
10	.437			−1.044	.451				−1.026	$\psi_{26,36}$	.153
11	.579			−1.359	.582				−1.346	$\psi_{29,36}$	.179
12	.521			−*.074*	.598				−*.077*	ϕ_{12}	*.067*
13		.755		.441		.789			.436	ϕ_{13}	*.018*
14		.782		.646		.845			.644	ϕ_{23}	−.233
15		.663		.648		.667			.654		
16		.635		1.297		.645			1.267	Step Two	
17		.645		.289		.652			.271	ϕ_{12}	*.087*
18		.720		−.482		.855		−**.288**	−.473	ϕ_{13}	*.032*
19	−***.190***	.663		1.335	−**.209**	.679			1.289	ϕ_{23}	−.250
20		.703		−.237		.823		−**.304**	−.221	ϕ_{14}	−*.007*
21		.585		.576		.582			.560	ϕ_{24}	*.109*
22		.773		.303		.833			.303	ϕ_{34}	−*.095*
23		.802		.197		.816			.192		
24		.756		.447		.789			.448		
25			.607	.472			.642	**.609**	.486		
26			.480	.343			.498	**.596**	.351		

TABLE 10.
continued

Item	Step One				Step Two					COV & COR
	P	E	N	Int	P	E	N	MT	Int	Step One
27	***.173***		.566	−.369	**.171**		.568	**.400**	−.386	
28			.588	.600			.594		.585	
29			.604	*.050*			.641	**.561**	*.038*	
30		***.152***	.610	−*.004*		**.162**	.618		−*.019*	
31			.724	−.435			.820		−.412	
32	−***.200***		.773	.343	−**.207**		.813		.325	
33			.709	−.609			.750		−.579	
34	−**.245**		.603	.310	−**.279**		.616		.291	
35			.738	−.456			.831		−.428	
36			.512	−.478			.538	**.324**	−.485	

Note. COV & COR, covariance and correlation; P, psychoticism; E, extraversion; N, neuroticism; MT, minor trait; Int, intercepts; minor loadings were bolded; insignificant estimates were italicized.

TABLE 11.
Step Three and MIRT for the EPQ-R scale

Item	Step Three				MIRT					COV & COR	
	P	E	N	MT	P	E	N	MT	Int	Step Three	
1	.603			***.248***	.649				-1.127	$\psi_{2,12}$	.180
2	.532				.553				.252	$\psi_{18,20}$	.149
3	.560				.533				-1.144	$\psi_{14,22}$	.083
4	.392				.437				− .490	$\psi_{13,24}$	.094
5	.533				.498				-1.482	ϕ_{12}	*.093*
6	.416		**− .297**		.397		**− .288**		− .626	ϕ_{13}	*− .008*
7	.395		***− .189***		.492				-1.218	ϕ_{23}	− .225
8	.555				.505				-1.099	ϕ_{14}	*.011*
9	.527	**− .226**			.575	**− .318**			-1.320	ϕ_{24}	*.075*
10	.436				.497				-1.029	ϕ_{34}	*− .027*
11	.584				.643				-1.357		
12	.517				.517				*− .076*	MIRT	
13		.750				.685			.436	$\psi_{2,12}$	.251
14		.791				.764			.644	$\psi_{18,20}$	.393
15		.661				.692			.656	$\psi_{14,22}$	.260
16		.631				.710			1.271	$\psi_{13,24}$	.250
17		.648				.650			.271	ϕ_{12}	.143
18		.740		***− .204***		.744			− .480	ϕ_{13}	*.004*
19	**− .222**	.660			**− .242**	.756			1.291	ϕ_{23}	− .236
20		.726		**− .224**		.698		**− .134**	− .233	ϕ_{14}	.140
21		.585				.604			.564	ϕ_{24}	*.058*
22		.779				.755			.303	ϕ_{34}	*− .069*
23		.794				.868			.193		
24		.748				.766			.450		

TABLE 11.
continued

Item	Step Three				MIRT					COV & COR
	P	E	N	MT	P	E	N	MT	Int	Step Three
25			.567	**.583**			.519	**.708**	.490	
26			.428	**.594**			.449	**.579**	.354	
27	***.168***		.530	**.412**			.515	**.468**	− .383	
28			.560				.612		.589	
29			.566	**.549**			.578	**.611**	*.040*	
30		***.155***	.594				.558		− *.018*	
31			.780				.845		− .413	
32	− ***.171***		.786				.806		.331	
33			.737				.746		− .582	
34	− **.223**		.588		− **.268**		.670		.299	
35			.793				.828		− .430	
36			.488	**.350**			.521	**.417**	− .484	

Note. Intercept estimates (Int) for Step Three were close to those in Step One or Two and omitted.

development is conducted under a more exploratory setting in which the targeted traits are unclear for many items. Then the PCIRM or PCIRM-LD with a known loading structure can be used for local dependence as the second step. Simulation studies showed that, although consuming more running time, the proposed models outperformed existing models with or without the conditions of local dependence. Moreover, a three-step approach provides a road map to identify a minor trait, which can offer a simpler interpretation to the clustered local dependence and reduce the bias when estimating the residual covariance. In the real-life example, the three-step approach identified a simpler model with a minor trait and a few residual terms.

The proposed approach was still workable even when the sample size was about moderate (e.g., ~ 250), although a larger sample size performed better. This can be especially beneficial for applied researchers without a large sample size. Applied researchers can also add an intermediate state of test analysis in between content analysis with a small number of examinees and large-scale data collection. Note that due to the nature of the L_1-norm penalty, the PCIRM can perform loading selection even better with a larger number of latent traits. We encourage researchers to explore different sets of constraints under which the proposed models can function appropriately across various settings. Future studies can also extend the approach to accommodate polytomous responses or a mix of both dichotomous and partial-credit items, which are common in psychological or educational testing. Finally, more work is needed to fully understand the performance and utility of the proposed approach across a wider range of settings, such as higher dimensionality, more complex structures, and/or different amount of prespecified information.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendix A. Derivation of Conditional Distributions

