Psychometrika (2026), **91**, 335–359
doi:10.1017/psy.2025.10068

PSYCHOMETRIC SOCIETY

APPLICATION AND CASE STUDIES - ORIGINAL

Regularized Joint Maximum Likelihood Estimation of Latent Space Item Response Models

Dylan Molenaar[1] and Minjeong Jeon[2]

[1]Department of Psychology, University of Amsterdam, The Netherlands; [2]Department of Education, University of California, Los Angeles, USA
Corresponding author: Dylan Molenaar; Email: D.Molenaar@uva.nl

(Received 4 June 2025; revised 5 November 2025; accepted 21 November 2025)

Abstract

In latent space item response models (LSIRMs), subjects and items are embedded in a low-dimensional Euclidean latent space. As such, interactions among persons and/or items can be revealed that are unmodeled in conventional item response theory models. Current estimation approach for LSIRMs is a fully Bayesian procedure with Markov Chain Monte Carlo, which is, while practical, computationally challenging, hampering applied researchers to use the models in a wide range of settings. Therefore, we propose an LSIRM based on two variants of regularized joint maximum likelihood (JML) estimation: penalized JML and constrained JML. Owing to the absence of integrals in the likelihood, the JML methods allow for various models to be fit in limited amount of time. This computational speed facilitates a practical extension of LSIRMs to ordinal data, and the possibility to select the dimensionality of the latent space using cross-validation. In this study, we derive the two JML approaches and address different issues that arise when using maximum likelihood to estimate the LSIRM. We present a simulation study demonstrating acceptable parameter recovery and adequate performance of the cross-validation procedure. In addition, we estimate different binary and ordinal LSIRMs on real datasets pertaining to deductive reasoning and personality. All methods are implemented in R package 'LSMjml' which is available from CRAN.

Keywords: cross-validation; item response theory; latent space modeling; maximum likelihood

The latent space item response model (LSIRM) integrates latent space models from social network analysis (Hoff et al., 2002) with item response theory (IRT) from psychometrics, extending traditional IRT by embedding persons and items in a metric, multidimensional latent space. As such, LSIRMs may reveal item–person interactions that generally remain unnoticed in conventional models giving more insights about residual dependencies between persons, between items, and between items and persons. This has been shown valuable in the substantive fields of intelligence (Kang & Jeon, 2024; Kim et al., 2014), developmental psychology (Go et al., 2022), mental health (Jeon & Schweinberger, 2024), social influence (Park et al., 2023), national school policy (Jin et al., 2022), and student monitoring (Jeon et al., 2021). In addition, extensions of the LSIRM have enabled the analysis of multilevel structured data (Jin et al., 2022), longitudinal data (Jeon & Schweinberger, 2024; Park et al., 2022), and response time data (Jin et al. 2023; Kang & Jeon, 2024).

Estimation of LSIRMs has been dominated by the fully Bayesian, Markov Chain Monte Carlo (MCMC) estimation scheme by Jeon et al. (2021), which has been implemented in R (Go et al., 2023), JAGS, Stan, NIMBLE (Luo et al., 2023), and Shiny (Ho & Jeon, 2023). Although valuable due to its

© The Author(s), 2026. Published by Cambridge University Press on behalf of Psychometric Society.
This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/psy.2025.10068 Published online by Cambridge University Press

flexibility and its facilities for posterior diagnostics, the MCMC routines are numerically demanding, which may hamper (applied) researchers from using the LSIRM in a wide range of settings that may involve large-scale data. In addition, due to the numerical demanding nature of the MCMC approach, model fit comparison is relatively challenging as it involves multiple models to be fit to the data. Although there are alternatives like leave-one-out cross-validation using Pareto smoothing (Vehtari et al., 2017), these will still be computationally intensive for models like the LSIRM. Currently, researchers rely on spike and slab priors for model selection (e.g., George & McCulloch, 1997; Ishwaran & Rao, 2005) to compare a given LSIRM with R dimensions to a baseline model with $R = 0$. This approach is feasible in the MCMC framework as it does not increase the computational burden significantly. However, tools to compare multiple competing models that differ in R have not yet been developed.

Therefore, in this study, we propose different joint maximum likelihood (JML)-based approaches to fit various LSIRMs to data in a fast and efficient way, facilitating large-scale model application and model selection. In the early years of IRT, JML (Birnbaum, 1968; Lord, 1980; Mislevy & Stocking, 1989) was one of the dominant approaches to fit conventional IRT models to data using software packages such as LOGIST (Wingersky, 1983) and BICAL (Wright & Mead, 1976). As computers were not as fast as nowadays, a desirably practical property of JML was its numerical efficiency. That is, in JML, all parameters are assumed to be fixed effects so that the likelihood function does not include any integrals. These integrals make approaches such as MCMC and marginal maximum likelihood (MML; Bock & Aitkin, 1981) relatively time-consuming, as they require numerical approximation due to the lack of a closed form solution. However, over the years, popularity of JML decreased in favor of MML up until recently, when JML was revived in IRT by the work of Chen et al. (2019, 2020) and Bergner et al. (2022). In this work, the authors developed variations of JML that are suitable for estimation of high-dimensional IRT models on large datasets, which is—even with today's computers—still challenging for the state-of-the art MCMC and MML approaches. In addition to IRT, some JML approaches have been developed for latent space models for social network analysis. That is, Zhang et al. (2022) and Ma et al. (2020) focused on a latent space model with high-dimensional covariates and JML estimation of its parameters. In the approach by Zhang et al., the covariates enter the model via a generalized linear latent variable model, whereas in Ma et al., these covariates are included as predictors next to the latent space positions.

As the IRT models estimated by JML are typically high-dimensional, the complexity of these models is commonly managed through regularization. Regularization is a technique originating from ridge regression (Hoerl & Kennard, 1970) in which the regression parameters are pushed to zero to prevent overfitting and to stabilize parameter estimates in the case of multicollinearity. Currently, regularization includes a variety of techniques, such as penalization and constraints on the parameter space, to promote parameter shrinkage (Hastie et al., 2009; Tibshirani, 1996) or to improve finite-sample performance (Firth, 1993). For example, Chen et al. (2019) used constraints on the parameter space to estimate a multidimensional exploratory IRT model, and Bergner et al. (2022) used L_2 penalization of the parameter space of a general family of IRT models for collaborative filtering.

In this study, we apply these two regularization strategies, constraining the parameter space and penalizing regions of the parameter space, to provide an efficient and stable JML estimation algorithm for the relatively complex LSIRM model. As the effects of regularization are comparable to the effects of parameter priors on the likelihood function, we will show that our JML-based LSIRM is a special cases of the existing MCMC-based LSIRM, providing a highly comparable but easier to estimate variant of the LSIRM. As MCMC obviously has other practical advantages over JML (e.g., flexibility and full posterior information), we present our JML approach as an extension of the current LSIRM modeling toolbox, not as an alternative.

The models by Zhang et al. (2022) and Ma et al. (2020) discussed above are related but different from our approach. That is, both Zhang et al. and Ma et al. focused on a unipartite (only modeling one set of nodes, e.g., persons but not items) latent space model without IRT component, and an inner product distance measure. Our model, however, is bipartite, includes an IRT model component, and uses the Euclidean distance. All three of these aspects make our study to face very different challenges

https://doi.org/10.1017/psy.2025.10068 Published online by Cambridge University Press

in the development of the JML approach. However, besides differences in the underlying model, our estimation approach and that of Zhang et al. and Ma et al. are similar in spirit.

One of the key issues that arises in implementing a maximum likelihood scheme for the LSIRM is that the model can only be identified up to a rotation of the latent space. In the MCMC framework by Jeon et al. (2021), researchers addressed this problem by post-processing the MCMC chains using Procrustes matching (Gower 1975; Sibson, 1979), which involves rotating the latent space to the space from the MCMC iteration with the highest likelihood. Such an approach is infeasible in a maximum likelihood framework. We therefore propose alternative constraints based on the echelon structure from exploratory factor analysis (Dolan et al., 2009; McDonald, 2013). As these constraints are specified a priori, this approach has the advantage that the orientation on which the results are produced is explicitly defined, which facilitates comparisons to other methods.

Thus, using JML, the LSIRM can be estimated in a computationally fast and efficient way, allowing researchers to fit different LSIRM models in a limited amount of time. A resulting advantage that we demonstrate in this study is that the selection of the dimensionality of the latent space can be informed by a K-fold cross-validation routine (e.g., Bergner et al., 2022; Haslbeck & van Bork, 2024). As discussed above, it is currently not possible to compare models differing in the dimensions of the latent space. This possibility, therefore, seems a valuable addition to the toolbox of researchers interesting in LSIRM modeling. In addition, we demonstrate how it is straightforward to use our approach to fit LSIRMs to ordinal data in a limited amount of time. Ordinal LSIRM have recently been development in an MCMC framework (see De Carolis et al., 2025) but are time-consuming to fit.

The outline of this article is as follows: We first present the LSIRM and discuss MCMC estimation of the model parameters. Next, we present the two JML variants and discuss the constraints needed to solve the rotational indeterminacy of the latent space. Then, we outline the methods for parameter estimation, a generalization to ordinal data, and a cross-validation approach to model selection. In the simulation study, we demonstrate that the accuracy of the parameter recovery of our JML approach is comparable to that of the MCMC approach and that the cross-validation approach successfully selects the correct model in most cases. Next, in two illustrations, we apply a binary LSIRM to a dataset on deductive reasoning and an ordinal LSIRM to a dataset on personality. We end with a general discussion.

1. Latent space item response models

The LSIRM is a statistical model for the dichotomous item responses $X_{pi} \in \{0, \dots C\}$ of person $p = 1, \dots, N$ on item $i = 1, \dots, n$. It is assumed that, after accounting for the main effect of the person by person intercept θ_p, and for the main effect of the item by item intercept β_i, the person and item residuals can be embedded in an R-dimensional Euclidean latent space using the R-dimensional vector of person coordinates $\boldsymbol{z}_p \in \mathbb{R}^R$ and the R-dimensional vector of item coordinates $\boldsymbol{w}_i \in \mathbb{R}^R$. As a result, the conditional probability, $P(X_{pi} = c|\theta_p, \beta_i, \boldsymbol{z}_p, \boldsymbol{w}_i)$, is given by

$$P(X_{pi} = 1|\theta_p, \beta_i, \gamma, \mathbf{z}_p, \mathbf{w}_i) = \omega(\theta_p + \beta_i - \gamma d(\mathbf{z}_p, \mathbf{w}_i)), \tag{1}$$

where $\omega(.)$ is a logistic function, γ is the strictly positive weight parameter, and $d(.)$ is a distance function. Even though $d(.)$ can be any distance function that obeys to the mathematical principles of reflexivity, symmetry, and triangular inequality (e.g., Chebyshev distance, Minkowski distance, and Manhattan distance), LSIRMs have generally been applied using the Euclidian distance, that is,

$$d(\mathbf{z}_p, \mathbf{w}_i) = \|\mathbf{z}_p - \mathbf{w}_i\|_2 = \sqrt{\sum_{r=1}^{R} (z_{pr} - w_{ir})^2}, \tag{2}$$

where z_{pr} and w_{ir} are the rth element of $\mathbf{z}_p$ and $\mathbf{w}_i$, respectively. In the present framework, $d(.)$ is a distance function; however, it can be any other function that models the relation of the latent space coordinates. For instance, Zhang et al. (2022) and Ma et al. (2020) used an inner product similarity

https://doi.org/10.1017/psy.2025.10068 Published online by Cambridge University Press

measure for $d(.)$, which would give $d(\mathbf{z}_p, \mathbf{w}_i) = \mathbf{z}_p^T \mathbf{w}_i$. The advantage of the inner product is that it is of low rank, resulting in tractable theoretical properties of the statistical model. However, the inner product cannot be interpreted as a distance that is undesirable for the present aim. See Jeon et al. (2021) for a discussion of the interpretational challenges of the inner product measure in LSIRMs.

The LSIRM above can be interpreted as a Rasch model (Rasch, 1960) in which the residuals are further modeled using $\mathbf{z}_p$ and $\mathbf{w}_i$. As a result, the LSIRM can give more insight in local dependencies of the item response data due to interactions between items, interactions between persons, or interactions between items and persons. For instance, such interactions may arise because of some items requiring a more similar response processes than others, some persons sharing a relevant background characteristic more than others (e.g., educational attainment and motivation), and some persons use a different solution strategy on some of the items than others.

