

Article

Regularized Latent Class Analysis for Polytomous Item Responses: An Application to SPM-LS Data

Alexander Robitzsch ^{1,2*} ¹ IPN – Leibniz Institute for Science and Mathematics Education, Kiel, Germany² Centre for International Student Assessment (ZIB), Kiel, Germany

* Correspondence: robitzsch@leibniz-ipn.de

Abstract: The last series of Raven's standard progressive matrices (SPM-LS) test were studied with respect to its psychometric properties in a series of recent papers. In this paper, the SPM-LS dataset is analyzed with regularized latent class models (RLCM). For dichotomous item response data, an alternative estimation approach for RLCMs is proposed. For polytomous item responses, different alternatives for performing regularized latent class analysis are proposed. The usefulness of the proposed methods is demonstrated in a simulated data illustration and for the SPM-LS dataset. For the SPM-LS dataset, it turned out the regularized latent class model resulted in five partially ordered latent classes.

Keywords: regularized latent class analysis, regularization, fused regularization, fused grouped regularization, distractor analysis

1. Introduction

There has been recent interest in assessing the usefulness of short versions of the Raven's Progressive Matrices. [1] composed the last twelve matrices of the Standard Progressive Matrices (SPM-LS) and argued that it could be regarded as valid indicator of general intelligence g . As part of this special issue, the SPM-LS dataset that was analyzed in [1] was reanalyzed in a series of papers applying a wide range of psychometric approaches. In particular, [2] investigated item distractor analysis with a particular focus on reliability, [3] provided additional insights due to dimensionality analysis, [4] applied the Haberman interaction model using the R package dexter, Mokken scaling was employed by [5], and, finally, [6] presented Bayesian item response modeling using the R package brms.

This diversity of approaches have in common that quantitative latent variable models were utilized. In this paper, discrete latent variable models (i.e., latent class models) are applied for analyzing the SPM-LS dataset. With discrete latent variable models, the analysis of types instead of traits is the primary focus (see [7] and [8] papers). A disadvantage of discrete latent variable models is that they often have a large number of parameters to estimate. So-called regularization approaches automatically reduce the number of parameters to estimate (see [9] for the technique in structural equation modeling). In this paper, these regularization approaches are applied in discrete latent variable models, and some extensions for polytomous data are proposed.

The paper is structured as follows. In Section 2, we give a brief overview of latent class analysis. In Section 3, regularized latent class analysis for dichotomous and polytomous data is introduced. In Section 4, we apply proposed models of Section 3 in a simulated data illustration. In Section 5, we apply regularized latent class analysis to the SPM-LS dataset. Finally, in Section 6, we conclude with a discussion.

2. Latent Class Analysis

Latent variable models represent discrete items by a number of latent variables (see [10] for an overview). These latent variables can be categorical or quantitative or a mixture of both. Quantitative

latent variables are considered in factor analysis, structural equation models, or item response models. In this article, we focus on categorical latent variables. In this case, latent variables are labeled as latent classes and are extensively studied in the literature of latent class analysis (LCA; [11], [12], [13]).

A latent class model (LCM) represents the multivariate distribution of I categorical items $X = (X_1, \dots, X_I)$ by a fixed number of C latent classes. Let U denote the latent class variable that takes one of the values $1, 2, \dots, C$. It is assumed that items X_i are conditionally independent on the latent class variable U . This means that it holds that

$$P(\mathbf{X} = \mathbf{x} | U = c) = \prod_{i=1}^I P(X_i = x_i | U = c) \quad \text{for } \mathbf{x} = (x_1, \dots, x_I) \quad . \quad (1)$$

The multivariate probability distribution is then given as a mixture distribution

$$P(\mathbf{X} = \mathbf{x}) = \sum_{c=1}^C P(U = c) \prod_{i=1}^I P(X_i = x_i | U = c) \quad . \quad (2)$$

Applications of LCMs to intelligence tests can be found in [14] or [15].