Equation (2) can be rewritten as $y_{ij} = \boldsymbol{\lambda}_j^{\mathrm{T}}\boldsymbol{\theta}_i + \mu_j + \varepsilon_{ij}$. After rearranging the loading vector for item j, the trait vector is $\boldsymbol{\theta}_i = \begin{pmatrix}\boldsymbol{\theta}_i' \\ \boldsymbol{\theta}_i''\end{pmatrix}$, where the $K' \times 1$ vector $\boldsymbol{\theta}_i'$ and $K'' \times 1$ vector $\boldsymbol{\theta}_i''$ are associated with the unspecified loadings and zero/specified loadings, respectively. Similarly, one can denote the rearranged trait matrix as $\boldsymbol{\Theta} = \begin{pmatrix}\boldsymbol{\Theta}' \\ \boldsymbol{\Theta}''\end{pmatrix} = \begin{pmatrix}\boldsymbol{\theta}_1', \ldots \boldsymbol{\theta}_i', \ldots \boldsymbol{\theta}_N' \\ \boldsymbol{\theta}_1'', \ldots \boldsymbol{\theta}_i'', \ldots \boldsymbol{\theta}_N''\end{pmatrix}$. Denote $\mathbf{y}_{i(-j)}$ as the subvector of $\mathbf{y}_i$ with the jth element deleted. The conditional distribution for $\boldsymbol{\lambda}_j'$ can be obtained as:

$$
\begin{aligned}
&p(\boldsymbol{\lambda}_j' \,|\, \mathbf{Z}, \boldsymbol{\Lambda}_{-j}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\delta}_j) = p(\boldsymbol{\lambda}_j' \,|\, \mathbf{Y}_j, \mathbf{Y}_{-j}, \boldsymbol{\Theta}, \boldsymbol{\Lambda}_{-j}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\delta}_j) \\
&\propto p(\mathbf{Y}_j | \mathbf{Y}_{-j}, \boldsymbol{\Theta}, \boldsymbol{\lambda}_j', \boldsymbol{\Lambda}_{-j}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Psi}) p(\boldsymbol{\lambda}_j') \\
&= \prod_{i=1}^{N} p(y_{ij} | \mathbf{y}_{i(-j)}, \boldsymbol{\theta}_i, \boldsymbol{\lambda}_j', \boldsymbol{\Lambda}_{-j}, \boldsymbol{\mu}, \boldsymbol{\Psi}) p(\boldsymbol{\lambda}_j') \\
&\propto \prod_{i=1}^{N} \exp\left\{\frac{1}{2}\left[y_{ij} - \mu_j - (\boldsymbol{\lambda}_j')^{\mathrm{T}}\boldsymbol{\theta}_i' - \boldsymbol{\psi}_j^{\mathrm{T}}\boldsymbol{\psi}_{-jj}^{-1}(\mathbf{y}_{i(-j)} - \boldsymbol{\mu}_{-j} - \boldsymbol{\Lambda}_{-j}\boldsymbol{\theta}_i)\right]^{\mathrm{T}} \times \right. \\
&\left. (\psi_{jj} - \boldsymbol{\psi}_j^{\mathrm{T}}\boldsymbol{\Psi}_{-jj}^{-1}\boldsymbol{\psi}_j)^{-1}\left[y_{ij} - \mu_j - (\boldsymbol{\lambda}_j')^{\mathrm{T}}\boldsymbol{\theta}_i' - \boldsymbol{\psi}_j^{\mathrm{T}}\boldsymbol{\Psi}_{-jj}^{-1}(\mathbf{y}_{i(-j)} - \boldsymbol{\mu}_{-j} - \boldsymbol{\Lambda}_{-j}\boldsymbol{\theta}_i)\right]\right\} \times p(\boldsymbol{\lambda}_j').
\end{aligned}
$$

With $\psi_{jj}^{*} = \psi_{jj} - \boldsymbol{\psi}_j^{\mathrm{T}} \boldsymbol{\Psi}_{-jj}^{-1} \boldsymbol{\psi}_j$, $\mathbf{Y}_j^{*} = \mathbf{Y}_j - \boldsymbol{\psi}_j^{\mathrm{T}} \boldsymbol{\Psi}_{-jj}^{-1} \mathbf{Y}_{-j}$ or $y_{ij}^{*} = y_{ij} - \boldsymbol{\psi}_j^{\mathrm{T}} \boldsymbol{\Psi}_{-jj}^{-1} \mathbf{y}_{i(-j)}$, and $\boldsymbol{\mu}_j^{*} = \mu_j - \boldsymbol{\psi}_j^{\mathrm{T}} \boldsymbol{\Psi}_{-jj}^{-1} (\boldsymbol{\mu}_{-j} + \boldsymbol{\Lambda}_{-j} \boldsymbol{\Theta})$ or $\mu_{ij}^{*} = \mu_j - \boldsymbol{\psi}_j^{\mathrm{T}} \boldsymbol{\Psi}_{-jj}^{-1} (\boldsymbol{\mu}_{-j} + \boldsymbol{\Lambda}_{-j} \boldsymbol{\theta}_i)$, it becomes:

$$\begin{aligned}
&\prod_{i=1}^{N} \exp\left\{ -\frac{1}{2} (y_{ij}^{*} - \mu_{ij}^{*} - (\boldsymbol{\lambda}_j')^{\mathrm{T}} \boldsymbol{\theta}_i')^{\mathrm{T}} \psi_{jj}^{*-1} (y_{ij}^{*} - \mu_{ij}^{*} - (\boldsymbol{\lambda}_j')^{\mathrm{T}} \boldsymbol{\theta}_i') \right\} \times p(\boldsymbol{\lambda}_j') \\
&\quad = \exp\left\{ -\frac{1}{2} \psi_{jj}^{*-1} (\mathbf{Y}_j^{*} - \boldsymbol{\mu}_j^{*} - (\boldsymbol{\lambda}_j')^{\mathrm{T}} \boldsymbol{\Theta}')^{\mathrm{T}} (\mathbf{Y}_j^{*} - \boldsymbol{\mu}_j^{*} - (\boldsymbol{\lambda}_j')^{\mathrm{T}} \boldsymbol{\Theta}') \right\} \times p(\boldsymbol{\lambda}_j').
\end{aligned}$$

With $\boldsymbol{\lambda}_j' \sim N(\mathbf{0}, \mathbf{D}_{\tau_j})$, one has:

$$\begin{aligned}
&\boldsymbol{\lambda}_j' \Big| \mathbf{Z}, \boldsymbol{\Lambda}_{-j}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \delta_j \sim N[(\psi_{jj}^{*-1} (\boldsymbol{\Theta}'(\boldsymbol{\Theta}')^{\mathrm{T}}) + \mathbf{D}_{\tau_j}^{-1})^{-1} \psi_{jj}^{*-1} \boldsymbol{\Theta}' (\mathbf{Y}_j^{*} - \boldsymbol{\mu}_j^{*}), \\
&\quad (\psi_{jj}^{*-1} (\boldsymbol{\Theta}'(\boldsymbol{\Theta}')^{\mathrm{T}}) + \mathbf{D}_{\tau_j}^{-1})^{-1}].
\end{aligned}$$