1.1. Estimation

MCMC sampling methods have been developed to fit the LSIRM (see, e.g., Jeon et al., 2021). By specifying prior distributions for the model parameters presented above, it is possible to estimate an additional nonnegative parameter σ_θ^2, the variance of the person intercept. In addition, $\log(\gamma) = \gamma'$ is modeled to ensure that γ is strictly positive. Next, if we collect all θ_p parameters in the N-dimensional vector $\boldsymbol{\theta}$ and all β_i parameters in in the n-dimensional vector $\boldsymbol{\beta}$, and if we denote the full $N \times R$ matrix of stacked $\mathbf{z}_p^{\mathrm{T}}$ vectors by $\mathbf{Z}$, and the full $n \times R$ matrix of stacked $\mathbf{w}_i^{\mathrm{T}}$ vectors by $\mathbf{W}$, the posterior distribution of the model parameters is given by

$$f\left(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{W}, \mathbf{Z}, \gamma', \sigma^2_{theta} | \mathbf{X}\right) \propto \left[\prod_{p=1}^{N}\prod_{i=1}^{n} f_X\left(X_{pi} | \boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{W}, \mathbf{Z}, \gamma'\right)\right] \times \left[\prod_{p=1}^{N} f_\theta\left(\theta_p | \sigma_\theta^2\right) f_z\left(\mathbf{z}_p\right)\right] \times \left[\prod_{i=1}^{n} f_w\left(\mathbf{w}_i\right) f_\beta\left(\beta_i\right)\right] \times f_{\sigma^2}\left(\sigma_\theta^2\right) \times f_{\gamma'}\left(\gamma'\right), \tag{3}$$

where $f_X(.)$ is the Bernoulli distribution with success probability given by Equation (1), and where $f_\theta(.)$, $f_z(.)$, $f_w(.)$, $f_\beta(.)$, $f_{\sigma^2}(.)$, and $f_{\gamma'}(.)$ are (hyper) prior distributions for, respectively, θ_p, $\mathbf{z}_p$, $\mathbf{w}_i$, β_i, σ_θ^2, and γ'. In general, the following distributions are used (e.g., Go et al., 2023; Jeon et al., 2021): a univariate normal for θ_p with mean 0 and variance σ_θ^2 (as discussed above), a univariate normal for β_i with mean 0 and variance σ_β^2, a multivariate orthogonal standard normal distributions for $\mathbf{z}_p$ and $\mathbf{w}_i$, an inverse-gamma distribution for σ_θ^2 with shape parameter a_σ and scale parameter b_σ, and a normal distribution for γ' with mean μ_γ and variance σ_γ^2.

1.2. Identification

The item and person intercept parameters θ_p and β_i are readily identified by fixing the prior mean of θ_p to 0 in the above. For $\mathbf{z}_p$ and $\mathbf{w}_i$, fixing the R dimensions to be orthogonal with zero mean and unit variance ensures that the locations of the $\mathbf{z}_p$ and $\mathbf{w}_i$ coordinates are uniquely identified. However, there still exists a rotational indeterminacy. That is, any two of the R dimensions of $\mathbf{z}_p$ and $\mathbf{w}_i$ can be rotated by an arbitrary angle α, resulting in a different solution with the same Euclidean distances $\|\mathbf{z}_p - \mathbf{w}_i\|_2$ and the same data likelihood. In the MCMC framework to fit the LSIRM, this indeterminacy causes each sample from the posterior parameter distribution to be potentially subject to a different rotation. Posterior sample means are therefore confounded and cannot be used as parameter estimates. To solve this issue, researchers rely on Procrustes matching (Gower, 1975; Sibson, 1979) in which each posterior sample of $\mathbf{Z}$ and $\mathbf{W}$ is transformed to match a given target, respectively, $\mathbf{T}^{(z)}$ and $\mathbf{T}^{(w)}$. In general, these target matrices are chosen to be the $\mathbf{z}_p$ and $\mathbf{w}_i$ samples from the MCMC iteration with the largest likelihood (Jeon et al., 2021).

https://doi.org/10.1017/psy.2025.10068 Published online by Cambridge University Press

1.3. Model selection

For model selection, Jeon et al. (2021) proposed the use of a spike and slab prior for γ'. Specifically, instead of a normal prior, $f_{\gamma'}(.)$ is specified as a mixture of two normal distributions

$$f_{\gamma'}(.) = (1-\delta) \times f_{\gamma'}^{(0)}\left(\gamma'\right) + \delta \times f_{\gamma'}^{(1)}\left(\gamma'\right),$$

where $\delta \in \{0,1\}$ is a dichotomous parameter, $f_{\gamma'}^{(0)}(.)$ is the spike prior with mean μ_0 and variance σ_0^2, and $f_{\gamma'}^{(1)}(.)$ is the slab prior with mean μ_1 and variance σ_1^2. The spike prior mean and variance should be chosen so that the resulting distribution of untransformed γ has a mean close to 0 and a small variance, while the slab prior mean and variance should reflect the more conventional prior mean and variance. Next, by assuming δ to be Bernoulli distributed with success parameter π_δ, and by specifying a Beta prior for π_δ, parameter δ can be sampled along with the other parameters in the model. Model selection then involves the posterior probability that δ is equal to 1. If this probability is large (commonly a cutoff of 0.5 is used), it is concluded that the LSIRM with the R under consideration accounts better for the data than an IRT model without latent space (i.e., a one-parameter logistic model).

2. Joint maximum likelihood estimation

As discussed, the MCMC estimation scheme above can be time-consuming as sufficient samples need to be drawn from the posterior parameter distribution. To have a fast alternative available, below we present a JML estimation approach to estimate the parameters from the LSIRM. Implementing the LSIRM in a maximum likelihood framework brings specific identification challenges, which we address below. In addition, we discuss a model selection procedure based on K-fold cross-validation that is impractical or infeasible for the more time-consuming estimation algorithms, and we present a generalization to ordinal data.

In maximum likelihood estimation, the parameters of a statistical model are estimated by maximizing the likelihood of the data for the unknown model parameters. For LSIRM, this would involve the joint log-likelihood function, $\log\left(f_X\left(X_{pi}|\boldsymbol{\theta},\boldsymbol{\beta},\mathbf{W},\mathbf{Z},\boldsymbol{\gamma}'\right)\right)$ with $f_X(.)$ from Equation (3), summed over items and persons. However, without further constraints, this model is unidentified. First of all, in a JML framework, all parameters are fixed effects so that σ_θ^2 is not a model parameter as it is absorbed in the θ_p estimates. For the same reason, γ is not estimable as it is equivalent to the standard deviation of z_{pk} and w_{ik}. Thus, γ is absorbed in the z_{pk} and w_{ik} estimates. Therefore, the JML-based LSIRM is given by

$$P\left(X_{pi}=1|\theta_p,\beta_i,\mathbf{z}_p,\mathbf{w}_i\right) = \omega\left(\theta_p+\beta_i-d\left(\mathbf{z}_p,\mathbf{w}_i\right)\right), \tag{4}$$

where $d(.)$ is the Euclidean distance from Equation (2).

As discussed above, to introduce further sparsity into the model, we consider two different regularization approaches based on either penalizing the joint likelihood or on constraining the maximum norm of the parameter vectors in the joint likelihood. As both regularization effects can be conceived as the effects of prior distributions, we intend to enhance the comparability of the JML results and the MCMC results. Both JML variants are discussed below.

2.1. Penalized joint maximum likelihood

In the first approach referred to as penalized joint maximum likelihood (pJML), we use an L_2 regularization penalty in the likelihood function. Although there are other options possible, we use the L_2 due to its correspondence to a normal prior. The log-likelihood function we consider is

$$\begin{aligned}\ell\left(\boldsymbol{\theta},\boldsymbol{\beta},\mathbf{W},\mathbf{Z},|\mathbf{X}\right) = \sum_{p=1}^{N}\sum_{i=1}^{n}\log\left(f_X\left(X_{pi}|\boldsymbol{\theta},\boldsymbol{\beta},\mathbf{W},\mathbf{Z}\right)\right) - \\ \frac{1}{2}\lambda\left(\sum_{p=1}^{N}\theta_p^2+\sum_{i=1}^{n}\beta_i^2+\sum_{i=1}^{n}\sum_{k=1}^{K}w_{ik}^2+\sum_{p=1}^{N}\sum_{k=1}^{K}z_{pk}^2\right),\end{aligned} \tag{5}$$

https://doi.org/10.1017/psy.2025.10068 Published online by Cambridge University Press

where λ is a penalty parameter and $f_X(.)$ is a Bernoulli distribution with success probability given by Equation (4). Although, in the context of predictive accuracy, λ can be optimized using cross-validation, here we take λ to be a prespecified parameter reflecting the precision of a zero-centered normal prior distribution. In the remainder of this article, we use $\lambda = 1$ so that the regularization effect on $\theta_p, \beta_i, \mathbf{z}_p$, and $\mathbf{w}_i$ is comparable to that of standard normal priors typically used in MCMC estimation of the LSIRM. For instance, in the R package *lsirm12pl* (Go et al., 2023), which we use to compare our results to in the simulation study below, standard normal priors are used by default for β_i, $\mathbf{z}_p$, and $\mathbf{w}_i$ (θ_p has a normal prior with variance σ_θ^2 as discussed above).

The resulting approach can be seen as a variant of the Bayesian joint modal estimation approach used by Swaminathan and Gifford (1982, 1985, 1986) to fit one-, two-, and three-parameter logistic IRT models in a Bayesian framework. That is, Equation (5) is proportional to the log-posterior distribution of the parameters by which its maximum gives joint modal estimates of the parameters (i.e., MAP estimates). Moreover, the pJML approach in Equation (5) is a special case of the MCMC-based LSIRM in Equation (3). That is, fixing $\gamma' = 0$, $\sigma_\theta^2 = 1, and\ \sigma_\beta^2 = 1$ in the logarithm of Equation (3), the MCMC-based LSIRM is equivalent to the pJML model in Equation (5) with $\lambda = 1$. The key difference that arises in practice is that MCMC focusses on posterior means instead of posterior modes by which the pJML model can be estimated much faster.

2.2. Constrained joint maximum likelihood

A next option referred to as constrained joint maximum likelihood (cJML) is to regularize the LSIRM model in Equation (4) by constraining the maximum norm of the person and item parameter vectors to some prespecified value. This approach has been proposed by Chen et al. (2019) for JML estimation of the multidimensional two-parameter logistic IRT model. Thus, we constrain the vector of person parameters $\boldsymbol{\tau}_{1p} = [\theta_p, \mathbf{z}_p]$ and the vector of item parameters $\boldsymbol{\tau}_{2i} = [\beta_p, \mathbf{w}_p]$ in the following way:

$$\left\|\boldsymbol{\tau}_{1p}\right\|_2 \leq C_1 \text{ and } \left\|\boldsymbol{\tau}_{2i}\right\|_2 \leq C_2, \tag{6}$$

where C_1 and C_2 are the prespecified maximum parameter norms for the person and item parameter vectors, respectively, and $\|.\|_2$ is the Euclidean norm. Thus, if $\mathbf{T}_1$ denotes a matrix of stacked $\boldsymbol{\tau}_{1p}$ vectors and $\mathbf{T}_2$ denotes a matrix of stacked $\boldsymbol{\tau}_{2i}$ vectors, the likelihood function in this approach is given by

$$\ell(\mathbf{T}_1, \mathbf{T}_2|\mathbf{X}) = \sum_{p=1}^{N}\sum_{i=1}^{n} \log\left(f_X\left(X_{pi}|\boldsymbol{\tau}_{1\mathrm{p}}, \boldsymbol{\tau}_{2\mathrm{i}}\right)\right) \text{with } \boldsymbol{\tau}_{1\mathrm{p}}, \boldsymbol{\tau}_{2\mathrm{i}} \text{ s.t. Equation } (6). \tag{7}$$

For a D-dimensional two-parameter logistic IRT model, Chen et al. (2019) proposed $C_1 = C_2 = 5\sqrt{D}$. As both $\boldsymbol{\tau}_{1p}$ and $\boldsymbol{\tau}_{2i}$ have $R+1$ dimensions in our LSIRM model, we replace D by $R+1$. In addition, we use $C_2 = 3\sqrt{R+1}$ for the item parameters and $C_1 = \frac{1}{2}C_2$ for the person parameters. Below, we show that these constraints are still less restrictive compared with the MCMC and pJML approaches.

The cJML model in Equation (7) can similarly be conceived as a joint modal estimation approach. To this end, the constraints in Equation (6) are enforced by using priors $f_{\boldsymbol{\tau}_1}(.)$ and $f_{\boldsymbol{\tau}_2}(.)$ on $\boldsymbol{\tau}_{1p}$ and $\boldsymbol{\tau}_{2i}$, respectively. These priors are uniform on an $(R+1)$-dimensional Euclidean ball with radius C_1 for the person parameters and C_2 for the item parameters, that is,

$$f_{\boldsymbol{\tau}_1}(\boldsymbol{\tau}_{1p}) \propto I\left(\left\|\left|\boldsymbol{\tau}_{1p}\right|\right\|_2 < C_1\right) \tag{8}$$

and

$$f_{\boldsymbol{\tau}_2}(\boldsymbol{\tau}_{2i}) \propto I\left(\left\|\left|\boldsymbol{\tau}_{2i}\right|\right\|_2 < C_2\right), \tag{9}$$

where $I(.)$ is the indicator function. Although, in this study, we rely on Equation (7), the above joint modal formulation of the cJML model is equivalent and explicates the difference with the pJML and MCMC approaches, both of which use normal priors. Unlike the pJML, the cJML-based LSIRM model is not strictly a special case of the MCMC-based LSIRM with standard normal priors discussed above.

https://doi.org/10.1017/psy.2025.10068 Published online by Cambridge University Press

However, it is a special case in the more general framework of Equation (3) with $\gamma' = 0$ and θ_p and $\mathbf{z}_p$ following the prior in Equation (8) and β_i and $\mathbf{w}_i$ following the prior in Equation (9).