2.1. Exploratory Latent Class Analysis for Dichotomous Item Responses

In this subsection, we describe the LCM for dichotomous items. Let $p_{ic} = P(X_i = 1 | U = c)$ denote the item response probability for correctly solving item i if a person is located in class c . In the estimation, these bounded parameters ($p_{ic} \in [0, 1]$) are transformed onto the real line by using the logistic transformation (see also [14])

$$P(X_i = x | U = c) = \frac{\exp(x\gamma_{ic})}{1 + \exp(\gamma_{ic})} \quad (x = 0, 1) \quad . \quad (3)$$

Note that p_{ic} is a one-to-one function of γ_{ic} . For estimation purposes, it is sometimes more convenient to estimate models with unbounded parameters instead of estimating models with bounded parameters. For I items and C classes, $I \cdot C$ item parameters have to be estimated in the case of dichotomous items. In comparison to item response models (1PL model: one parameter, 2PL: two parameters, etc.), this results in many more parameters to be estimated. However, LCMs do not pose the assumption that classes are ordered, and no monotonicity assumptions of item response functions are posed.

Moreover, let $p_c = P(U = c)$ denote the probability that a person is in class c . As for item parameters, a logistic transformation is used to represent the class probabilities p_c by parameters δ_c . More formally, we set

$$p_c = \frac{\exp(\delta_c)}{1 + \sum_{j=2}^C \exp(\delta_j)} \quad (c = 1, \dots, C) \quad , \quad (4)$$

where $\delta_1 = 0$. Because the probabilities sum to one, only $C - 1$ distribution parameters have to be estimated. In total, the saturated distribution of I dichotomous items has $2^I - 1$ free possible parameters, which is represented by $I \cdot C + C - 1$ parameters in the LCM with C classes.

LCMs can be interpreted as pure exploratory models because no structure of item response probabilities among classes is posed. Confirmatory LCMs assume additional equality constraints on item response probabilities ([16]; [17]; [18]; [19]). Like in confirmatory factor analysis, it could be assumed that some items load only on some classes, which translates into equal item response probabilities. Cognitive diagnostic models can be seen as particular confirmatory LCMs [20].

It should be emphasized that a restricted LCM forms the basis of almost all popular latent variable models for discrete item responses that are nowadays very popular. Researcher Anton Formann [14] represented the vector $\gamma = (\gamma_{11}, \dots, \gamma_{1C}, \dots, \gamma_{I1}, \dots, \gamma_{IC})$ of item response functions as linear combinations $\gamma_{ic} = \mathbf{q}_{ic}\boldsymbol{\alpha}$ ($i = 1, \dots, I; c = 1, \dots, C$) using a parameter vector $\boldsymbol{\alpha}$ and known weight

vectors q_{ic} . In addition, the distribution parameter $\delta = (\delta_1, \dots, \delta_C)$ is represented by $\delta_c = w_c \beta$ using a parameter vector β and known weight vectors w_c . The resulting so-called structured LCM [21] includes unidimensional and multidimensional 1PL and 2PL models as special cases as well as mixture item response models [22]. Notably, representing a discrete version of the normal distribution as latent classes only needs two components in α (i.e., the mean and the standard deviation as the first two moments) to represent the vector δ of distribution parameters. Similar generalizations of restricted LCM were proposed by researcher Matthias von Davier [23,24]. In the rest of this article, we will focus on the simple exploratory LCM, although the proposed extension also applies to the more general structured latent class model.

2.2. Exploratory Latent Class Analysis for Polytomous Item Responses

Now assume that there are I polytomous items and each item has $K_i + 1$ nominal categories $0, 1, \dots, K_i$. Item response probabilities are then given as $p_{ikc} = P(X_i = k | U = c)$ and are again transformed into unbounded parameters γ_{ikc} by a logistic transformation. In more detail, it is assumed that

$$P(X_i = x | U = c) = \frac{\exp(\gamma_{ixc})}{1 + \sum_{k=1}^{K_i} \exp(\gamma_{ikc})} \quad (x = 0, 1, \dots, K_i) \quad , \quad (5)$$

where $\gamma_{i0c} = 0$ for all items i and all latent classes c . Instead of estimating $K_i + 1$ probabilities for item i and class c , K_i free parameters γ_{ihc} have to be estimated. If all polytomous items have $K + 1$ categories, the multidimensional contingency table of observations has $(K + 1)^I - 1$ free parameters while in the LCM $I \cdot K \cdot C + C - 1$ parameters are estimated. It should be emphasized that the LCM for polytomous items has many free parameters and is typically much more data hungry than unidimensional and multidimensional item response models for polytomous data.