The conditional distribution for τ_{jk} can be expressed as:

$$\begin{aligned}
&p(\tau_{jk}^2 \,|\mathbf{Y}, \boldsymbol{\Lambda}, \boldsymbol{\Theta}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Psi}) \propto p(\lambda_{jk} |\, \tau_{jk}^2) p(\tau_{jk}^2) \\
&\quad \propto (\tau_{jk}^2)^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2\tau_{jk}^2} \lambda_{jk}^2 \right\} \exp\left\{ -\frac{\delta_j^2}{2} \tau_{jk}^2 \right\}. \\
&\quad \Rightarrow p\left(\frac{1}{\tau_{jk}^2} \middle| \mathbf{Y}, \boldsymbol{\Lambda}, \boldsymbol{\Theta}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Psi} \right) \propto \text{Inv-Gaussian}\left(\sqrt{\frac{\delta_j^2}{\lambda_{jk}^2}}, \delta_j^2 \right).
\end{aligned}$$

The conditional distribution for δ_j can be expressed as:

$$\begin{aligned}
&p(\delta_j^2 \,\Big| \mathbf{Y}, \boldsymbol{\Lambda}, \boldsymbol{\Theta}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Psi}) \propto p(\tau_{jk}^2 \Big| \, \delta_j^2) p(\delta_j^2) \\
&\quad \propto \frac{\delta_j^2}{2} \exp\left\{ -\frac{\delta_j^2}{2} \tau_{jk}^2 \right\} \times (\delta_j^2)^{\alpha_j - 1} \exp\left\{ -\beta_j (\delta_j^2) \right\} \\
&\quad \propto (\delta_j^2)^{(\alpha_j + 1) - 1} \exp\left\{ -(\frac{\tau_{jk}^2}{2} + \beta_j)(\delta_j^2) \right\} \\
&\quad \propto Gamma(\alpha_j + 1, \frac{\tau_{jk}^2}{2} + \beta_j).
\end{aligned}$$

Similarly, with $\boldsymbol{\lambda}_j'' \sim N(\lambda_{0j}, \mathbf{H}_{0j})$, the conditional distribution for $\boldsymbol{\lambda}_j''$ can be obtained as:

$$\begin{aligned}
&\boldsymbol{\lambda}_j'' \Big| \mathbf{Z}, \boldsymbol{\Lambda}_{-j}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Psi} \sim N[(\psi_{jj}^{*-1} (\boldsymbol{\Theta}'(\boldsymbol{\Theta}')^{\mathrm{T}}) + \mathbf{H}_{0j}^{-1})^{-1} \psi_{jj}^{*-1} \boldsymbol{\Theta}' ((\mathbf{Y}_j^{*} - \boldsymbol{\mu}_j^{*}) + \lambda_{0j} \mathbf{H}_{0j}^{-1}), \\
&\quad (\psi_{jj}^{*-1} (\boldsymbol{\Theta}'(\boldsymbol{\Theta}')^{\mathrm{T}}) + \mathbf{H}_{0j}^{-1})^{-1}].
\end{aligned}$$

The conditional distributions $p(\theta_i|\mathbf{Y}, \mathbf{\Lambda}, \mu, \mathbf{\Phi}, \mathbf{\Psi})$ and $p(\mu|\mathbf{Z}, \mathbf{\Lambda}, \mathbf{\Phi}, \mathbf{\Psi})$ are similar to those presented in Lee (2007, pp. 146–147), as:

$$\theta_i|\mathbf{X}, \mathbf{Y}, \mathbf{\Lambda}, \mu, \mathbf{\Phi}, \mathbf{\Psi} \sim N[(\mathbf{\Phi}^{-1} + \mathbf{\Lambda}^{\mathrm{T}}\mathbf{\Psi}^{-1}\mathbf{\Lambda})^{-1}\mathbf{\Lambda}^{\mathrm{T}}\mathbf{\Psi}^{-1}(\mathbf{y}_i - \mu), (\mathbf{\Phi}^{-1} + \mathbf{\Lambda}^{\mathrm{T}}\mathbf{\Psi}^{-1}\mathbf{\Lambda})^{-1}],$$
$$\mu|\mathbf{Z}, \mathbf{\Lambda}, \mathbf{\Phi}, \mathbf{\Psi} \sim N[(\mathbf{H}_{\mu 0}^{-1} + N\mathbf{\Psi}^{-1})^{-1}(\mathbf{\Psi}^{-1}\mathbf{V} + \mu_0\mathbf{H}_{\mu 0}^{-1}), (\mathbf{H}_{\mu 0}^{-1} + N\mathbf{\Psi}^{-1})^{-1}],$$

where $\mathbf{V} = \sum_{i=1}^{N}(\mathbf{y}_i - \mathbf{\Lambda}\theta_i)$.

For PCIRM-LI, $\mathbf{\Psi} = \mathrm{diag}(\psi_{jj})$ is modeled as a diagonal matrix. The conditional distribution for ψ_{jj} can be expressed as: $p(\psi_{jj}^{-1}\big|\mathbf{Z}, \mathbf{\Lambda}, \mu, \mathbf{\Phi}) \propto p(\mathbf{y}_j|\mathbf{y}_{-j}, \mathbf{\Lambda}, \mathbf{\Theta}, \mu, \mathbf{\Phi})p(\psi_{jj}^{-1})$. It can be further simplified as:

$$p(\psi_{jj}\big|\mathbf{Z}, \mathbf{\Lambda}, \mu, \mathbf{\Phi}) \propto \text{Inv-Gamma}(\alpha_{0j} + \frac{N}{2} - 1, \beta_{0j} + \frac{1}{2}\sum_{i=1}^{N}[y_{ij} - \mu_j - (\lambda_j'')^{\mathrm{T}}\theta_i'']^2).$$