To see that the cJML approach is the least restrictive approach in terms of parameter regularization, we compare the prior variance across the three approaches. For the MCMC and pJML approaches, all priors have variance 1 (except for θ in the MCMC approach). The variances of the priors in Equations (8) and (9) are $\frac{C_1^2}{R+3}$ for the person parameters and $\frac{C_2^2}{R+3}$ for the item parameters. Given our choices for C_1 and C_2 above, this results in a prior variance of 4.5, 5.4, and 6 for the item parameters and of 1.125, 1.35, and 1.5 for the person parameters for $R = 1$, $R = 2$, and $R = 3$, respectively. These priors are, thus, less informative in terms of prior variance than those of the MCMC and pJML approaches (although differences are small for the person parameters). This will also be illustrated in the simulation study.

2.3. Consistency

Traditional unregularized JML is known to be theoretically inconsistent (Andersen, 1973; Haberman, 1977) as asymptotic theory is violated because of the number of parameters increasing with N. In this study, it is expected that, due to the known correspondence of our L_2 penalization to the effects of normal parameter priors, the finite-sample properties of the JML procedure will improve. However, we are unaware of theoretical proofs of consistency of pJML in the literature, so whether penalized JML is strictly consistent in the asymptotic sense has yet to be established.

For the cJML approach described above, it is known that the procedure is theoretically consistent under a double-asymptotic regime for structured generalized latent factor models (Chen et al., 2020) and multidimensional unstructured two-parameter IRT models (see Chen et al., 2019; Chen & Li, 2024). Similarly, approaches related to cJML are shown to be consistent for unipartite latent space models with low-rank inner-product distances (Ma et al., 2020; Zhang et al., 2022). However, these results do not necessarily apply to the present model due to the unstructured nature of the latent space, the nonlinear and high-rank character of the natural parameter arising from Euclidean distances, and the bipartite nature of our item–respondent interactions. As the JML consistency issues are known to decrease with an increasing number of items (Haberman, 1977), and our simulation study below shows that our JML approaches perform comparable to the existing Bayesian approaches, we believe that any inconsistency (if any) is relatively unproblematic given the aim of the present study.

2.4. Rotation

As discussed above, the distances in Equation (2) are subject to a rotational indeterminacy and can be arbitrarily rotated to produce the same distances. For instance, for $R = 2$,

$$\mathbf{W}' = \mathbf{W}\mathbf{R}^{\mathbf{T}} \text{ and } \mathbf{Z}' = \mathbf{Z}\mathbf{R}^{\mathbf{T}}$$

with

$$\boldsymbol{R} = \begin{bmatrix} \cos r & -\sin r \\ \sin r & \cos r \end{bmatrix}$$

will result in the same Euclidean distances $d(\mathbf{z}_p, \mathbf{w}_i) = d\left(\mathbf{z}'_p, \mathbf{w}'_i\right)$ for all p and all i for any arbitrary angle of rotation $r \in [-2\pi, 2\pi]$. As discussed above, this problem is solved in an MCMC framework by Procrustes matching in which the samples from the posterior are all transformed to a target solution. As our approaches do not involve posterior samples, we will use specific constraints on $\mathbf{W}$ to fix $\mathbf{W}$ and $\mathbf{Z}$ to an arbitrary rotation without affecting the data likelihood.

Specifically, the problem of rotational indeterminacy of $\mathbf{W}$ and $\mathbf{Z}$ is similar to the problem of rotational indeterminacy of the factor loadings and the factor scores in factor analysis (e.g., Jennrich, 1978). Therefore, we will rely on what is referred to as "echelon rotation" in the factor analysis literature (Dolan et al., 2009; McDonald, 2013). That is, we will fix the $\frac{R(R-1)}{2}$ elements of the upper triangle of

https://doi.org/10.1017/psy.2025.10068 Published online by Cambridge University Press

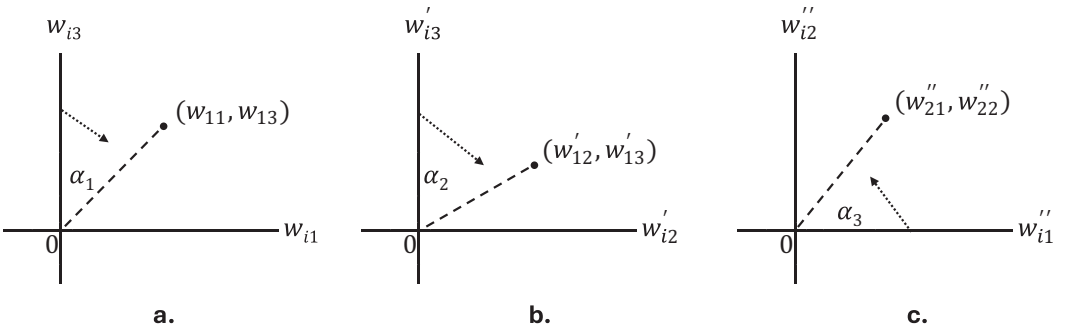


Figure 1. Graphical illustration of the echelon rotation for $R = 3$. The dotted arrows indicate the direction of the rotation, the solid dot denotes a specific coordinate (x,y), and the striped lines give an indication of the new (i.e., rotated) position of the axis connected to the dotted arrow. The angle of rotation is indicated by α_1, α_2, and α_3.

submatrix $\mathbf{W}_{1:\mathrm{R}-1,1:\mathrm{R}-1}$ to 0. For instance, for $R = 3$,

$$\mathbf{W} = \begin{bmatrix} 0 & 0 & w_{13} \\ w_{21} & 0 & w_{23} \\ w_{31} & w_{32} & w_{33} \\ \vdots & \vdots & \vdots \\ w_{n1} & w_{n2} & w_{n3} \end{bmatrix}. \tag{10}$$

An advantage of using this constraint compared with Procrustes matching is that the solution is more explicitly defined and comparable across different methods that fit the LSIRM. As for Procrustes matching, the solution depends on the largest likelihood encountered during MCMC sampling, the solution may depend on the sampling scheme, estimation algorithm, or even on random fluctuations within the same method applied to the same dataset. This is unproblematic for a single application as any rotation solution is arbitrary. However, if applications or algorithms need to be compared, using Procrustes matching may confound the comparison. Using echelon rotation, any solution obtained with any estimation algorithm (MCMC or maximum likelihood, but also potential other procedures such as least squares, variational inference, and minorization) can be rotated to the structure above to facilitate comparison.

For $R = 3$, the following three rotations are carried out to transform $\mathbf{W}$ to an echelon structure:

$$\mathbf{W}' = \mathbf{W}\mathbf{R}_1^{\mathrm{T}}, \mathbf{W}'' = \mathbf{W}'\mathbf{R}_2^{\mathrm{T}}, \text{ and } \mathbf{W}''' = \mathbf{W}''\mathbf{R}_3^{\mathrm{T}}, \tag{11}$$

where $\mathbf{R}_1$, $\mathbf{R}_2$, and $\mathbf{R}_3$ are given by

$$\mathbf{R}_1 = \begin{bmatrix} \cos(\alpha_1) & 0 & \sin(\alpha_1) \\ 0 & 1 & 0 \\ -\sin(\alpha_1) & 0 & \cos(\alpha_1) \end{bmatrix}, \mathbf{R}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha_2) & \sin(\alpha_2) \\ 0 & -\sin(\alpha_2) & \cos(\alpha_2) \end{bmatrix}, \text{ and } \mathbf{R}_3 = \begin{bmatrix} \cos(\alpha_3) & \sin(\alpha_3) & 0 \\ -\sin(\alpha_3) & \cos(\alpha_3) & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{12}$$

The angles of rotation α_1, α_2, and α_3 can be obtained by standard trigonometry and are given by

$$\alpha_1 = \arctan\left(\frac{w_{11}}{w_{13}}\right), \alpha_2 = \arctan\left(\frac{w'_{12}}{w'_{13}}\right), \text{ and } \alpha_3 = -\arctan\left(\frac{w''_{22}}{w''_{21}}\right), \tag{13}$$

where w'_{ik} and w''_{ik} are the elements from the rotated $\mathbf{W}'$ and $\mathbf{W}''$ matrices. See Figure 1 for a graphical illustration of the rotation scheme.

That is, in the figure, it can be seen that, first, w_{i1} and w_{i3} are rotated around w_{i2} using angle α_1 so that $w'_{11} = 0$. Next, w'_{i2} and w'_{i3} are rotated around w'_{i1} using angle α_2 so that $w''_{12} = 0$. Finally, w''_{i1} and

https://doi.org/10.1017/psy.2025.10068 Published online by Cambridge University Press

w''_{i2} are rotated around w''_{i3} using angle α_3 so that $w'''_{22} = 0$. Final matrix $\mathbf{W}'''$ will have the structure in Equation (10). If $\mathbf{R}_0 = \mathbf{R_3}\,(\mathbf{R}_2\mathbf{R}_1)$ is denoted to be the overall rotation matrix, then the same rotation can be applied to $\mathbf{Z}$, resulting in $\mathbf{Z}''' = \mathbf{Z}\mathbf{R}_0^{\mathrm{T}}$. As noted before, the Euclidean distances of the original parameters from $\mathbf{W}$ and $\mathbf{Z}$ are equivalent to the Euclidean distances of the rotated parameters from $\mathbf{W}'''$ and $\mathbf{Z}'''$, that is, $d\,(\mathbf{z}_p, \mathbf{w}_i) = d\left(\mathbf{z}'''_p, \mathbf{w}''''_i\right)$ for all p and all i. Generalization to any R is straightforward and can be practically performed using the algorithm by, for instance, Wansbeek and Meijer (2000).

2.5. Model selection

Owing to its diagnostic character, in many situations, an LSIRM with $R = 2$ will suffice because of its suitability to visualize the results in $2D$ plots. However, if a more statistical informed decision needs to be made about R, there are no methods available yet. In a maximum likelihood framework, direct comparison of the models using common fit indices, such as Akaike information criterion and Bayesian information criterion, is challenging because of the ambiguity about the definition of the parameter penalty for models with different dimensions of $\mathbf{W}$ and $\mathbf{Z}$. As we rely on JML estimation, a suitable fit index may be provided by the Joint-likelihood-based information criteria (JIC) by Chen and Li (2022). However, in simulations, it turned out that the JIC only works for our model if the ratio n/N is larger than considered here (see below and Appendix B of the Supplementary Material).

Therefore, we propose a model selection procedure based on K-fold cross-validation. Similar procedures have been proposed for model selection among multidimensional two-parameter logistic models (Bergner et al., 2022) and network models, including latent space models (Li et al., 2020). Specifically, $\mathbf{X}^{(\mathrm{k})}$ denotes the $N \times n$ matrix of data in fold $k = 1, \ldots, K$, and Q denotes the number of observed elements in $\mathbf{X}$. Then, the data in fold k, $\mathbf{X}^{(\mathrm{k})}$, are obtained by randomly selecting floor$\left(\frac{Q}{K}\right)$ elements X_{pi} from the full data matrix $\mathbf{X}$ without replacement, and assigning these to elements $X^{(k)}_{pi}$ in $\mathbf{X}^{(k)}$. The elements of $\mathbf{X}^{(k)}$ that have no value from $\mathbf{X}$ assigned are replaced by missing values. As such, $\mathbf{X}$ is partitioned in K subsamples. If Q is not a multiple of K, the elements from $\mathbf{X}$ that are yet unselected are assigned to the first $Q \bmod K$ folds.

Next, the LSIRM is fit K times to the data matrix $\mathbf{X}$, each time leaving out the data from fold k by assuming these data to be Missing Completely at Random (Rubin, 1976). Finally, for each of the K results obtained, we determine the predictive accuracy of the LSIRM in predicting the data from fold K. Specifically, we consider three performance metrics.