Like for dichotomous data, restricted LCM were formulated that represented the vector of all item response functions by $\gamma_{ikc} = q_{ikc} \alpha$ ($i = 1, \dots, I; k = 1, \dots, K_i; c = 1, \dots, C$) using a parameter vector α and known weight vectors q_{ikc} [25]. It is advisable to induce some structure on item response functions, especially for polytomous data, because many parameters have to be estimated without any structural assumptions.

In many applications, the allocation of persons to classes should be predicted by person variables Z [11]. In more detail, class probabilities $p_c = P(U = c)$ are replaced by subject-specific conditional probabilities $P(U = c | Z = z)$ (so-called latent class regression). These models further ease the interpretation of latent classes.

3. Regularized Latent Class Analysis

As LCMs are exploratory models, interpretation of results could sometimes be challenging. Moreover, in not too large samples, parameter estimation gets instable, and findings are sometimes not generalizable across different samples. Alternatively, confirmatory latent models could be estimated for obtaining more stable and more interpretable parameter estimates. However, such confirmatory approaches need assumptions that have to be known in advance of the data analysis. Hence, alternative approaches are sought.

Regularized latent class models (RLCM; [26]) estimate item response probabilities under the presupposition that similar item response probabilities in these models are grouped and receive the same value. The main idea of using the regularization technique (see [27] for an overview) to LCMs is that by subtracting an appropriate penalty term from the log-likelihood function, some simpler structure on item response probabilities is posed. Different penalty terms typically result in different estimated parameter structures. In a recent Psychometrika paper, [26] proposed the RLCM for dichotomous item responses. Related work for dichotomous data can be found in [28] and [29].

The regularization technique has also been applied for factor models with continuous items ([30]; [31]) and discrete items ([32]; [33]) in order to fit exploratory factor models with the goal of estimating

as many zero loadings as possible. In this respect, regularization is a viable alternative to factor rotation methods ([34]).

The regularization technique has also been applied to Gaussian mixture models in which cluster means are estimated to be equal for some variables among clusters [35,36]. RLCA is also referred to as penalized latent class analysis (see [37]). Under this label, LCMs are typically meant by that apply regularization to the estimation of regression coefficients of the latent class regression model [37–42]. [43] provide a recent review of applications of the regularization technique in mixture models.

In the following, we describe the RLCM at first for dichotomous items. Afterward, we consider the more complex case of polytomous items in which more possibilities for setting equality constraints among item response probabilities are present.

3.1. Regularized Latent Class Analysis for Dichotomous Item Responses

At first, we consider the case of dichotomous items X_i ($i = 1, \dots, I$). In a RLCM, not all item response probabilities p_{ic} ($c = 1, \dots, C$) are assumed to be unique. [26] subtracted a penalty term from the log-likelihood function that penalizes differences in ordered item response probabilities. In more detail, denote by $p_{i(c)}$ ordered item response probabilities such that $p_{i(1)} \leq p_{i(2)} \leq \dots p_{i(C)}$, and collect all parameters in \mathbf{p}_i^* . Then, Chen and colleagues used the following penalty function for item i

$$\text{Pen}(\mathbf{p}_i^*; \lambda) = \sum_{c=2}^C H_{\text{SCAD}}(p_{i(c)} - p_{i(c-1)}; \lambda) \quad , \quad (6)$$

where H_{SCAD} denotes the smoothly clipped absolute deviation penalty (SCAD; [44]). The SCAD penalty takes a value of zero if $p_{i(c)} - p_{i(c-1)} = 0$ and is positive otherwise. The parameter λ is a regularization parameter that governs the strength of the penalty function. With small values of λ , differences are barely penalized, but with large values of λ , differences are heavily penalized.

If \mathbf{X} denotes the matrix of observed data and \mathbf{p}^* denotes the vector of all ordered item response probability and δ the vector that represents the skill class probabilities, the following function is maximized in [26]:

$$l(\mathbf{p}^*, \delta; \mathbf{X}) - N \sum_{i=1}^I \text{Pen}(\mathbf{p}_i^*; \lambda) \quad , \quad (7)$$

where l denotes the log-likelihood function of the data. By employing a penalty function Pen in the estimation, some item response probabilities are merged, which, in turn, eases the interpretation of resulting latent classes. It should be noted that for estimating model parameters, the regularization parameter λ has to be fixed. In practice, the regularization parameter λ also has to be estimated. Hence, the maximization is performed on a grid of λ values (say, $\lambda = .01, .02, \dots, .30$), and that model is selected that is optimal with respect to some criterion. Typical criteria are the cross-validated log-likelihood or information criteria like the Akaike information criterion (AIC), Bayesian information criterion (BIC), or others [27].