Appendix B. Block Gibbs Sampler to Draw Σ and Ψ with Local Dependence

The conditional distribution $p(\mathbf{\Sigma}|\mathbf{Y}, \mu, \mathbf{\Theta}, \mathbf{\Lambda}, \mathbf{\Phi}, \boldsymbol{\tau}_s, \delta_s)$ can be decomposed as follows:

$$p(\mathbf{\Sigma}|\mathbf{Y}, \mu, \mathbf{\Theta}, \mathbf{\Lambda}, \mathbf{\Phi}, \boldsymbol{\tau}_s, \delta_s) \propto p(\mathbf{Y}|\mathbf{\Sigma}, \mu, \mathbf{\Theta}, \mathbf{\Lambda}, \mathbf{\Phi})p(\boldsymbol{\tau}_s, \delta_s)$$
$$\propto |\mathbf{\Sigma}|^{N/2}\exp\left[-\mathrm{tr}\left(-\frac{1}{2}\mathbf{S}\mathbf{\Sigma}\right)\right]\prod_{i<j}\exp\left(-\frac{\sigma_{ij}^2}{2\tau_{ij}}\right) \times \prod_{j=1}^{J}\exp\left(-\frac{\delta_s\sigma_{jj}}{2}\right)\mathrm{I}(\mathbf{\Sigma} > 0)$$

where $\boldsymbol{\tau}_s = (\tau_{ij})_{i<j}$ is the vector of the latent scale parameters, and

$$\mathbf{S} = \sum_{i=1}^{N}(\mathbf{y}_i - \mu - \mathbf{\Lambda}\boldsymbol{\theta}_i)(\mathbf{y}_i - \mu - \mathbf{\Lambda}\boldsymbol{\theta}_i)^{\mathrm{T}}.$$

For $j = 1, \dots, J$ and without loss of generality, one can partition and rearrange the columns of $\mathbf{\Sigma}$ and $\mathbf{S}$ as follows:

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{-jj} & \boldsymbol{\sigma}_j \\ \boldsymbol{\sigma}_j^{\mathrm{T}} & \sigma_{jj} \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} \mathbf{S}_{-jj} & \mathbf{s}_j \\ \mathbf{s}_j^{\mathrm{T}} & s_{jj} \end{pmatrix}$$

where σ_{jj} is the jth diagonal element of $\mathbf{\Sigma}$, $\boldsymbol{\sigma}_j = (\sigma_{j1}, \dots, \sigma_{j,j-1}, \sigma_{j,j+1}, \dots, \sigma_{jJ})^{\mathrm{T}}$ is the vector of all off-diagonal elements of the jth column, and $\mathbf{\Sigma}_{-jj}$ is the $(J-1) \times (J-1)$ matrix resulting from deleting the jth row and jth column of $\mathbf{\Sigma}$. Similar, s_{-jj} is the jth diagonal element of $\mathbf{S}$, $\mathbf{s}_j$ is the vector of all-diagonal elements of the jth column of $\mathbf{S}$, and $\mathbf{S}_{-jj}$ is the matrix with the jth row and jth column of $\mathbf{S}$ deleted. Then we have:

$$p\left(\boldsymbol{\sigma}_j, \sigma_{jj}|\mathbf{\Sigma}_{-jj}, \mathbf{Y}, \mathbf{\Theta}, \mu, \mathbf{\Lambda}, \mathbf{\Phi}, \boldsymbol{\tau}_s, \delta_s\right)$$
$$\propto \left(\sigma_{jj} - \boldsymbol{\sigma}_j^{\mathrm{T}}\mathbf{\Sigma}_{-jj}^{-1}\boldsymbol{\sigma}_j\right)^{\frac{N}{2}} \times \exp\left\{-\frac{1}{2}[\boldsymbol{\sigma}_j^{\mathrm{T}}\mathbf{M}_\tau^{\mathrm{T}}\boldsymbol{\sigma}_j + 2\mathbf{s}_j^{\mathrm{T}}\boldsymbol{\sigma}_j + \left(s_{jj} + \delta_s\right)\sigma_{jj}]\right\}$$

where $\mathbf{M}_\tau$ is the diagonal matrix with diagonal elements $\tau_{j1}, \dots, \tau_{j,j-1}, \tau_{j,j+1}, \dots, \tau_{jJ}$.

Let $\boldsymbol{\beta} = \boldsymbol{\sigma}_j$ and $\boldsymbol{\gamma} = \sigma_{jj} - \boldsymbol{\sigma}_j^{\mathrm{T}} \boldsymbol{\Sigma}_{-jj}^{-1} \boldsymbol{\sigma}_j$. It can be shown that:

$$p\left(\boldsymbol{\beta} | \boldsymbol{\Sigma}_{-jj}, \mathbf{Y}, \boldsymbol{\Theta}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\Phi}, \boldsymbol{\tau}_s, \delta_s\right)$$

$$\propto N\left(-\left[\left(s_{jj} + \delta_s\right) \boldsymbol{\Sigma}_{-jj}^{-1} + \mathbf{M}_\tau^{-1}\right]^{-1} \mathbf{s}_j, \left[\left(s_{jj} + \delta_s\right) \boldsymbol{\Sigma}_{-jj}^{-1} + \mathbf{M}_\tau^{-1}\right]^{-1}\right),$$

$$p\left(\boldsymbol{\gamma} | \boldsymbol{\Sigma}_{-jj}, \mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\tau}_s, \delta_s\right) \propto Gamma\left(\frac{N}{2} + 1, \frac{s_{jj} + \delta_s}{2}\right).$$

After drawing from the above conditional distributions, we can obtain $\boldsymbol{\sigma}_j = \boldsymbol{\beta}$, $\boldsymbol{\sigma}_j^{\mathrm{T}} = \boldsymbol{\beta}^{\mathrm{T}}$ and $\sigma_{jj} = \boldsymbol{\gamma} + \boldsymbol{\sigma}_j^{\mathrm{T}} \boldsymbol{\Sigma}_{-jj}^{-1} \boldsymbol{\sigma}_j$, then the jth column and row of $\boldsymbol{\Sigma}$ can be updated one at a time. At the end, $\boldsymbol{\Psi} = \boldsymbol{\Sigma}^{-1}$is computed.
The conditional distribution for $\boldsymbol{\tau}_s = (\tau_{ij})_{i<j}$ can be expressed as:

$$p(\boldsymbol{\tau}_s | \mathbf{Y}, \boldsymbol{\Lambda}, \boldsymbol{\Theta}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}, \delta_s) \propto p(\boldsymbol{\Sigma} | \boldsymbol{\tau}_s, \delta_s) p(\boldsymbol{\tau}_s | \delta_s)$$

$$\propto \prod_{i<J} \tau_{ij}^{-\frac{1}{2}} \exp\left\{-\frac{\sigma_{ij}^2 + \tau_{ij}^2 \delta_s^2}{2\tau_{ij}}\right\}.$$