2.5.1. Unnormalized classification error

The first metric is what we refer to as the unnormalized classification error (*UCE*):

$$UCE = \sum_{(p,i)\in S^{(k)}} \sum_{k=1}^{K} I\left(X^{(k)}_{pi} \neq \dot{X}^{(k)}_{pi}\right), \tag{14}$$

where $S^{(k)}$ is the set containing all (p,i) combinations for which $X^{(k)}_{pi}$ is observed in $\mathbf{X}^{(\boldsymbol{k})}$ and $\dot{X}^{(k)}_{pi}$ is the model predicted score (0 or 1) for $X^{(k)}_{pi}$, that is,

$$\dot{X}^{(k)}_{pi} = \mathrm{round}\left(P\left(X^{(k)}_{pi} = 1 | \theta^{(-k)}_p, \beta^{(-k)}_i, \mathbf{z}^{(-k)}_p, \mathbf{w}^{(-k)}_i\right)\right), \tag{15}$$

where the parameters contain superscript $(-k)$ to denote that the data in $\mathbf{X}^{(k)}$ were not used in its estimation. Note that *UCE* is inversely related to the classification accuracy, which is generally defined as the overall proportion of scores correctly classified as 0 or 1. We focus on classifications error so that smaller values indicate better-performing models, which is better in line with the residual sum of squares that we introduce later (which is naturally smaller for better-performing models). In addition, while the classification accuracy is commonly normalized, we focus on the unnormalized metric to increase the range of the metric and to be better in line with log-likelihood-based metrics, which are

https://doi.org/10.1017/psy.2025.10068 Published online by Cambridge University Press

typically unnormalized (at least in psychometrics). The disadvantage of using unnormalized metrics is that its exact value is harder to interpret compared with, for example, prediction accuracies. However, prediction accuracies can be very similar across models due to their narrow range and upper bound (which is not necessarily 1; see Bergner et al., 2022), and can therefore also be challenging to interpret.

2.5.2. Unnormalized ROC error

The next fit metric proposed is based on the area under the receiver operator curve (ROC). An ROC gives the relation among the true positive rates and the false positive rates of a model across different thresholds in a classification. For a perfect classification, the area under this curve is equal to 1, indicating that the true positive rate is 1 and the false positive rate is 0. In the present study, we use the area under the ROC curve to see how well the data in a given fold are predicted using the LSIRM as estimated on the remaining folds. As discussed above, we focus on unnormalized metrics for which lower values indicate better model performance; therefore, we focus on the unnormalized ROC error (*URE*):

$$URE = \sum_{k=1}^{K} N^{(k)} n^{(k)} \left[1 - \text{AUC}\left(\mathbf{X}^{(k)}, \mathbf{P}^{(k)}\right)\right], \tag{16}$$

where $AUC(.)$ is the function that determines the area under the ROC curve and $\mathbf{P}^{(k)}$ is an $N \times n$ matrix containing the model predictions for fold k in elements $S^{(k)}$ of that matrix. Similarly to $\mathbf{X}^{(k)}$, the elements of $\mathbf{P}^{(k)}$ that are not in $S^{(k)}$ are set to missing.

2.5.3. Residual sum of squares

Finally, we use a metric based on the residual sum of squares (*RSS*) in the prediction of $\mathbf{X}^{(k)}$ by $\mathbf{P}^{(k)}$. That is,

$$RSS = \sum_{(p,i) \in S^{(r)}} \sum_{k=1}^{K} \left(X_{pi}^{(k)} - P_{pi}^{(k)}\right)^2, \tag{17}$$

where $P_{pi}^{(k)}$ is element (p,i) from matrix $\mathbf{P}^{(k)}$. For the *RSS*, it also holds that lower values indicate a better-performing model.

2.6. Likelihood optimization

We implemented the methods above in the *R* package *LSMjml*, which is available from CRAN and from www.dylanmolenaar.nl. We use a gradient ascent algorithm to maximize the likelihood in Equation (5) for pJML and in Equation (7) for cJML with $\mathbf{W}$ subject to the echelon structure discussed above. For pJML, parameters $\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{Z}$, and $\mathbf{W}$ are updated sequentially by holding the other parameters constant. The likelihood is penalized as shown in Equation (5), with $\lambda = 1$ as discussed above. For cJML, parameter pairs $[\boldsymbol{\beta}, \mathbf{W}]$ and $[\boldsymbol{\theta}, \mathbf{Z}]$ are updated alternatingly. The constraints in Equation (6) are introduced by gradient projection (Chen et al., 2019; Nocedal & Wright, 2006, p. 485) in which the parameter values are transformed to comply to Equation (6) after each parameter vector update (if necessary). As both JML procedures are full information, missing data can be accommodated by assuming MAR and summing the log-likelihood over all available data points.

Starting values are based on a preliminary exploratory item factor analysis with $R + 1$ factors in which the first factor has equal factor loadings for all items and the factor loadings of the remaining R factors follow an echelon structure. Cai (2010) proposes a similar strategy for multidimensional item factor analysis. However, like Bergner et al. (2022), we found that, in general, random initialization performs equally well. See Algorithms 1 and 2 below for a description of, respectively, the pJML and cJML procedures using pseudocode.

https://doi.org/10.1017/psy.2025.10068 Published online by Cambridge University Press

Algorithm 1 – Penalized JML

1. **Input** Matrix of observed binary responses $\mathbf{X}$ or matrix of observed dummy coded ordinal responses $\mathbf{X}'$, R, tolerance, λ.
2. **Starting values** Obtain the starting values $\mathbf{Z}^{(0)}$, $\mathbf{W}^{(0)}$, $\boldsymbol{\theta}^{(0)}$, and $\boldsymbol{\beta}^{(0)}$ for respectively $\mathbf{Z}$, $\mathbf{W}$, $\boldsymbol{\theta}$, and $\boldsymbol{\beta}$ by either 1) randomly sampling from standard normal distributions; or by 2) fitting an $R+1$ constrained factor model and use these estimates as starts (see text).
3. **Iteration t=0:** Determine $\ell\left(\boldsymbol{\theta}^{(0)}, \boldsymbol{\beta}^{(0)}, \mathbf{W}^{(0)}, \mathbf{Z}^{(0)} \middle| \mathbf{X}\right)$ from Equation 5; Determine maximum step size for parameters $\eta^{(\boldsymbol{\beta})}, \eta^{(\boldsymbol{\theta})}, \eta^{(\mathbf{Z})}$, and $\eta^{(\mathbf{W})}$ by a line search.
4. **Iterate $\boldsymbol{t = 1, 2, \ldots}$, until convergence** (difference in $\ell(.)$is smaller than tolerance)**:**

   For $i = 1, 2, \ldots, n$**:**

   Update $\beta_i^{(t)} = \beta_i^{(t-1)} + \eta^{(\boldsymbol{\beta})} \boldsymbol{g}_i^{(\boldsymbol{\beta})}$ where $\boldsymbol{g}^{(\boldsymbol{\beta})} = \frac{\partial \ell\left(\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\beta}^{(t-1)}, \mathbf{W}^{(t-1)}, \mathbf{Z}^{(t-1)} \middle| \mathbf{X}\right)}{\partial \boldsymbol{\beta}^{(t-1)}}$

   For $p = 1, 2, \ldots, N$:

   Update $\theta_p^{(t)} = \theta_p^{(t-1)} + \eta^{(\boldsymbol{\theta})} \boldsymbol{g}_p^{(\boldsymbol{\theta})}$ where $\boldsymbol{g}^{(\boldsymbol{\theta})} = \frac{\partial \ell\left(\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\beta}^{(t)}, \mathbf{W}^{(t-1)}, \mathbf{Z}^{(t-1)} \middle| \mathbf{X}\right)}{\partial \boldsymbol{\theta}^{(t-1)}}$

   For $p = 1, 2, \ldots, N$:

   For $s = 1,2, \ldots R$:

   Update $z_{pr}^{(t)} = z_{pr}^{(t-1)} + \eta^{(\mathbf{Z})} \boldsymbol{G}_{pr}^{(\mathbf{Z})}$ where $\boldsymbol{g}^{(\mathbf{Z})} = \frac{\partial \ell\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\beta}^{(t)}, \mathbf{W}^{(t-1)}, \mathbf{Z}^{(t-1)} \middle| \mathbf{X}\right)}{\partial \mathbf{Z}^{(t-1)}}$

   For $i = 1, 2, \ldots, n$:

   For $r = 1,2, \ldots R$:

   If w_{ir} is not fixed in Echelon structure:

   Update $w_{ir}^{(t)} = w_{ir}^{(t-1)} + \eta^{(\mathbf{W})} \boldsymbol{G}_{ir}^{(\mathbf{W})}$ where $\boldsymbol{g}^{(\mathbf{W})} = \frac{\partial \ell\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\beta}^{(t)}, \mathbf{W}^{(t-1)}, \mathbf{Z}^{(t)} \middle| \mathbf{X}\right)}{\partial \mathbf{W}^{(t-1)}}$

   Determine $\ell\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\beta}^{(t)}, \mathbf{W}^{(t)}, \mathbf{Z}^{(t)} \middle| \mathbf{X}\right)$

   In every update, $\eta^{(.)}$ is determined from a line search starting at the maximum step size established in step 3. For every 100th iteration the maximum step size parameters from step 3 are updated
5. **Output parameter values from final iteration**

https://doi.org/10.1017/psy.2025.10068 Published online by Cambridge University Press

Algorithm 2 – Constrained JML

1. **Input** Matrix of observed binary responses $\mathbf{X}$ or matrix of observed dummy coded ordinal responses $\mathbf{X}'$, R, tolerance, C_1, C_2.
2. **Starting values** See Algorithm 1.
3. **Iteration t=0:** Determine $\ell\left(\boldsymbol{\theta}^{(0)}, \boldsymbol{\beta}^{(0)}, \mathbf{W}^{(0)}, \mathbf{Z}^{(0)} \middle| \mathbf{X}\right)$ from Equation 7; Determine maximum step size for parameters $\eta^{(1)}$, and $\eta^{(2)}$ by a line search.
4. **Iterate $t = 1, 2, \ldots$, until convergence** (difference in $\ell(.)$ is smaller than tolerance)**:**

   For $i = 1, 2, \ldots, n$**:**

   Update $\beta_i^{(t)} = \beta_i^{(t-1)} + \eta^{(1)} g_i^{(\beta)}$ where $\boldsymbol{g}^{(\boldsymbol{\beta})} = \frac{\partial \ell\left(\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\beta}^{(t-1)}, \mathbf{W}^{(t-1)}, \mathbf{Z}^{(t-1)} \middle| \mathbf{X}\right)}{\partial \boldsymbol{\beta}^{(t-1)}}$

   For $r = 1, 2, \ldots R$:

   If w_{ir} is not fixed in Echelon structure:

   Update $w_{ir}^{(t)} = w_{ir}^{(t-1)} + \eta^{(1)} G_{ir}^{(\mathbf{W})}$ where $\boldsymbol{g}^{(\mathbf{W})} = \frac{\partial \ell\left(\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\beta}^{(t-1)}, \mathbf{W}^{(t-1)}, \mathbf{Z}^{(t-1)} \middle| \mathbf{X}\right)}{\partial \mathbf{W}^{(t-1)}}$

   If Euclidean norm of $\tau_{2i} = [\beta_i \ \mathbf{w}_i] > C_2$:

   $\tau'_{2i} = \frac{\tau_{2i}}{\|\tau_{2i}\|_2} \times C_2$ and update β_i and $\mathbf{w}_i$ accordingly

   For $p = 1, 2, \ldots, N$:

   Update $\theta_p^{(t)} = \theta_p^{(t-1)} + \eta^{(2)} g_p^{(\boldsymbol{\theta})}$ where $\boldsymbol{g}^{(\boldsymbol{\theta})} = \frac{\partial \ell\left(\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\beta}^{(t)}, \mathbf{W}^{(t)}, \mathbf{Z}^{(t-1)} \middle| \mathbf{X}\right)}{\partial \boldsymbol{\theta}^{(t-1)}}$

   For $s = 1, 2, \ldots R$:

   Update $z_{pr}^{(t)} = z_{pr}^{(t-1)} + \eta^{(2)} G_{pr}^{(\mathbf{Z})}$ where $\boldsymbol{g}^{(\mathbf{Z})} = \frac{\partial \ell\left(\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\beta}^{(t)}, \mathbf{W}^{(t)}, \mathbf{Z}^{(t-1)} \middle| \mathbf{X}\right)}{\partial \mathbf{Z}^{(t-1)}}$

   If Euclidean norm of $\tau_{1i} = [\theta_p \ \mathbf{z}_p] > C_1$:

   $\tau'_{1p} = \frac{\tau_{1p}}{\|\tau_{1p}\|_2} \times C_1$ and update θ_p and $\mathbf{z}_p$ accordingly

   Determine $\ell\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\beta}^{(t)}, \mathbf{W}^{(t)}, \mathbf{Z}^{(t)} \middle| \mathbf{X}\right)$

   In every update, $\eta^{(.)}$ is determined from a line search starting at the maximum step size established in step 3. For every 100th iteration the maximum step size parameters from step 3 are updated.