The maximization in Equation 7 is conducted using an expectation-maximization (EM) algorithm (see Section 3.3 for general description). The estimation approach of [26] is implemented in the R package CDM [45,46].

3.1.1. Fused Regularization among Latent Classes

Though the estimation approach of [26] is successful, it is not clear how it could be generalized to polytomous data because it is not evident how item response probabilities of several categories should be ordered. Hence, we propose a different estimation approach. We apply the technique of fused

regularization [47,48] that penalizes all pairwise differences of item response probabilities. In more detail, for a vector \mathbf{p}_i of item response probabilities, we replace the penalty in Equation 6 of [26] by

$$\text{Pen}(\mathbf{p}_i; \lambda) = \sum_{c < d} H_{\text{MCP}}(p_{ic} - p_{id}; \lambda) \quad , \quad (8)$$

where h_{MCP} denotes the minimax concave penalty (MCP; [49]). We do not expect dramatic differences to the SCAD penalty but we would expect less biased estimators than using the often employed least absolute shrinkage and selection operator (LASSO) penalty $H_{\text{LASSO}}(x; \lambda) = \lambda|x|$ (see [27]). By using pairwise differences in Equation 8, item response probabilities are essentially merged into item-specific clusters of values that are equal within each cluster. Hence, the same goal as in [26] is achieved. As explained in Subsection 2.2, our estimation approach uses transformed item response probabilities γ . Therefore, operationally, we replace Equation 8 by

$$\text{Pen}(\boldsymbol{\gamma}_i; \lambda) = \sum_{c < d} H_{\text{MCP}}(\gamma_{ic} - \gamma_{id}; \lambda) \quad . \quad (9)$$

In Figure 1, the LASSO, MCP, and SCAD penalty functions are depicted. It can be seen for x values near to 0, the MCP and the SCAD penalty equal the LASSO penalty (i.e., $f(x) = \lambda|x|$). For sufficiently large x values MCP and SCAD reach an upper asymptote, which is not the case for the LASSO penalty. Hence, for the MCP and SCAD penalty, the penalty is relatively constant for large values of x . This property explains why the MCP and SCAD penalty typically results in less biased estimates. It should be noted that the application of the regularization presupposes some sparse structure in the data for obtaining unbiased estimates. In other words, the true data generating mechanism consists of a sufficiently large number of equal item parameters. If all item response probabilities would be different in the data generating model, employing a penalty that forces many item parameters to be equal to each other would conflict the data generating model.