It can be shown that for $i < j$,

$$p\left(\frac{1}{\tau_{ij}} \middle| \mathbf{Y}, \boldsymbol{\Lambda}, \boldsymbol{\Theta}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}, \delta_s\right) \propto \text{Inv-Gaussian}\left(\sqrt{\frac{\delta_s^2}{\sigma_{ij}^2}}, \delta_s^2\right).$$

The conditional distribution for δ_s can be expressed as:

$$p(\delta_s | \mathbf{Y}, \boldsymbol{\Lambda}, \boldsymbol{\Theta}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}, \boldsymbol{\tau}) \propto p(\boldsymbol{\tau} | \delta_s) p(\delta_s)$$

$$\propto Gamma\left(\alpha_s + \frac{J(J+1)}{2}, \beta_s + \frac{1}{2} \sum_{i=1}^{J} \sum_{j=1}^{J} \left|\sigma_{ij}\right|\right).$$

Appendix C. Details of Nonzero Loading Estimates

See Tables 12, 13, 14 and 15.

TABLE 12.
Nonzero loading estimates for PCIRM-LI and PCIRM under local independence

Parameter	TRUE	PCIRM-LI				PCIRM			
		BIAS	RMSE	SE	SIG%	BIAS	RMSE	SE	SIG%
λ_{11}	.5	− .002	.051	.079	1.000	.006	.045	.095	1.000
λ_{21}	.5	.001	.047	.079	1.000	.008	.042	.095	1.000
λ_{31}	.7	− .002	.047	.077	1.000	− .004	.044	.102	1.000
λ_{41}	.7	− .013	.051	.077	1.000	− .014	.047	.102	1.000
λ_{51}	.7	.001	.047	.077	1.000	− .001	.043	.102	1.000
λ_{61}	.7	.000	.046	.077	1.000	− .003	.041	.103	1.000
λ_{71}	.7	− .001	.049	.078	1.000	− .005	.047	.103	1.000
λ_{81}	.7	.012	.051	.079	1.000	− .003	.045	.102	1.000
$\lambda_{17,1}$	.3	− .036	.064	.092	.885	− .028	.055	.107	.845
$\lambda_{18,1}$	.3	− .030	.060	.092	.910	− .024	.052	.107	.855
λ_{12}	.5	.000	.050	.080	1.000	− .027	.053	.098	1.000
λ_{22}	.5	.001	.052	.081	1.000	− .025	.054	.098	1.000
λ_{92}	.7	.013	.044	.074	1.000	− .024	.044	.086	1.000
$\lambda_{10,2}$	.5	.011	.048	.074	1.000	.002	.045	.090	1.000
$\lambda_{11,2}$	.7	.013	.046	.073	1.000	− .001	.041	.095	1.000
$\lambda_{12,2}$	.7	.008	.048	.072	1.000	− .005	.045	.095	1.000
$\lambda_{13,2}$	.7	.007	.049	.073	1.000	− .006	.045	.096	1.000
$\lambda_{14,2}$	.7	.005	.044	.072	1.000	− .006	.042	.096	1.000
$\lambda_{15,2}$	.7	.014	.051	.073	1.000	.001	.044	.095	1.000
$\lambda_{16,2}$	.7	.014	.048	.073	1.000	− .009	.046	.095	1.000
λ_{93}	.3	− .026	.054	.087	.935	− .024	.050	.104	.890
$\lambda_{10,3}$	.5	− .009	.046	.078	1.000	− .026	.051	.094	1.000
$\lambda_{17,3}$	.7	.012	.043	.074	1.000	− .022	.043	.086	1.000
$\lambda_{18,3}$	.7	.016	.051	.074	1.000	− .021	.046	.086	1.000
$\lambda_{19,3}$	.7	.006	.046	.072	1.000	− .003	.044	.094	1.000
$\lambda_{20,3}$	.7	.008	.046	.071	1.000	− .002	.042	.094	1.000
$\lambda_{21,3}$	.7	.008	.049	.072	1.000	− .001	.045	.094	1.000
$\lambda_{22,3}$	.7	.007	.043	.071	1.000	− .001	.040	.094	1.000
$\lambda_{23,3}$	.7	.013	.050	.072	1.000	.002	.046	.094	1.000
$\lambda_{24,3}$	.7	.018	.049	.072	1.000	− .002	.041	.093	1.000

Note. TRUE, true values; BIAS, bias of the parameter estimates; SE, mean of the standard error estimates; RMSE, root mean squares error between the estimates and the true values; SIG%, percentage of estimates significantly different from zero at $\alpha = .05$.

TABLE 13.
Nonzero loading estimates for BSEMs and GLFM-DR under local independence