5. **Output parameter values from final iteration**

https://doi.org/10.1017/psy.2025.10068 Published online by Cambridge University Press

2.7. Ordered categorical data

With the methods above in place, a generalization to ordered categorical data is straightforward. That is, the LSIRM is formulated as a sequential item IRT model (Tutz, 1990) as follows:

$$P(X_{pi} \geq h|\theta_p, \beta_{ic}, \mathbf{z}_p, \mathbf{w}_i) = \prod_{c=0}^{h} \omega\left(\theta_p + \beta_{ic} - d(\mathbf{z}_p, \mathbf{w}_i)\right), \tag{18}$$

where $X_{pi} \in \{0, \ldots, C_i - 1\}$ are the ordered categorical item responses and β_{ic} is the effect of item category $c = 0, \ldots, C_i - 1$, which are ordered as follows: $\infty = \beta_{i0} > \beta_{i1} > \cdots > \beta_{ic} > \cdots > \beta_{i(C_i-1)}$. Note that $\gamma = 1$ as discussed above.

This ordinal LSIRM can readily be fit using the methodology discussed above by recoding the $N \times n$ data matrix $\mathbf{X}$ with order categorical elements $X_{pi} \in \{0, \ldots, C_i - 1\}$ into an $N \times \sum_{i=1}^{n}(C_i - 1)$ matrix of cumulative binary dummy coded variables $\mathbf{X}'$ with elements $X'_{pij} \in \{0, 1\}$, for $j = 1, \ldots, C_i - 1$. Specifically,

$$X'_{pij} = 1 \text{ if } X_{pi} \geq j \text{ and } X'_{pij} = 0 \text{ otherwise.} \tag{19}$$

For instance, if $C_i = 3$, this results in $X'_{pi} = [0,0]$ for $X_{pi} = 0$, $X'_{pi} = [1,0]$ for $X_{pi} = 1$, and $X'_{pi} = [1,1]$ for $X_{pi} = 2$. Now, by treating the variables X'_{pij} as dummy items in the binary LSIRM in Equation (18), and by fixing $w_{ijr} = w_{ir}$ for dummy items that correspond to the same original item, fitting the cJML or pJML models to these dummy items will be equivalent to fitting Equation (18) to the original data $\mathbf{X}$. The estimates of β_{ij} on $\mathbf{X}$' will correspond to those of β_{ic} in Equation (18) obtained on $\mathbf{X}'$. Note that by using the approach above, the number of categories for the item responses is allowed to differ across items.

3. Simulation studies

3.1. Design

In this simulation study, we focus on parameter recovery of JML approaches proposed in this article. We focus on an LSIRM with $R = 2$, as this is arguably the model that is used in practice most. In addition, we focus on binary data to be able to compare the results to the existing MCMC approach, which is only available for binary data. Moreover, the binary model underlies the ordinal model, as discussed above; therefore, the results will apply to the ordinal case as well. We simulate the data using the parameterization in Equations (1) and (2) (including the γ parameter). Note that the JML models do not include γ as a free parameter as explained above; however, we do simulate data using γ to manipulate the effect of the latent space on the data and to demonstrate that γ is indeed absorbed in the z_{pr} and w_{ir} JML estimates.

We consider two sample sizes: $N = 1{,}000$ to reflect a more practical setting and $N = 10{,}000$ to study the large sample behavior of the model estimates. We fully cross these sample size conditions with $n = 24$ (practical setting) and $n = 96$ (large sample). Below, we discuss the true values for the parameters, which are inspired by or equal to the MCMC estimates of the deductive reasoning data analyzed in the illustration section below. Specifically, true values for β_i are between 1.5 and 5 and are identical to the deductive reasoning data estimates. In addition, w_{i1} are decreasing from 2 to -2 in equally sized steps and w_{i2} are between -2 and 2 and randomly assigned to the different items. For the conditions with $n = 96$, the item parameters from the conditions with $n = 24$ are repeated four times. For the person parameters, θ_p is specified to follow a normal distribution with mean 0 and variance 0.8, and z_{p1} and z_{p2} are specified to follow a normal distribution with mean 0 and variance, respectively, 0.9 and 0.7. The correlation between z_{p1} and z_{p2} was specified to be equal to -0.14. Finally, γ is equal to 1.7. As the above values are based on the MCMC results from the deductive reasoning data, we also wanted to add a data condition where $\gamma = 1$ to establish how parameter recovery is affected by this parameter. All other parameters in this condition are the same as above. We use 50 replications for each condition

To the data from the different conditions, we fit the LSIRM with $R = 2$ using both pJML and cJML as discussed above. We also fit the LSIRM using MCMC estimation in the R package *lsirm12pl* (Go et al., 2023). We did so to have a benchmark approach available to compare JML results to. That is, we want to

https://doi.org/10.1017/psy.2025.10068 Published online by Cambridge University Press

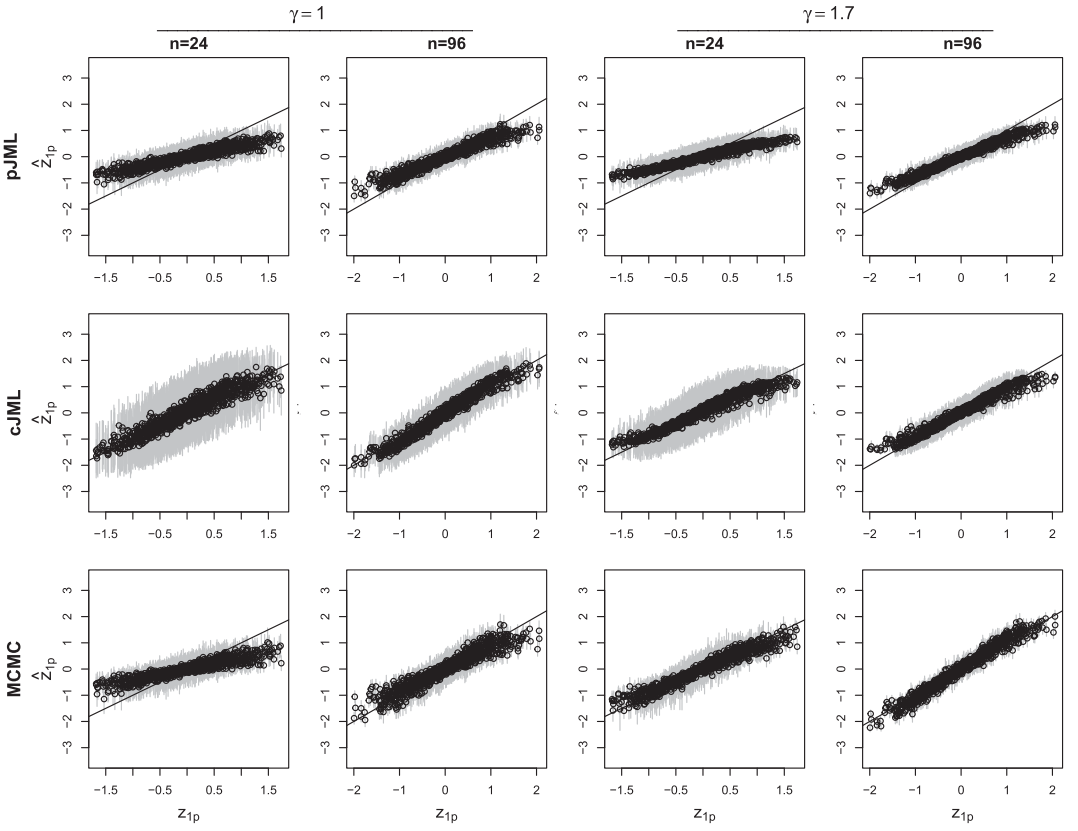


Figure 2. Plot of the true values of z_{1p} (x-axis) and the mean estimated values (y-axis) for penalized joint maximum likelihood (pJML), constrained joint maximum likelihood (cJML), and Markov Chain Monte Carlo for 24 and 96 items in the conditions $\gamma = 1$ and $\gamma = 1.7$. In addition, $N = 1,000$ in all plots and the gray vertical lines indicate the range of the estimates within one standard deviation from the mean. Note that the pJML and cJML estimates are divided by the true value of γ (see the text).

demonstrate that parameter estimates are highly comparable across the different fitting methods after appropriate rotation. The aim of this study is not to show superiority of one method over the other. We used all default settings of *lsirm12pl*. That is, we relied on 15,000 posterior samples of which 2,500 are burn-in and which is thinned by 5. In addition, the default priors of the package are a standard normal prior for β_i, z_{rp}, and w_{ri}, a normal prior with mean 0 and variance σ^2_θ for θ_p, a normal prior with mean 0.5 and variance 1.0 for γ', and an inverse-gamma prior with scale and shape 0.001 for σ^2_θ.

3.2. Results

As the results do not differ importantly across the different dimensions of $\mathbf{z}_p$ and $\mathbf{w}_i$, in the below, we focus on z_{p1} and w_{i1}. In addition, due to space limitations, in the main text, we mostly provide graphical and verbal displays of the results, but we will make references to tables that can be found in the Supplementary Material.

3.2.1. Parameter recovery

Figure 2 contains the average estimates and standard deviations across replications for z_{p1} in the cases of $n = 24$ and $n = 96$ in the $\gamma = 1$ and $\gamma = 1.7$ conditions for $N = 1,000$. In addition, similarly, Figure 3 contains the average estimates and standard deviations across replications for w_{i1} in the cases of $N = 1,000$ and $N = 10,000$ in the $\gamma = 1$ and $\gamma = 1.7$ conditions for 24 items. In the figures, the true values are the plots

https://doi.org/10.1017/psy.2025.10068 Published online by Cambridge University Press

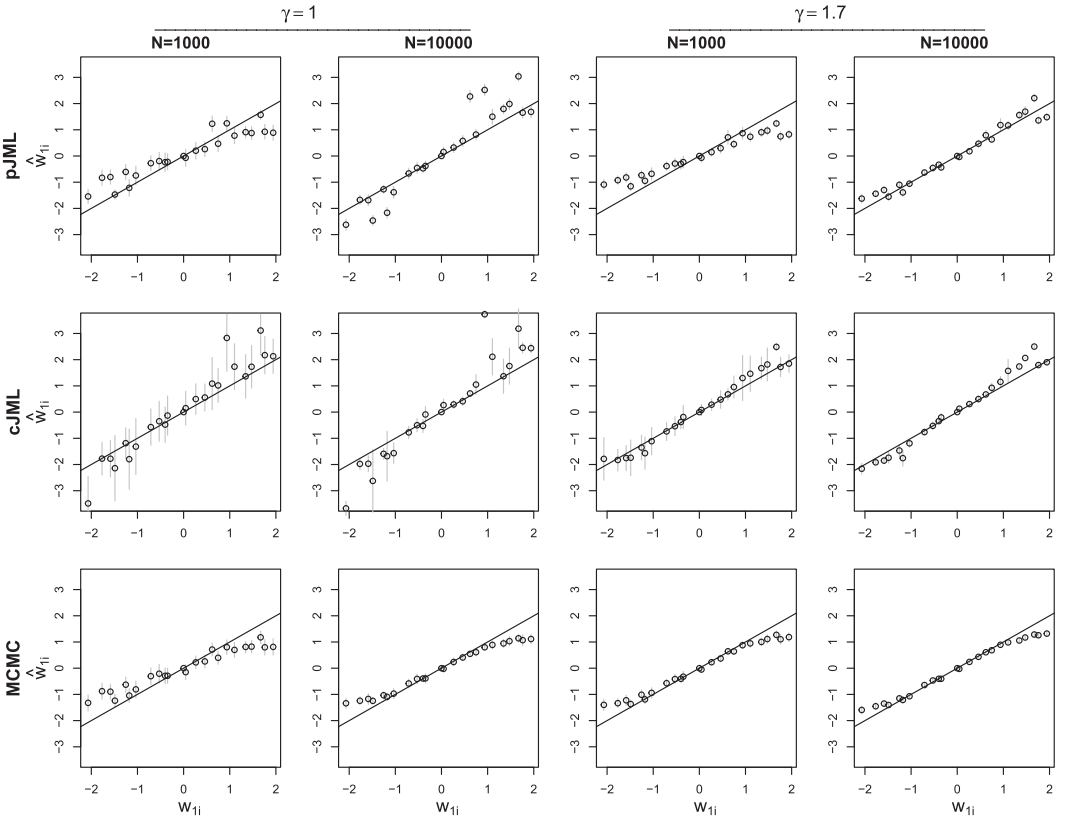


Figure 3. Plot of the true values of w_{1i} (x-axis) and the mean estimated values (y-axis) for penalized joint maximum likelihood (pJML), constrained joint maximum likelihood (cJML), and Markov Chain Monte Carlo for $N = 1{,}000$ and $N = 10{,}000$ in the conditions $\gamma = 1$ and $\gamma = 1.7$. In addition, $n = 24$ in all plots and the gray vertical lines indicate the range of the estimates within one standard deviation from the mean. Note that the pJML and cJML estimates are divided by the true value of γ (see the text).

on the x-axis and the mean estimates on the y-axis. For the pJML and cJML estimates in the condition $\gamma = 1.7$, the estimates are divided by 1.7, as the γ parameter is absorbed in these estimates (as it is not a free parameter in the JML approaches, see above). As can be seen, pJML and MCMC estimates tend to be unbiased up to a shrinkage effect that decreases for increasing N and n. For cJML, there is no shrinkage effect noticeable, which is understandable given the constraints used in this approach (Equations (8) and (9)), which operate as uniform priors. Shrinkage of the MCMC estimates diminishes at a higher rate for increasing n or N and γ compared to pJML especially for z_{p1}.