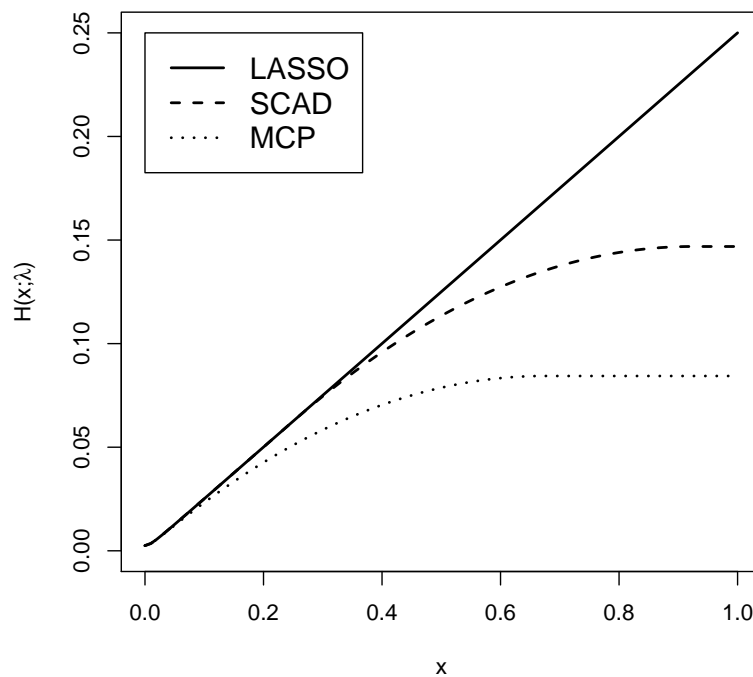


Figure 1. Different penalty functions used in regularization

3.1.2. Hierarchies in Latent Class Models

The RLCM can be used to derive a hierarchy among latent classes. The main idea is depicted in Figure 2. In an RLCM with $C = 4$ classes, a partial order of latent classes is defined. Class 1 is smaller than Classes 2, 3, and 4. Classes 2 and 3 cannot be ordered. Finally, Classes 2 and 3 are smaller than Class 4. More formally, in an RLCM, we define class c to be *smaller* than class d (or: class d is *larger* than class c) if all item response probabilities in class c are at most as large as in class d , i.e., $p_{ic} \leq p_{id}$ for all items $i = 1, \dots, I$. We use the notation $c \preceq d$ to indicate that c is smaller than d . In a test with many items, fulfilling these inequalities for all items might be a too strong requirement. Hence, one weakens the concept of partial ordering a bit. Given a tolerable for at most ι items, we say that class c is *approximately smaller* than class d if $p_{ic} \leq p_{id}$ is fulfilled for at least $I - \iota$ items.

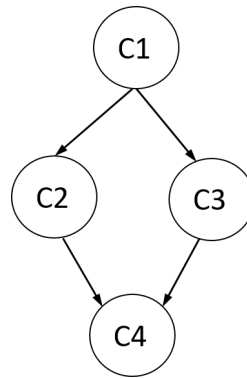


Figure 2. Illustration of a Partial Order with 4 Latent Classes

The partial ordering of latent classes substantially eases the interpretation of the results in RLCMs. [26] used the RLCM to derive partially ordered latent classes in cognitive diagnostic modeling. [50] also applied the RLCM for estimating hierarchies among latent classes (see also [46]). Using the RLCM with an analysis of hierarchies may be considered as a preceding method of confirmatory approaches to latent class modeling.

3.2. Regularized Latent Class Analysis for Polytomous Item Responses

In the following, we propose an extension of RLCM for polytomous item responses. It has been shown that using information from item distractors [1,51] could increase the reliability compared to using only dichotomous item responses that only distinguishes between correct and incorrect item responses. Assume that 0 denotes the category that refers to a correct response and $1, \dots, K_i$ refer to the categories of the distractors. In our parameterization of the LCM for polytomous data (see Subsection 2.2), only parameters γ_{ikc} of distractors k for item i in classes c are parameterized. Given the relatively small sample size of the SPM-LS application data (i.e., $N = 499$), the number of estimated parameters in an unrestricted LCM turn out to be quite large because there are 7 distractors per item. Moreover, it could be supposed that the distractors of an item behave similarly. Hence, it would make sense to estimate some item parameters to be equal to each other.

We now outline alternatives for structural assumptions on item response probabilities. Let us fix item i . For $K_i + 1$ categories and C classes, $K_i \cdot C$ item parameters are modeling (omitting the category 0). Hence, we can distinguish between different strategies to the setting of equalities of item parameters. First, for a fixed category k , one can merge some item response probabilities among classes. This means that some of the differences $\gamma_{ikc} - \gamma_{ikd}$ ($c \neq d$) are zero. Hence, a penalty on differences $\gamma_{ikc} - \gamma_{ikd}$ has to be posed. This is just the penalty as for dichotomous items (see Equation 8), but the regularization is applied for K_i categories instead of one category. Second, for a fixed class c , some item response probabilities among categories could be merged. In this case, one would impose a penalty

on differences $\gamma_{ikc} - \gamma_{ihc}$ ($k \neq h$). Third, penalization among classes and among categories can be simultaneously applied. In the remainder, we outline the different strategies in more detail.

3.2.1. Fused Regularization among Latent Classes

Let $\gamma_{ik*} = (\gamma_{ik1}, \dots, \gamma_{ikC})$ denote the vector of item parameters for item i in category k . Again, let γ_i denote the vector of all item parameters of item i . For a regularization parameter λ_1 and item i , we define the penalty

$$Pen(\gamma_i; \lambda_1) = \sum_{k=1}^{K_i} Pen(\gamma_{ik*}; \lambda_1) = \sum_{k=1}^{K_i} \sum_{c < d} H_{MCP}(\gamma_{ikc} - \gamma_{ikd}; \lambda_1) \quad . \quad (10)$$

As a result, for a category, some item response probabilities will be merged across latent classes. However, the merging of item parameters (also referred to as fusing) is independently applied for all categories of an item. In practice, it is maybe not plausible that all distractors of an item would function differently, and item parameters should be more regularized.