Par	TRUE	BSEM-DRt				BSEM-DRi				GLFM-DR	
		BIAS	RMSE	SE	SIG%	BIAS	RMSE	SE	SIG%	BIAS	RMSE
λ_{11}	.5	.275	.276	.091	1.000	.204	.210	.054	1.000	.059	.233
λ_{21}	.5	− .263	.264	.058	1.000	.199	.204	.054	1.000	.044	.220
λ_{31}	.7	− .539	.539	.074	.935	.056	.071	.055	1.000	.009	.317
λ_{41}	.7	− .545	.545	.075	.690	.050	.066	.057	1.000	− .005	.314
λ_{51}	.7	− .539	.539	.073	.930	.063	.075	.058	1.000	.007	.322
λ_{61}	.7	− .539	.539	.073	.895	.060	.072	.058	1.000	.012	.319
λ_{71}	.7	− .543	.543	.072	.880	.054	.070	.057	1.000	.009	.321
λ_{81}	.7	− .541	.541	.075	.815	.050	.066	.057	1.000	.129	.231
$\lambda_{17,1}$	.3	− .294	.294	.062	.000	− .209	.211	.039	.760	.033	.170
$\lambda_{18,1}$	.3	− .067	.068	.061	1.000	− .206	.207	.040	.810	.038	.172
λ_{12}	.5	− .498	.498	.063	.000	− .311	.313	.046	1.000	.026	.229
λ_{22}	.5	− .269	.269	.060	1.000	− .308	.310	.046	1.000	.034	.223
λ_{92}	.7	.086	.089	.085	1.000	.128	.133	.045	1.000	.030	.290
$\lambda_{10,2}$	.5	− .261	.261	.055	1.000	.142	.149	.055	1.000	.043	.210
$\lambda_{11,2}$	.7	− .525	.525	.071	.990	.037	.056	.055	1.000	− .007	.305
$\lambda_{12,2}$	.7	− .532	.532	.069	.995	.034	.055	.057	1.000	.000	.323
$\lambda_{13,2}$	.7	− .519	.520	.072	.995	.036	.058	.057	1.000	− .011	.302
$\lambda_{14,2}$	.7	− .531	.531	.072	.975	.036	.055	.057	1.000	− .003	.309
$\lambda_{15,2}$	.7	− .529	.529	.070	.995	.041	.061	.056	1.000	− .004	.316
$\lambda_{16,2}$	.7	− .529	.529	.070	.985	.038	.059	.056	1.000	.104	.252
λ_{93}	.3	− .300	.300	.064	.000	− .219	.220	.039	.565	.023	.162
$\lambda_{10,3}$	.5	− .254	.254	.058	1.000	− .268	.270	.044	1.000	.022	.215
$\lambda_{17,3}$	.7	.072	.077	.090	1.000	.117	.122	.045	1.000	.025	.323
$\lambda_{18,3}$	.7	− .457	.457	.061	1.000	.116	.122	.045	1.000	.011	.322
$\lambda_{19,3}$	.7	− .525	.525	.072	.990	.031	.052	.052	1.000	− .025	.357
$\lambda_{20,3}$	.7	− .528	.528	.071	1.000	.037	.056	.053	1.000	− .017	.348
$\lambda_{21,3}$	.7	− .530	.530	.078	.915	.040	.061	.053	1.000	− .019	.351
$\lambda_{22,3}$	.7	− .528	.528	.077	.930	.036	.054	.057	1.000	− .020	.354
$\lambda_{23,3}$	.7	− .528	.529	.074	.980	.037	.058	.052	1.000	− .015	.352
$\lambda_{24,3}$	.7	− .522	.522	.073	.985	.035	.056	.053	1.000	.082	.250

Note. TRUE, true values; BIAS, bias of the parameter estimates; SE, mean of the standard error estimates; RMSE, root mean squares error between the estimates and the true values; SIG%, percentage of estimates significantly different from zero at $\alpha = .05$.

TABLE 14.
Nonzero loading estimates for PCIRM-LI, PCIRM, and CIRM-LD under local dependence

Par	T	PCIRM-LI				PCIRM				CIRM-LD			
		B	R	SE	SIG%	B	R	SE	SIG%	B	R	SE	SIG%
λ_{11}	.5	− .020	.053	.079	1.000	.004	.044	.096	1.000	− .016	.045	.058	1.000
λ_{21}	.5	− .022	.055	.080	1.000	.005	.044	.099	1.000	− .014	.043	.059	1.000
λ_{31}	.7	.104	.129	.083	.995	.065	.082	.117	1.000	.017	.037	.048	1.000
λ_{41}	.7	.102	.128	.083	.995	.065	.083	.116	1.000	.020	.038	.048	1.000
λ_{51}	.7	− .047	.088	.077	.995	− .031	.058	.104	1.000	− .008	.032	.046	1.000
λ_{61}	.7	− .054	.087	.076	.995	− .035	.057	.105	1.000	− .010	.033	.046	1.000
λ_{71}	.7	− .034	.075	.077	.995	− .010	.045	.101	1.000	− .002	.031	.047	1.000
λ_{81}	.7	− .027	.073	.078	.995	− .017	.049	.101	1.000	− .007	.031	.046	1.000
$\lambda_{17,1}$	.3	− .040	.067	.091	.900	− .041	.059	.110	.765	.000	.039	.058	1.000
$\lambda_{18,1}$	.3	− .036	.066	.091	.905	− .035	.058	.110	.825	− .001	.046	.058	1.000
λ_{12}	.5	− .002	.061	.081	.995	− .034	.058	.105	1.000	− .008	.041	.056	1.000
λ_{22}	.5	− .003	.064	.082	.995	− .036	.058	.104	1.000	− .009	.044	.057	1.000
λ_{92}	.7	− .007	.053	.074	1.000	− .030	.047	.086	1.000	− .031	.045	.051	1.000
$\lambda_{10,2}$	.5	− .003	.048	.074	1.000	.002	.043	.091	1.000	− .011	.040	.055	1.000
$\lambda_{11,2}$	.7	.094	.117	.077	.995	.048	.065	.097	1.000	.015	.035	.045	1.000
$\lambda_{12,2}$	.7	.095	.122	.076	.995	.049	.068	.097	1.000	.015	.033	.046	1.000
$\lambda_{13,2}$	.7	− .022	.073	.073	.995	− .023	.049	.093	1.000	− .005	.027	.044	1.000
$\lambda_{14,2}$	.7	− .031	.070	.072	.995	− .041	.060	.092	1.000	− .011	.034	.044	1.000
$\lambda_{15,2}$	.7	− .011	.067	.073	.995	− .007	.045	.090	1.000	− .005	.031	.045	1.000
$\lambda_{16,2}$	.7	.001	.070	.074	.995	− .015	.045	.097	1.000	− .002	.031	.045	1.000
λ_{93}	.3	− .025	.056	.087	.980	− .030	.054	.102	.920	.002	.040	.054	1.000
$\lambda_{10,3}$	.5	− .015	.062	.077	.995	− .032	.056	.095	1.000	− .006	.039	.055	1.000
$\lambda_{17,3}$	.7	.000	.055	.073	1.000	− .023	.043	.087	1.000	− .025	.041	.052	1.000
$\lambda_{18,3}$	.7	− .002	.058	.073	1.000	− .025	.048	.091	1.000	− .026	.045	.053	1.000
$\lambda_{19,3}$	.7	.081	.110	.076	.995	.031	.054	.098	1.000	.011	.031	.045	1.000
$\lambda_{20,3}$	.7	.087	.115	.075	.995	.039	.057	.100	1.000	.018	.035	.044	1.000
$\lambda_{21,3}$	.7	− .026	.080	.072	.995	− .027	.052	.098	1.000	− .010	.034	.044	1.000
$\lambda_{22,3}$	.7	− .026	.075	.071	.995	− .031	.055	.098	1.000	− .008	.032	.044	1.000
$\lambda_{23,3}$	.7	− .004	.069	.072	.995	− .002	.046	.100	1.000	− .002	.028	.044	1.000
$\lambda_{24,3}$	.7	.002	.071	.071	.995	.000	.044	.096	1.000	− .002	.032	.044	1.000

Note. T, True; B, BIAS; R, RMSE; SE, mean of the standard error estimates; RMSE, root mean squares error between the estimates and the true values; SIG%, percentage of estimates significantly different from zero at $\alpha = .05$.