In Supplementary Table A1, the mean absolute bias (MAB), the variance (VAR), and the mean squared error (MSE) of the estimates for person parameters z_{1p} and θ_p averaged over persons are reported. To separate the shrinkage effects from these statistics, for each parameter, the estimates are scaled to have a standard deviation equal to that of the true parameters. The recovery of the standard deviation of the estimates is studied separately later.

The most important result is that for all three methods, the MAB, VAR, and MSE decrease for z_{p1} and θ_p for an increasing number of items, n. In addition, pJML performs comparable to MCMC especially for larger N and n. For cJML, the VAR and MSE are slightly higher compared with cJML and MCMC, particularly in conditions with smaller N and n.

Supplementary Table A2 contains the MSE, VAR, and MAB for item parameters w_{i1} and b_{i1} averaged over items. For all three methods, these statistics decrease for both w_{i1} and b_i for an increasing sample size, N. One exception is for the estimates of b_i for pJML in the $\gamma = 1$ condition where the VAR

https://doi.org/10.1017/psy.2025.10068 Published online by Cambridge University Press

increases slightly. Similar as for the person parameters, for w_i, pJML performs comparable to MCMC especially for larger N and n. For cJML, the VAR and MSE are slightly higher compared with cJML and MCMC particularly in conditions with smaller N and n. For b_i, results are generally somewhat better for MCMC in the case of smaller N, n, and γ.

Next, we focus on accuracy of the standard deviation of the estimates. As the variance of these standard deviations is very small, we only report the mean bias in Supplementary Table A3. Previously, we focused on the MAB to prevent biasing effects from canceling out across items or persons. However, here, the mean bias is more interesting as it gives an indication of the direction of the effect. That is, a mean bias smaller than 0 indicates shrinkage of the parameter estimates, whereas a mean bias larger than 0 indicates increased parameter variability. In addition, as we are not aggregating over persons or items, effects cannot cancel out.

As the standard deviations of the parameters are not explicit model parameters for pJML and cJML, and these parameters are fixed for MCMC (except for σ^2_θ), the mean bias of the standard deviations has been determined by focussing on the theoretically expected standard deviation. For MCMC, the expected standard deviation is the standard deviations of the true values used to simulate the data (see above). For pJML and cJML, the expected value of $SD(w_{1i})$ and $SD(z_{1p})$ are, respectively, $\gamma \times SD(w_{1i})$ and $\gamma \times SD(z_{1p})$ as γ is expected to be absorbed in w_{1i} and z_{1p} as discussed above. For SD (b_i) and $SD(\theta_p)$, the expected values are the standard deviations of the true values used to simulate the data. The results show that, as was already concluded from Figures 2 and 3, shrinkage effects are present for pJML and MCMC with slightly negative biases that decrease for increasing N and n. This decrease generally follows a higher rate for MCMC compared with pJML. For cJML, the estimates are mostly associated with increased variability with positive biases that decrease for increasing N and n.

3.2.2. Estimation time

Figure 4 contains a bar plot of the mean estimation time of the different approaches for $\gamma = 1$ in minutes. For $\gamma = 1.7$, pJML and cJML are even faster, while MCMC is about as fast (see Supplementary Table A4). It can be concluded that pJML is the fastest, and MCMC is the slowest. These averages are just given as an indication; the comparison between cJML/pJML and MCMC may not be fully fair as the settings for the MCMC routine (e.g., number of samples) may be tuned in specific conditions so that the average times shown in the figure are an overestimate. However, looking at the number of samples you can take from the posterior parameters in the average time pJML took to converge (also in the figure), it is clear that the JML-based approaches require a substantial smaller amount of time. As discussed, this brings possibilities for establishing model fit, which are discussed below.

3.2.3. Cross-validation

We conducted additional simulations to test the appropriateness of our proposed K-folds cross-validation procedure. The full description of this study and the results can be found in Appendix B of the Supplementary Material. The main results indicate that cross-validation with the UCE, URE, and RSS metrics as proposed above can distinguish well between the models for $\gamma = 1$, irrespective of the number of items, the number of persons, or the number of folds considered. For the smaller data scenario ($N = 1{,}000$, $n = 24$), the *UCE* has difficulties, correctly identifying the $R = 1$ model for five folds (true positive rate of 0.66). However, using 10 folds, the true positive rate increases to 0.78, which is more acceptable. For $\gamma = 1.7$, all true positives were at least 0.98, with one exception of a true positive of 0.86 for the *UCE* when $R = 1$ and $K = 5$.

It should be noted that our cross-validation procedure tends to overselect, and not underselect. That is, although, in general, the correct model is picked most of the times, if an incorrect model is being selected, it is always a model with a larger R than the true model. This is a known property of cross-validation in general (see Li et al., 2020), but it should be kept in mind when using cross-validation for dimensionality selection.

https://doi.org/10.1017/psy.2025.10068 Published online by Cambridge University Press

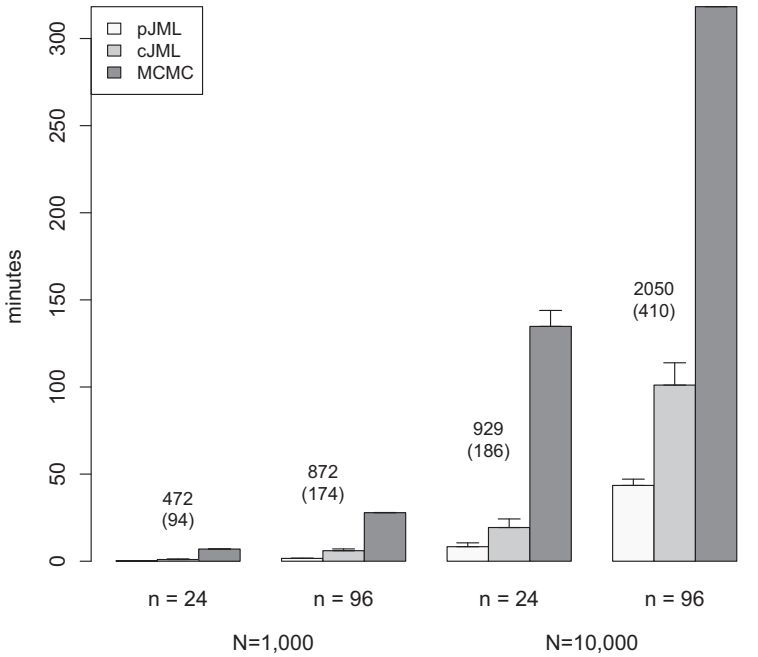


Figure 4. Bar plot of the mean estimation time in minutes for penalized joint maximum likelihood (pJML), constrained joint maximum likelihood, and Markov Chain Monte Carlo for the different conditions and $\gamma = 1$. The error bars indicate one standard deviation. The numbers indicate the rounded number of samples you can take from the posterior parameters in the average time pJML took to converge (after thinning in brackets).

3.3. Conclusion

In general, our cJML and pJML estimation approaches perform as intended. The parameter recovery is comparable to that of MCMC with some minor differences in specific conditions due to differences in the priors and restrictions imposed. The cJML estimates are generally more variable because of the constraints in this approach being the least restrictive among the three approaches. However, it should be noted that in practice, C_1 and C_2 may be chosen by cross-validation, which may decrease parameter variability. The pJML approach is generally faster than cJML, with MCMC being slower than the JML approaches. Finally, the cross-validation procedure works well in selecting the number of dimensions of the latent space.

4. Illustration 1: dichotomous data

4.1. Data

In this section, we analyze the scores of $N = 418$ children to 24 items from a Piagetian syllogistic deductive reasoning test (see Spiel et al., 2001). These data have previously been analyzed using an MCMC LSIRM by Jeon et al. (2021). In the present application, we use these data to illustrate the similarities between the results of JML and MCMC after appropriate rotation. In addition, Jeon et al. tested the fit of an $R = 0$ LSIRM against an $R = 2$ LSIRM using the spike and slab prior approach discussed above. Here, we verify these results using the present cross-validation approach with $K = 10$ and extend these results by also considering $R = 1$ and $R = 3$. In addition, we explore the results of the $R = 3$ model. Models are fit using the same methods and settings as discussed for Simulation Study 1.

4.2. Results

We first consider the LSIRM with $R = 2$ as studied by Jeon et al. (2021) in an MCMC framework. After the rotation of the MCMC results to an echelon structure, correlations among the parameter estimates

https://doi.org/10.1017/psy.2025.10068 Published online by Cambridge University Press

Table 1. Model fit indices as based on a 10-fold cross-validation.

R	*UCE*	*URE*	*RSS*
0	3,209	2,424	2,049
1	2,536	1,571	1,729
2	**2,403**	1,416	**1,688**
3	2,426	**1,364**	1,696

Note: For each metric, the smallest value is in boldface. *RSS*, residual sum of squares; *UCE*, unnormalized classification error; *URE*, unnormalized ROC error.

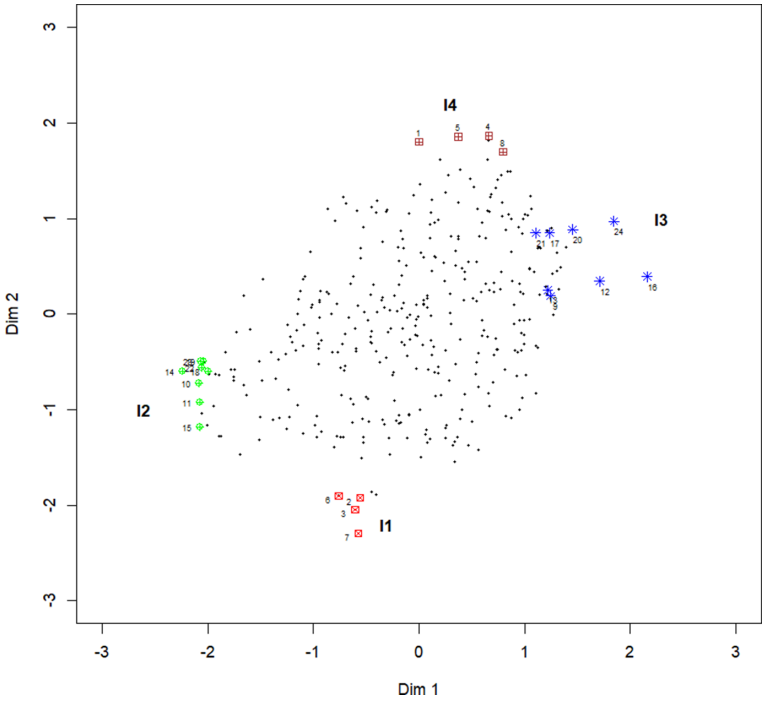


Figure 5. Plot of the estimates of vector $\mathbf{w}_i$ from an $R = 2$ model for all items. I1–I4 are the item groups identified by Jeon et al. (2021).

across the different approaches are close to 1.0 (i.e., at least 0.989 for the MCMC—JML parameter estimate correlations) or practically 1.0 (i.e., at least 0.999 for the pJML and cJML parameter estimate correlations). We, therefore, only consider the pJML results below.

Table 1 contains the results of the cross-validation for models with $R = 0, R = 1$, $R = 2$, *and* $R = 3$. As can be seen, the *UCE* and *RSS* select the $R = 2$ model to be the best fitting, whereas the *URE* favors the $R = 3$ model. As Jeon et al. (2021) considered an $R = 2$ model, we here also explore if an $R = 3$ model will give some additional insights. First, in Figure 5, the $\mathbf{w}_i$ and $\mathbf{z}_p$ estimates are plotted for an $R = 2$ LSIRM. The figure is highly comparable to Figure 8a of Jeon et al., up to a difference in orientation due to the differences in rotation between pJML and MCMC. Following Jeon et al., we label the four item clusters that are observed as "I1" to "I4" in the figure. These are interpreted as, respectively, concrete complex inference items, abstract and counterfactual logical fallacy items, bi-conditional items, and complex algebra items.

https://doi.org/10.1017/psy.2025.10068 Published online by Cambridge University Press

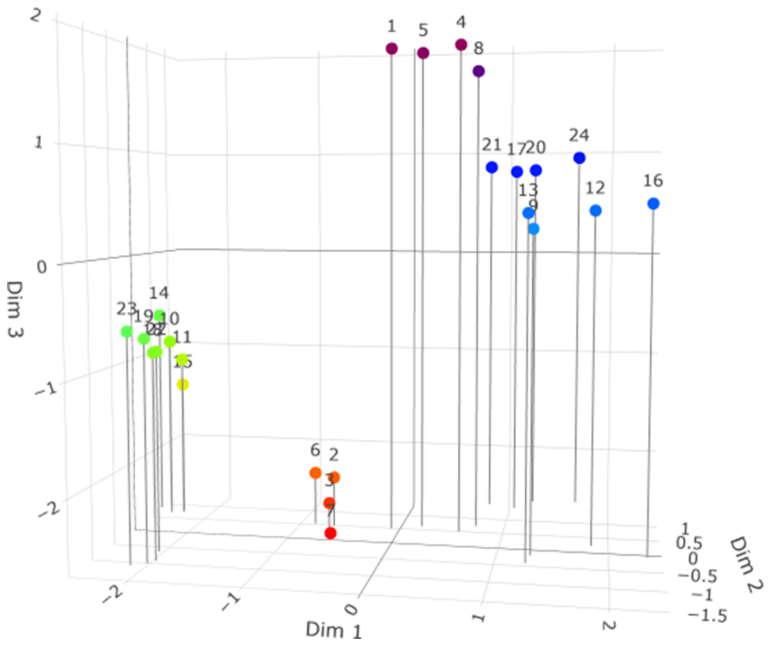


Figure 6. Plot of the estimates of vector $\mathbf{w}_i$ from an $R = 3$ model for all items. The view has been adjusted to demonstrate correspondence with the results for the $K = 2$ model, which lacks dimension 1 (see the text).