3.2.2. Fused Regularization among Categories

As a second alternative, we now merge categories. Let $\gamma_{i*c} = (\gamma_{i1c}, \dots, \gamma_{iK_ic})$ denote the vector of item parameters for item i in class c . For a regularization parameter λ_2 and an item i , we define the penalty

$$Pen(\gamma_i; \lambda_2) = \sum_{c=1}^C Pen(\gamma_{i*c}; \lambda_2) = \sum_{c=1}^C \sum_{k < h} H_{MCP}(\gamma_{ikc} - \gamma_{ihc}; \lambda_2) \quad . \quad (11)$$

As a result, some of the item response probabilities of categories are set equal to each other. As an outcome of applying this penalty, atypical distractors could be detected. However, by using the penalty in Equation 11, no equalities among latent classes are imposed.

3.2.3. Fused Regularization among Latent Classes and Categories

The apparent idea is to combine the regularization among latent classes and categories. By doing so, the penalties in Equations 10 and 11 have to be added. In more detail, for regularization parameters λ_1 and λ_2 , we use the penalty

$$Pen(\gamma_i; \lambda_1, \lambda_2) = \sum_{k=1}^{K_i} \sum_{c < d} H_{MCP}(\gamma_{ikc} - \gamma_{ikd}; \lambda_1) + \sum_{c=1}^C \sum_{k < h} H_{MCP}(\gamma_{ikc} - \gamma_{ihc}; \lambda_2) \quad . \quad (12)$$

It can be seen that the penalty in Equation 12 now depends on two regularization parameters. In the estimation, the one-dimensional grid of regularization parameters has then to be substituted by a two-dimensional grid. This substantially increases the computational demand.

3.2.4. Fused Group Regularization among Categories

We can now proceed to pose additional structural assumptions on item parameters. One could suppose that two distractors k and h of item i show the same behavior. In the RLCM, this means that $\gamma_{ikc} - \gamma_{ihc} = 0$ holds for all classes $c = 1, \dots, C$. The group regularization technique allows us to estimate all parameters in a subset of parameters to be zero (see [52] for a review). A fused group regularization approach presupposes that either all differences $\gamma_{ikc} - \gamma_{ihc}$ equal zero or all differences are estimated to be different from zero [53,54]. This property can be achieved by substituting a norm of the difference of the two vectors in the penalty. In more detail, one considers the penalty

$$Pen(\gamma_i; \lambda_1) = \sum_{k < h} H_{MCP}(\|\gamma_{ik*} - \gamma_{ih*}\|; \lambda_1) \quad (13)$$

where for a vector $\mathbf{x} = (x_1, \dots, x_p)$, the norm $\|\mathbf{x}\|$ is defined as $\|\mathbf{x}\| = \sqrt{p} \sqrt{\sum_{k=1}^p x_k^2}$. In practice, using the penalty in Equation 13 could provide a more parsimonious estimation than the penalty defined in Equation 12. In principle, model comparisons can be carried out to decide which assumption is better represented in the data.

3.2.5. Fused Group Regularization among Classes

Alternatively, one could also assume that latent classes function the same among classes. In the RLCM, then it would hold that $\gamma_{ikc} - \gamma_{ikd} = 0$ for all categories $k = 1, \dots, K_i$. A fused group regularization results in the property that either all item parameters of classes c and d are equal to each other or all estimated to be different from each other. The following penalty is used in this case:

$$Pen(\gamma_i; \lambda_2) = \sum_{c < d} H_{MCP}(\|\gamma_{i*c} - \gamma_{i*d}\|; \lambda_2) \quad (14)$$

3.3. Estimation

We now describe the estimation of the proposed RLCM for polytomous data. Let $\mathbf{X} = (x_{nik})$ denote the observed dataset where x_{nik} equals 1 if person n ($n = 1, \dots, N$) chooses category k for item i . Let γ_i denote item parameters of item i and the vector that contains item parameters of all items. The vector δ represents the skill class distribution. Furthermore, let $p_{ic}(x; \gamma_i) = P(X_i = x | U = c)$ and $p_c(\delta) = P(U = c)$.