TABLE 15.
Nonzero loading estimates for BSEMs and GLFM-DR under local dependence

Par	TRUE	BSEM-DRt				BSEM-DRi				GLFM-DR	
		BIAS	RMSE	SE	SIG%	BIAS	RMSE	SE	SIG%	BIAS	RMSE
λ_{11}	.5	.270	.271	.092	1.000	.170	.176	.054	1.000	− .007	.230
λ_{21}	.5	− .265	.265	.058	1.000	.161	.168	.054	1.000	− .010	.224
λ_{31}	.7	− .537	.537	.073	.910	.139	.143	.049	1.000	.112	.434
λ_{41}	.7	− .541	.541	.075	.830	.140	.146	.052	1.000	.114	.431
λ_{51}	.7	− .536	.536	.072	.955	.022	.051	.058	1.000	− .076	.302
λ_{61}	.7	− .535	.535	.072	.935	.022	.052	.058	1.000	− .084	.302
λ_{71}	.7	− .545	.545	.072	.895	.033	.055	.056	1.000	− .074	.310
λ_{81}	.7	− .542	.542	.075	.815	.022	.051	.056	1.000	.043	.250
$\lambda_{17,1}$	.3	− .294	.294	.063	.000	− .199	.201	.039	.865	.015	.182
$\lambda_{18,1}$	.3	− .068	.069	.061	1.000	− .196	.197	.041	.865	.015	.181
λ_{12}	.5	− .498	.498	.063	.000	− .289	.291	.046	1.000	.043	.232
λ_{22}	.5	− .271	.271	.060	1.000	− .287	.289	.046	1.000	.030	.235
λ_{92}	.7	.079	.082	.086	1.000	.099	.107	.046	1.000	.006	.293
$\lambda_{10,2}$	.5	− .261	.261	.055	1.000	.120	.128	.054	1.000	.024	.185
$\lambda_{11,2}$	.7	− .523	.523	.071	1.000	.113	.119	.051	1.000	.061	.384
$\lambda_{12,2}$	.7	− .527	.527	.069	.995	.115	.122	.053	1.000	.063	.399
$\lambda_{13,2}$	.7	− .515	.515	.071	1.000	.004	.046	.057	1.000	− .052	.320
$\lambda_{14,2}$	.7	− .527	.527	.072	.990	.006	.048	.056	1.000	− .057	.310
$\lambda_{15,2}$	.7	− .531	.531	.070	.985	.021	.048	.055	1.000	− .039	.314
$\lambda_{16,2}$	.7	− .530	.530	.070	.990	.027	.051	.055	1.000	.076	.239
λ_{93}	.3	− .300	.300	.064	.000	− .204	.206	.040	.785	.040	.162
$\lambda_{10,3}$	.5	− .254	.254	.058	1.000	− .257	.259	.043	1.000	.031	.185
$\lambda_{17,3}$	.7	.068	.073	.091	1.000	.100	.106	.045	1.000	.030	.279
$\lambda_{18,3}$	.7	− .458	.458	.061	1.000	.095	.102	.046	1.000	.015	.278
$\lambda_{19,3}$	.7	− .522	.523	.071	1.000	.101	.108	.049	1.000	.061	.358
$\lambda_{20,3}$	.7	− .524	.524	.070	.990	.109	.115	.050	1.000	.065	.370
$\lambda_{21,3}$	.7	− .525	.525	.077	.940	.009	.047	.053	1.000	− .035	.287
$\lambda_{22,3}$	.7	− .524	.524	.076	.970	.013	.046	.057	1.000	− .039	.299
$\lambda_{23,3}$	.7	− .529	.529	.074	.965	.025	.049	.052	1.000	− .023	.300
$\lambda_{24,3}$	.7	− .524	.524	.073	.980	.024	.048	.053	1.000	.077	.180

Note. TRUE, true values; BIAS, bias of the parameter estimates; SE, mean of the standard error estimates; RMSE, root mean squares error between the estimates and the true values; SIG%, percentage of estimates significantly different from zero at $\alpha = .05$.

Appendix D. The Revised Eysenck Personality Questionnaire Short Scale

No.	Content	T
1	Would you take drugs which may have strange or dangerous effects?	P
2	Do you prefer to go your own way rather than act by the rules?	P
3	Do you think marriage is old fashioned and should be done away with?	P
4	Do you think people spend too much time safeguarding their future with savings and insurance?	P
5	Would you like other people to be afraid of you?	P
6(R)	Do you take much notice of what people think?	P
7(R)	Would being in debt worry you?	P
8(R)	Do good manners and cleanliness matter much to you?	P
9(R)	Do you enjoy co-operating with others?	P
10(R)	Does it worry you if you know there are mistakes in your work?	P
11(R)	Do you try not to be rude to people?	P
12(R)	Is it better to follow society's rules than go your own way?	P
13	Are you a talkative person?	E
14	Are you rather lively?	E
15	Can you usually let yourself go and enjoy yourself at a lively party?	E
16	Do you enjoy meeting new people?	E
17	Do you usually take the initiative in making new friends?	E
18	Can you easily get some life into a rather dull party?	E
19	Do you like mixing with people?	E
20	Can you get a party going?	E
21	Do you like plenty of bustle and excitement around you?	E
22	Do other people think of you as being very lively?	E
23(R)	Do you tend to keep in the background on social occasions?	E
24(R)	Are you mostly quiet when you are with other people?	E
25	Does your mood often go up and down?	N
26	Do you ever feel 'just miserable' for no reason?	N
27	Are you an irritable person?	N
28	Are your feelings easily hurt?	N
29	Do you often feel 'fed-up'?	N
30	Are you often troubled about feelings of guilt?	N
31	Would you call yourself a nervous person?	N
32	Are you a worrier?	N
33	Would you call yourself tense or 'highly-strung'?	N
34	Do you worry too long after an embarrassing experience?	N
35	Do you suffer from 'nerves'?	N
36	Do you often feel lonely?	N

Note. Item responses were dichotomous (i.e., Yes/No); T = targeted trait; P = psychoticism; E = extraversion; N = neuroticism; negatively worded items were marked by "R."

References

Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement*, *12*(3), 261–280. https://doi.org/10.1177/014662168801200305.

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. Reading, MA: Addison-Wesley.