Next, we fit an $R = 3$ LSIRM. Figure 6 contains a three-dimensional plot of the estimates of $\mathbf{w}_i$ from this model. We first focus on an orientation of the view of the plot in which variability on dimension 1 (w_{i1}) is mostly masked to show the correspondence with the $R = 2$ plot in Figure 5. It is important to note that by using the echelon restrictions in the way illustrated in Equation (10), the first dimension from an $R = 2$ model corresponds to the second dimension in an $R = 3$ model. Therefore, we masked dimension 1 in Figure 6 to show the correspondence to Figure 5. As these figures indeed correspond, we now demonstrate the differences on the first dimension in Figure 7. As can be seen, an additional item cluster seems to arise splitting cluster I2 into I2a and I2b where I2a are items 10, 11, 14, and 15, and I2b are items 18, 19, 22, and 23. In Figure 8, we next add the estimates of $\mathbf{z}_p$ to the three-dimensional plot. This figure illustrates that persons differ in their proximity to the different item clusters, with the I3 items (blue) being more isolated, for instance, compared with the I1 items (red).

4.3. Conclusion

The main purpose of this real data analysis was to illustrate our JML-based approach on real data. A cross-validation indicated that an $R = 2$ model seems, indeed, appropriate for the deductive reasoning task analyses by Jeon et al. (2021) using an MCMC approach. As results indicated that there may be variability on an additional dimension, we explored a three-dimensional model and found some differences among the items in the three-dimensional space.

5. Illustration 2: ordered categorical data

5.1. Data

In this section, we illustrate the application of the present methodology to the analysis of ordinal data. To this end, we analyze data from the Adjective Check List (ACL; Gough & Heilbrun, 1980). The original ACL consists of 300 adjectives for which respondents need to indicate on a 5-point scale to

https://doi.org/10.1017/psy.2025.10068 Published online by Cambridge University Press

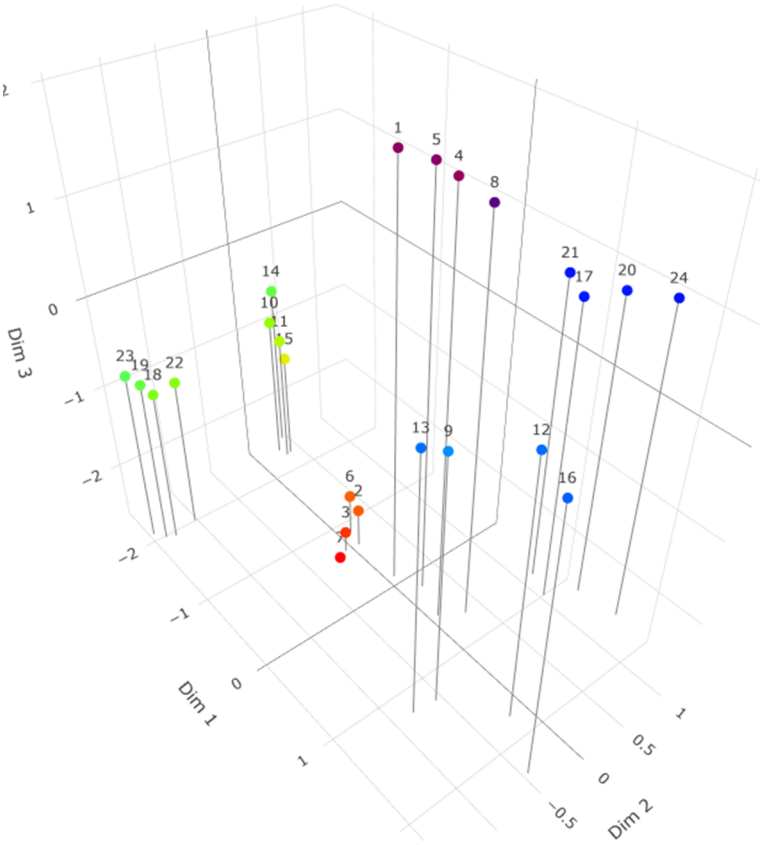


Figure 7. Plot of the estimates of vector $\mathbf{w}_i$ from an $R = 3$ model for all items. The view has been chosen to demonstrate the differences across dimension 1.

what degree the adjective describes the personality of the respondent. The item scores of 433 students on an adapted version of the ACL is available as part of the Mokken package in R (Van der Ark, 2007). Here, we focus on the Communality and the Dominance scale, which both consist of 10 items. We fit the ordinal LSIRM in Equation (18) to each scale separately using the cumulative binary dummy coding approach described above. We focus on $R = 2$ for ease of interpretation and visualization. The contraindicative items (indicated by an asterisk in the item name in the presentation of the results) are reversely coded. For the item "reliable" from the Communality scale, respondents did not use the lowest response category. Therefore, this item only has three threshold parameters, and all other items have four threshold parameters.

5.2. Results

Item parameter estimates for β_{ic} and $\boldsymbol{w}_i$ of the ordinal LSIRM can be found in Supplementary Table C1. In Figure 9, the latent space, including $\boldsymbol{z}_p$, is displayed for the two scales. For the item positions $\boldsymbol{w}_i$, the figure contains 99% confidence ellipses based on bootstrapped standard errors using 1,000 samples. Even though the scales are intended to be unidimensional, as can be seen from the figure, for both scales, the items seem to meaningfully cluster. For instance, for the Communality scale, omitting the "unscrupulous*" and "honest" items that have relatively large standard errors, three clusters arise: the "unintelligent*" item on the on side, which is apart from all the other items; the "reliable"–"dependable" cluster; and a cluster with among others "cruel*" and "unfriendly*." Items within a cluster are close to

https://doi.org/10.1017/psy.2025.10068 Published online by Cambridge University Press

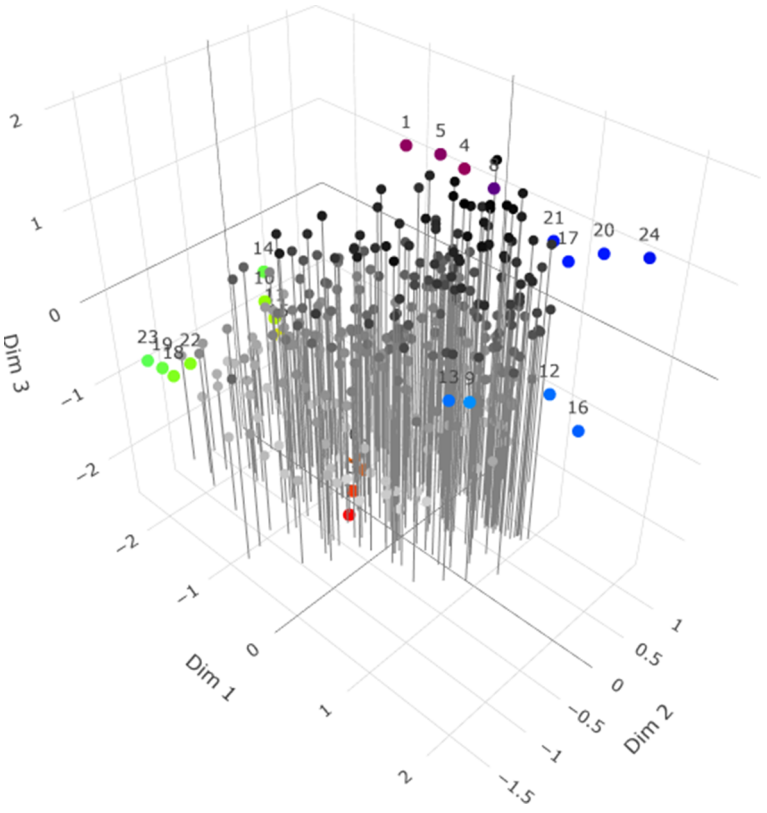


Figure 8. Same as Figure 7, but with the estimates of $\mathbf{z}_p$ added.

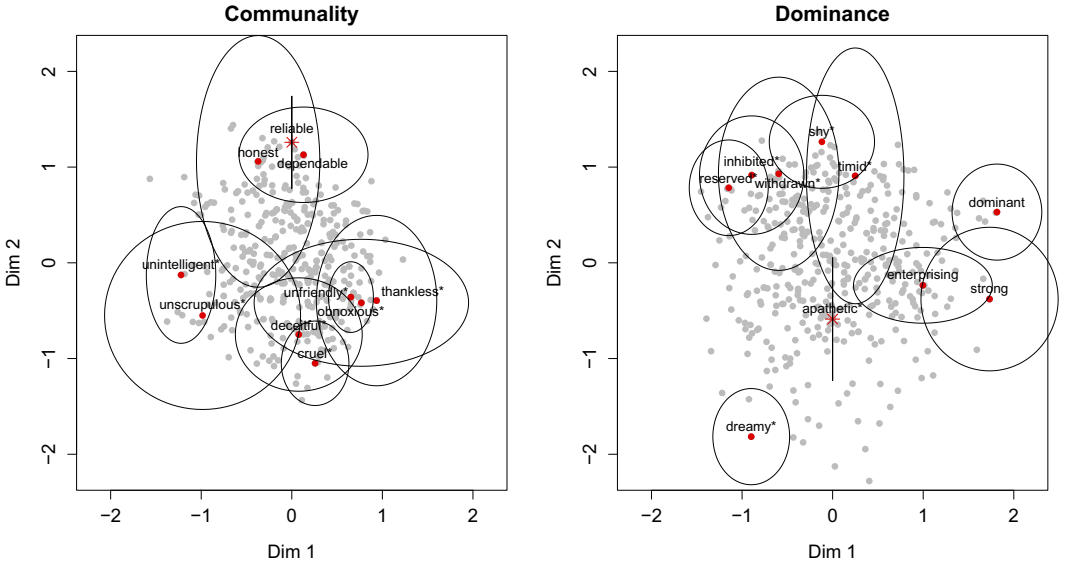


Figure 9. Graphical representation of the ordinal latent space item response model in application 2 for the Communality scale (left) and the Dominance scale (right) of the ACL. The red dots represent the item locations $\boldsymbol{w}_i$, and the gray dots represent the person locations $\boldsymbol{z}_p$. The ellipses around the item locations give the range of the 99% confidence intervals. The location of the first item ("reliable" in Communality scale and "apathic" in the Dominance scale) is indicated by a star instead of a dot as w_{11} is fixed to 0 for this item, and a 99% confidence line is displayed for this item.

https://doi.org/10.1017/psy.2025.10068 Published online by Cambridge University Press

each other while being further away from the other clusters. For the Dominance scale, something similar is observed, omitting the "timid*" item, which has a relatively large standard error, "dreamy*" stands out as it is apart from most of the items, while "dominant," "strong," and "enterprising" form a cluster, and the other items form a separate cluster. "Apathic*" is separate but in the middle of the three clusters.

For the person positions, respondents are scattered across the clusters, with some persons have an estimate closer to, for example, the "reserved*–withdrawn*" item cluster than to other items. Respondents near the origin of the coordinate system (0,0) score high or low on all items (which could be inferred from their θ_p estimate).

5.3. Conclusion

We demonstrated the viability of the ordinal LSIRM on a real dataset. In practice, results like the above can be used to make profiles. For instance, using the adjectives check list above, in job applicant selection, one may want to aim for respondents close to the "enterprising" and "strong" items, but also to the shy* and timid* items, while not necessarily close to the "dreamy*" item. Using the present approach such distances between persons and items can readily be interpreted. Other statistical methods like factor analysis are valuable but serve different purposes compared with the latent space approach adopted here. That is, factor analysis parameters cannot be interpreted in terms of item–person distances as in factor analysis triangular inequality is violated. For a more elaborate discussion on the differences, see Jeon et al. (2021).