Following [26] and [33], an EM algorithm is applied for estimating model parameters. The complete-data log-likelihood function is given

$$l_{\text{com}}(\gamma, \delta, \mathbf{U}) = \sum_{n=1}^N \sum_{i=1}^I \sum_{k=1}^{K_i} \sum_{c=1}^C x_{nik} u_{nc} \log p_{ic}(k; \gamma_i) + \sum_{n=1}^N \sum_{c=1}^C u_{nc} \log p_c(\delta) \quad , \quad (15)$$

where $\mathbf{u}_n = (u_{n1}, \dots, u_{nC})$ is the vector of latent class indicators for person n . It holds that $u_{nc} = 1$ if person n is located in class c . Obviously, the true class membership is unknown and, hence, 15 cannot be used for maximization.

In the EM algorithm, the estimation of l_{com} is replaced by the expected complete-data log-likelihood function by integrating over the posterior distribution. In more detail, unobserved values u_{nc} are replaced by their conditional expectations:

$$u_{nc}^* = E(u_{nc} | \mathbf{x}_n; \gamma^{(t)}, \delta^{(t)}) = \frac{p_c(\delta^{(t)}) \prod_{i=1}^I \prod_{k=1}^{K_i} p_{ic}(k; \gamma_i^{(t)})^{x_{nki}}}{\sum_{d=1}^D p_d(\delta^{(t)}) \prod_{i=1}^I \prod_{k=1}^{K_i} p_{id}(k; \gamma_i^{(t)})^{x_{nki}}} \quad (c = 1, \dots, C) \quad , \quad (16)$$

where $\gamma^{(t)}$ and $\delta^{(t)}$ are parameter estimates from a previous iteration t . The EM algorithm alternates between the E-step and the M-step. By replacing the unobserved values u_{ni} by their expected values u_{ni}^* , the following Q-function is obtained that is used for maximization in the M-step

$$Q(\gamma, \delta | \gamma^{(t)}, \delta^{(t)}) = \sum_{n=1}^N \sum_{i=1}^I \sum_{k=1}^{K_i} \sum_{c=1}^C x_{nik} u_{nc}^* \log p_{ic}(k; \gamma_i) + \sum_{n=1}^N \sum_{c=1}^C u_{nc}^* \log p_c(\delta) \quad . \quad (17)$$

From this Q-function, the penalty function is subtracted such that the following function is minimized for some regularization parameter λ in the M-step

$$Q(\gamma, \delta | \gamma^{(t)}, \delta^{(t)}) - N \sum_{i=1}^I Pen(\gamma_i; \lambda) \quad . \quad (18)$$

It can be seen that item parameters γ_i are separately obtained for each item i in the M-step and one maximizes

$$\sum_{n=1}^N \sum_{k=1}^{K_i} \sum_{c=1}^C x_{nik} u_{nc}^* \log p_{ic}(k; \gamma_i) - N \text{Pen}(\gamma_i; \lambda) \quad . \quad (19)$$

Latent class probability parameters δ are also obtained independently from item parameters in the M-step.

The penalty function *Pen* turns out to be non-differentiable. Here, we use a differentiable approximation of the penalty function ([55]; see also [56]). As it is well known that the log-likelihood function in LCMs is prone to multiple maxima, using multiple starting values in the estimation is advised.

The described EM algorithm is included in an experimental version of the function `regpolca()` in the R package `sirt` [57]. The function is under current development for improving computational efficiency.

4. Simulated Data Illustration

Before we illustrate the application of the method to the SPM-LS dataset, we demonstrate the technique using a simulated data set. This helps to better understand the proposed method of regularized latent class modeling under ideal conditions.

4.1. Dichotomous Item Responses

4.1.1. Data Generation

First, we consider the case of dichotomous items. To mimic the situation in the SPM-LS dataset, we also chose $I = 12$ items for simulating a dataset. Moreover, to reduce sampling uncertainty somewhat, a sample size of $N = 1000$ subjects was chosen. There were $C = 4$ latent classes with true class probabilities .30, .20, .10, and .40. In Table 1, we present the item response probabilities with each cluster. We only specified parameters for six items and duplicated these parameters for the remaining six items in the test. It can be seen in Table 1 that many item response probabilities were set equal to each other. Indeed, for the first four items, there are only two instead of four unique probabilities. Moreover, it is evident from Table 1 that the four classes are partially ordered. The first class has the lowest probabilities for all items and is, therefore, the smallest class that consists of the least proficient subjects. The fourth class has the highest probabilities, constitutes the largest class, and contains the most proficient subjects.