Casella, G., & George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician*, *46*(3), 167–174. https://doi.org/10.1080/00031305.1992.10475878.

Chen, Y., Li, X., Liu, J., & Ying, Z. (2018). Robust measurement via a fused latent and graphical item response theory model. *Psychometrika*, *83*(3), 538–562. https://doi.org/10.1007/s11336-018-9610-4.

Chen, Y., Li, X., & Zhang, S. (2019). Structured latent factor analysis for large-scale data: Identifiability, estimability, and their implications. *Journal of the American Statistical Association*,. https://doi.org/10.1080/01621459.2019.1635485.

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis–Hastings algorithm. *The American Statistician*, *49*(4), 327–335. https://doi.org/10.1080/00031305.1995.10476177.

Eaton, M. L. (1983). *Multivariate statistics: A vector space approach*. Beachwood, OH: Institute of Mathematical Statistics.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: L. Erlbaum Associates.

Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika*, *82*(4), 904–927. https://doi.org/10.1007/s11336-017-9557-x.

Eysenck, S. B., & Barrett, P. (2013). Re-introduction to cross-cultural studies of the EPQ. *Personality and Individual Differences*, *54*(4), 485–489. https://doi.org/10.1016/j.paid.2012.09.022.

Eysenck, S. B., Eysenck, H. J., & Barrett, P. (1985). A revised version of the psychoticism scale. *Personality and Individual Differences*, *6*(1), 21–29. https://doi.org/10.1016/0191-8869(85)90026-1.

Forero, C. G., & Maydeu-Olivares, A. (2009). Estimation of IRT graded response models: Limited versus full information methods. *Psychological Methods*, *14*(3), 275–299. https://doi.org/10.1037/a0015825.

Gelman, A. (1996). Inference and monitoring convergence. In W. R. Gilks, S. Richardson, & D. J. Spiegelharter (Eds.), *Markov Chain Monte Carlo in practice* (pp. 131–144). London: Chapman & Hall.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). London: Chapman & Hall.

Gelman, A., Meng, X.-L., & Stern, H. S. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, *6*, 733–807.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*(6), 721–741. https://doi.org/10.1109/TPAMI.1984.4767596.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (Eds.). (1996). *Markov chain Monte Carlo in practice*. London: Chapman & Hall.

Gill, J. (2002). *Bayesian methods: A social and behavioral sciences approach*. Boca Raton, FL: Chapman & Hall/CRC. https://doi.org/10.1201/9781420057478.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their application. *Biometrika*, *57*(1), 97–109. https://doi.org/10.1093/biomet/57.1.97.

Jennrich, R. I., & Sampson, P. F. (1966). Rotation for simple loadings. *Psychometrika*, *31*(3), 313–323. https://doi.org/10.1007/BF02289465.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*(2), 183–202. https://doi.org/10.1007/BF02289343.

Khondker, Z. S., Zhu, H., Chu, H., Lin, W., & Ibrahim, J. G. (2013). The Bayesian covariance lasso. *Statistics and Its Interface*, *6*(2), 243–259. https://doi.org/10.4310/SII.2013.v6.n2.a8.

Lee, S.-Y. (2007). *Structural equation modeling: A Bayesian approach*. Hoboken, NJ: Wiley. https://doi.org/10.1002/9780470024737.

Liu, X. (2008). Parameter expansion for sampling a correlation matrix: An efficient GPX-RPMH algorithm. *Journal of Statistical Computation and Simulation*, *78*(11), 1065–1076. https://doi.org/10.1080/00949650701519635.

Liu, X., & Daniels, M. J. (2006). A new efficient algorithm for sampling a correlation matrix based on parameter expansion and re-parameterization. *Journal of Computational and Graphical Statistics*, *15*(4), 897–914. https://doi.org/10.1198/106186006X160681.

Lu, Z. H., Chow, S. M., & Loken, E. (2016). Bayesian factor analysis as a variable-selection problem: Alternative priors and consequences. *Multivariate Behavioral Research*, *51*(4), 519–539. https://doi.org/10.1080/00273171.2016.1168279.

Meng, X.-L. (1994). Posterior predictive p-values. *Annals of Statistics*, *22*(3), 1142–1160. https://doi.org/10.1214/aos/1176325622.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equations of state calculations by fast computing machine. *The Journal of Chemical Physics*, *21*(6), 1087–1092. https://doi.org/10.1063/1.1699114.

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17*(3), 313–335. https://doi.org/10.1037/a0026802.

Muthén, B. O., & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multi-group and growth modeling in Mplus. *Mplus web notes: No. 4*. Retrieved January 20, 2020, from http://www.statmodel.com/download/webnotes/CatMGLong.pdf.

Muthén, L. K., & Muthén, B. O. (1998–2015). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Pan, J., Ip, E. H., & Dubé, L. (2017). An alternative to post hoc model modification in confirmatory factor analysis: The Bayesian lasso. *Psychological Methods*, *22*(4), 687–704. https://doi.org/10.1037/met0000112.

Park, T., & Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, *103*(482), 681–686. https://doi.org/10.1198/016214508000000337.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, *6*, 7–11.

R Development Core Team. (2010). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement*, *9*(4), 401–412. https://doi.org/10.1177/014662168500900409.

Reckase, M. D. (2009). *Multidimensional item response theory*. New York, NY: Springer. https://doi.org/10.1007/978-0-387-89976-3.

Sun, J., Chen, Y., Liu, J., Ying, Z., & Xin, T. (2016). Latent variable selection for multidimensional item response theory models via L_1 regularization. *Psychometrika*, *81*(4), 921–939. https://doi.org/10.1007/s11336-016-9529-6.

Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association*, *82*(398), 528–540. https://doi.org/10.1080/01621459.1987.10478458.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B. Methodological*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

Wang, H. (2012). Bayesian graphical lasso models and efficient posterior computation. *Bayesian Analysis*, *7*(4), 867–886. https://doi.org/10.1214/12-BA729.

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, *12*(1), 58–79. https://doi.org/10.1037/1082-989X.12.1.58.

Yuan, M., & Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika*, *94*(1), 19–35. https://doi.org/10.1093/biomet/asm018.

Zhang, S., Chen, Y., & Li, X. (2019). mirtjml: Joint maximum likelihood estimation for high-dimensional item factor analysis. Retrieved January 20, 2020, from https://cran.r-project.org/web/packages/mirtjml/index.html.

Manuscript Received: 29 JUL 2019
Final Version Received: 12 AUG 2020
Published Online Date: 26 SEP 2020