6. Discussion

In this study, we proposed regularized JML estimation to fit the LSIRM to data. A key advantage is that the estimation is fast, which hopefully facilitates the use of these models by the applied researcher. Owing to its fast nature, model selection can be conducted using K-folds cross-validation, which is valuable as such model selection tools are still relatively limited in the LSIRM literature.

These advantages come at the cost of reduced information about the posterior parameter distribution. That is, the MCMC approach results in estimates of the full posteriors parameter distributions, while, as discussed, the estimates in our approach can be conceived as MAP estimates (e.g., Swaminathan & Gifford, 1985). Our approach does not give such detailed information about the shape of the posterior parameter distributions. In the illustration section, we bootstrapped the standard errors, but this increases the computational burden. The observed and expected information matrices may be useful in obtaining uncertainty measures; however, for our approach, the theoretical properties of these matrices are strictly unknow. At best, they give a local approximation if the posterior is normal near its mode. As MCMC can account for any form of the parameter distribution in principle, MCMC is superior in its measures of parameter uncertainty. Therefore, as noted earlier, we do not view our approach as a replacement for existing LSIRM modeling tools, but rather as an extension. For example, when uncertainty measures are of primary importance, MCMC estimation remains preferable. Conversely, if the main goal is to visualize latent space positions or determine the dimensions of the latent space, our JML-based approaches are recommended.

As discussed in this article, traditional JML estimation as proposed by Birnbaum (1968) is known to be asymptotically inconsistent (e.g., Haberman, 1977). For a multidimensional unstructured two-parameter logistic model, Chen et al. (2019) proofed JML to be consistent in the double asymptotic sense under appropriate constraints. As mentioned, we did not study such theoretical consistency for our approach. However, the simulation results suggest that parameter bias decreases for increasing number of items and sample size. Moreover, theoretical consistency was not our main objective in this study. That is, we studied JML as a less time-consuming complement to the existing MCMC procedures but with comparable results. In that respect, our approach performs as desired. It should be noted, however, that our implementation of the constrained JML approach uses the results from Chen et al. (2019),

https://doi.org/10.1017/psy.2025.10068 Published online by Cambridge University Press

which are known to be asymptotically consistent in the average sense (the overall MSE approaches 0). For entry-wise consistency (i.e., MSE approaches 0 for each parameter), additional post processing steps are needed (see Chen and Li, 2024). Here, we did not consider such post processing, but we note that our approach is equally amenable to such a treatment.

Supplementary material. To view supplementary material for this article, please visit http://doi.org/10.1017/psy.2025.10068.

Disclosure of use of AI tools. In this research, ChatGPT has been used to assist in literature searches and some minor writing edits.

Funding statement. There are no funders to report for this submission. Open access funding provided by University of Amsterdam.

Competing interests. The authors declare no competing interests.

References

Andersen, E. B. (1973). *Conditional inference and models for measuring*. Mentalhygiejnisk Forlag.

Bergner, Y., Halpin, P., & Vie, J. J. (2022). Multidimensional item response theory in the style of collaborative filtering. *Psychometrika*, *87*(1), 266–288. https://doi.org/10.1007/s11336-021-09788-9

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

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*(4), 443–459. https://doi.org/10.1007/BF02293801

Cai, L. (2010). High-dimensional exploratory item factor analysis by a Metropolis–Hastings Robbins–Monro algorithm. *Psychometrika*, *75*, 33–57. https://doi.org/10.1007/s11336-009-9136-x

Chen, Y., & Li, X. (2022). Determining the number of factors in high-dimensional generalized latent factor models. *Biometrika*, *109*(3), 769–782. https://doi.org/10.1093/biomet/asab044

Chen, Y., & Li, X. (2024). A note on entrywise consistency for mixed-data matrix completion. *Journal of Machine Learning Research 25*(343), 1–66.

Chen, Y., Li, X., & Zhang, S. (2019). Joint maximum likelihood estimation for high-dimensional exploratory item factor analysis. *Psychometrika*, *84*(1), 124–146. https://doi.org/10.1007/s11336-018-9646-5

Chen, Y., Li, X., and Zhang, S. (2020). Structured latent factor analysis for large-scale data: Identifiability, estimability, and implications. *Journal of the American Statistical Association 115*, 1756–1770. https://doi.org/10.1080/01621459.2019.1635485

De Carolis, L., Kang, I., & Jeon, M. (2025). A latent space graded response model for Likert scale psychological assessments. Manuscript submitted for publication.

Dolan, C. V., Oort, F. J., Stoel, R. D., & Wicherts, J. M. (2009). Testing measurement invariance in the target rotated multigroup exploratory factor model. *Structural Equation Modeling*, *16*, 295–314. https://doi.org/10.1080/10705510902751416

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, *80*(1), 27–38. https://doi.org/10.1093/biomet/80.1.27

George, E. I., & McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica*, *7*(2), 339–373.

Go, D., Jeon, M., Lee, S., Jin, I.-H., & Park, H.-J. (2022). Analyzing differences between parent- and self-report measures with a latent space approach. *PLoS One*, *17*, Article e0269376. https://doi.org/10.1371/journal.pone.0269376

Go, D., Kim, G., & Park, J. (2023). *lsirm12pl: Latent space item response model*. R package version 1.3.1. https://CRAN.R-project.org/package=lsirm12pl

Gough, H. G., & Heilbrun, A. B. (1980). *The adjective check list, manual 1980 edition*. Consulting Psychologists Press.

Gower, J. C. (1975). Generalized Procrustes Analysis. *Psychometrika*, *40*(1), 33–51. https://doi.org/10.1007/BF02291478

Haberman, S. J. (1977). Maximum likelihood estimates in exponential response models. *The Annals of Statistics*, *5*(5), 815–841. https://doi.org/10.1214/aos/1176343941

Haslbeck, J., & van Bork, R. (2024). Estimating the number of factors in exploratory factor analysis via out-of-sample prediction errors. *Psychological Methods*, *29*(1), 48–64. https://doi.org/10.1037/met0000528

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). Springer.

Ho, E., & Jeon, M. (2023). Interaction map: A visualization tool for personalized learning based on assessment data. *Psych*, *5*(4), 1140–1155. https://doi.org/10.3390/psych5040076

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association*, *97*(460), 1090–1098. https://doi.org/10.1198/016214502388618906

Ishwaran, H., & Rao, J. S. (2005). Spike and slab variable selection: Frequentist and Bayesian strategies. *The Annals of Statistics*, *33*, 730–773. https://doi.org/10.1214/009053604000001147

Jennrich, R. I. (1978). Rotational equivalence of factor loading matrices with specified values. *Psychometrika*, *43*, 421–426. https://doi.org/10.1007/BF02293650

Jeon, M., Jin, I. H., Schweinberger, M., & Baugh, S. (2021). Mapping unobserved item–respondent interactions: A latent space item response model with interaction map. *Psychometrika*, *86*(2), 378–403. https://doi.org/10.1007/s11336-021-09762-5

Jeon, M., & Schweinberger, M. (2024). A latent process model for monitoring progress toward hard to-measure targets with applications to mental health and online educational assessments. *The Annals of Applied Statistics*, *18*(3), 2123–2146. https://doi.org/10.1214/24-AOAS1873

Jin, I. H., Jeon, M., Schweinberger, M., Yun, J., & Lin, L. (2022). Multilevel network item response modelling for discovering differences between innovation and regular school systems in Korea. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *71*(5), 1225–1244. https://doi.org/10.1111/rssc.12569

Jin, I. H., Yun, J., Kim, H., & Jeon, M. (2023). A latent space accumulator model for response time: Applications to cognitive assessment data. *Stat*, *12*(1), e632. https://doi.org/10.1002/sta4.632

Kang, I., & Jeon, M. (2024). A recent development of a network approach to assessment data: Latent space item response modeling for intelligence studies. *Journal of Intelligence*, *12*(4), Article 38. https://doi.org/10.3390/jintelligence12040038

Kim, N., Jeon, M., & Partchev, I. (2024). Conditional dependence across slow and fast item responses: With a latent space item response modeling approach. *Journal of Intelligence*, *12*(2), Article 23. https://doi.org/10.3390/jintelligence12020023

Li, T., Levina, E., & Zhu, J. (2020). Network cross-validation by edge sampling. *Biometrika*, *107*(2), 257–276. https://doi.org/10.1093/biomet/asaa006

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Lawrence Elbaum.

Luo, J., De Carolis, L., Zeng, B., & Jeon, M. (2023). Bayesian estimation of latent space item response models with JAGS, Stan, and NIMBLE in R. *Psych*, *5*(2), 396–415. https://doi.org/10.3390/psych5020027

Ma, Z., Ma, Z., & Yuan, H. (2020). Universal latent space model fitting for large networks with edge covariates. *Journal of Machine Learning Research*, *21*(4), 1–67. Retrevied from: https://www.jmlr.org/papers/v21/17-470.html

McDonald, R. P. (2013). *Test theory: A unified treatment*. Psychology Press. https://doi.org/10.4324/9781410601087

Mislevy, R. J., & Stocking, M. L. (1989). A consumer's guide to LOGIST and BILOG. *Applied psychological measurement*, *13*(1), 57–75.

Nocedal, J., & Wright, S. J. (2006). *Numerical optimization*. New York, NY: Springer New York.

Park, J., Jeon, Y., Shin, M., Jeon, M., & Jin, I. H. (2022). Bayesian Shrinkage for Functional Network Models, With Applications to Longitudinal Item Response Data. *Journal of Computational and Graphical Statistics*, *31*(2), 360–377. https://doi.org/10.1080/10618600.2021.1999823

Park, J., Jin, I. H., & Jeon, M. (2023). How Social Networks Influence Human Behavior: An Integrated Latent Space Approach for Differential Social Influence. *Psychometrika*, *88*(4), 1529–1555. https://doi.org10.1007/s11336-023-09934-5

Rasch, G. (1960). *Probabilistic model for some intelligence and achievement tests*. Danish Institute for Educational Research.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. https://doi.org/10.1093/biomet/63.3.581

Sibson, R. (1979). Studies in the robustness of multidimensional scaling: Perturbational analysis of classical scaling. *Journal of the Royal Statistical Society: Series B (Methodological)*, *41*(2), 217–229. https://doi.org/10.1111/j.2517-6161.1979.tb01076.x

Spiel, C., Glück, J., & Gösslers, H. (2001). Stability and change of unidimensionality: The sample case of deductive reasoning. *Journal of Adolescent Research*, *16*(2), 150–168. https://doi.org/10.1177/0743558401162003

Swaminathan, H., & Gifford, J. A. (1982). Bayesian estimation in the Rasch model. *Journal of Educational Statistics*, *7*(3), 175–191. https://doi.org/10.3102/10769986007003175

Swaminathan, H., & Gifford, J. A. (1985). Bayesian estimation in the two-parameter logistic model. *Psychometrika*, *50*, 349–364. https://doi.org/10.1007/BF02294110

Swaminathan, H., & Gifford, J. A. (1986). Bayesian estimation in the three-parameter logistic model. *Psychometrika*, *51*, 589–601. https://doi.org/10.1007/BF02295598

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

Tutz, G. (1990). Sequential item response models with an ordered response. *British Journal of Mathematical and Statistical Psychology*, *43*(1), 39–55. https://doi.org/10.1111/j.2044-8317.1990.tb00925.x

Van der Ark, L. A. (2007). Mokken scale analysis in R. *Journal of Statistical Software*, *20*(11), 1–19. https://doi.org/10.18637/jss.v020.i11

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one out cross-validation and WAIC. *Statistics and Computing*, *27*, 1413–1432.

Wansbeek, T., & Meijer, E. (2000). *Measurement error and latent variables in econometrics*. North-Holland.

Wingersky, M. S. (1983). LOGIST: A program for computing maximum likelihood procedures for logistic test models. In R. K. Hambleton (Ed.), *Applications of item response theory*. (pp. 45–56). Educational Research Institute of British Columbia.

https://doi.org/10.1017/psy.2025.10068 Published online by Cambridge University Press

Wright, B. D., & Mead, R. J. (1976). *Rasch model analysis with the BICAL computer program*. U.S. Army Research Institute for Behavioral and Social Sciences. https://apps.dtic.mil/sti/tr/pdf/ADA119877.pdf

Zhang, X., Xu, G., & Zhu, J. (2022). Joint latent space models for network data with high-dimensional node variables. *Biometrika*, *109*(3), 707–720. https://doi.org/10.1093/biomet/asab063

Cite this article: Molenaar, D. and Jeon, M. (2026). Regularized Joint Maximum Likelihood Estimation of Latent Space Item Response Models. *Psychometrika*, 335–359. https://doi.org/10.1017/psy.2025.10068

https://doi.org/10.1017/psy.2025.10068 Published online by Cambridge University Press