Table 1. Data Illustration Dichotomous Data: True Item Response Probabilities p_{ic}

Item	Class			
	1	2	3	4
1,7	.10	.82	.82	.82
2,8	.22	.88	.88	.88
3,9	.16	.79	.16	.79
4,10	.25	.85	.25	.85
5,11	.10	.10	.46	.91
6,12	.22	.22	.22	.79

4.1.2. Results

In the first step, we estimated exploratory latent class models with $C = 2, 3, 4, 5$, and 6 classes. The model comparison is presented in Table 6. While the decision based on the AIC was ambiguous and selected the incorrect number of classes, the BIC correctly selected model with $C = 4$ latent classes. This observation is consistent with the literature that argues that model selection in LCMs should be based on the BIC instead of the AIC [11].

Table 2. Data Illustration Dichotomous Data: Model Comparison for Exploratory LCMs

C	#np	AIC	BIC
2	25	13636	13759
3	38	13169	13356
4	51	12979	13229
5	64	12981	13295
6	76	12976	13349

Note. C = number of classes; #np = number of estimated parameters.

In the solution with four classes, estimated class probabilities were .290, .204, .120, and .386, respectively, which closely resembled the data generating values. In Table 3, estimated item response probabilities are shown. The estimates were very similar to the data generating parameters that are presented in Table 1. It can be seen that some deviations from the simulated equality constraints are obtained. It is important to emphasize that latent class solutions are not invariant with respect to their class labels (so-called label switching). Class labels in the estimated model have to be permuted in order to match the class label in the simulated data.

Table 3. Data Illustration Dichotomous Data: Estimated Item Response Probabilities in Exploratory LCM with C = 4 Classes

Item	Class			
	1	2	3	4
1	.08	.79	.79	.82
2	.20	.84	.89	.91
3	.15	.76	.19	.81
4	.27	.90	.29	.86
5	.10	.09	.44	.92
6	.23	.23	.30	.77
7	.07	.79	.82	.85
8	.24	.87	.91	.87
9	.14	.82	.18	.81
10	.19	.90	.42	.83
11	.10	.13	.54	.89
12	.25	.19	.19	.77

Finally, we estimated the RLCM for regularization parameters from 0.01 to 1.00 in steps of 0.01 for C = 4 classes in order to obtain the best-fitting solution. The regularization parameter $\lambda = .21$ provided the best-fitting model in terms of the BIC (BIC = 13104, AIC = 12957). Notably, this model showed a substantially better model fit than the exploratory LCM with four classes due to the more parsimonious estimation of item parameters. In the model with $\lambda = .21$, in total, 21 item parameters were regularized (i.e., they were set equal to item parameters in other classes for the respective item), resulting in 30 freely estimated model parameters. With respect to the AIC, the best-fitting model was obtained for $\lambda = .15$ (BIC = 13110, AIC = 12953). This model resulted in 19 regularized item parameters and 32 freely estimated item parameters. The model selected by the minimal BIC ($\lambda = .21$) resulted in estimated class probabilities of .29, .21, .11, and .39. Estimated item response probabilities (shown in Table 4) demonstrate that the equality constraints that were posed in the data generating were correctly identified in the estimated model.

Table 4. Data Illustration Dichotomous Data: Estimated Item Response Probabilities in the RLCM with $C = 4$ Classes Based on the Minimal BIC ($\lambda = .21$)

Item	Class			
	1	2	3	4
1	.08	.80	.80	.80
2	.20	.88	.88	.88
3	.15	.79	.15	.79
4	.27	.87	.27	.87
5	.09	.09	.44	.92
6	.23	.23	.23	.76
7	.07	.82	.82	.82
8	.24	.87	.87	.87
9	.14	.81	.14	.81
10	.19	.85	.40	.85
11	.11	.11	.55	.89
12	.22	.22	.22	.76

Lastly, we want to illustrate the behavior of regularization. For the sequence of specified regularization parameters λ in the estimation, the estimated item response probabilities $p_{ic}(\lambda)$ can be plotted. Such a plot is also referred to as a regularization path [27]. With very small λ values, no classes were merged, and all item parameters were estimated differently from each other. With increasing values of λ , item parameters were subsequently merged. For Item 1 and Classes 2, 3, and 4, the regularization path is shown in Figure 3. At first, item parameters of Class 2 and Classes 3 and 4 are merged at $\lambda = .04$. Afterwards, all three item response probabilities are merged at $\lambda = .09$.

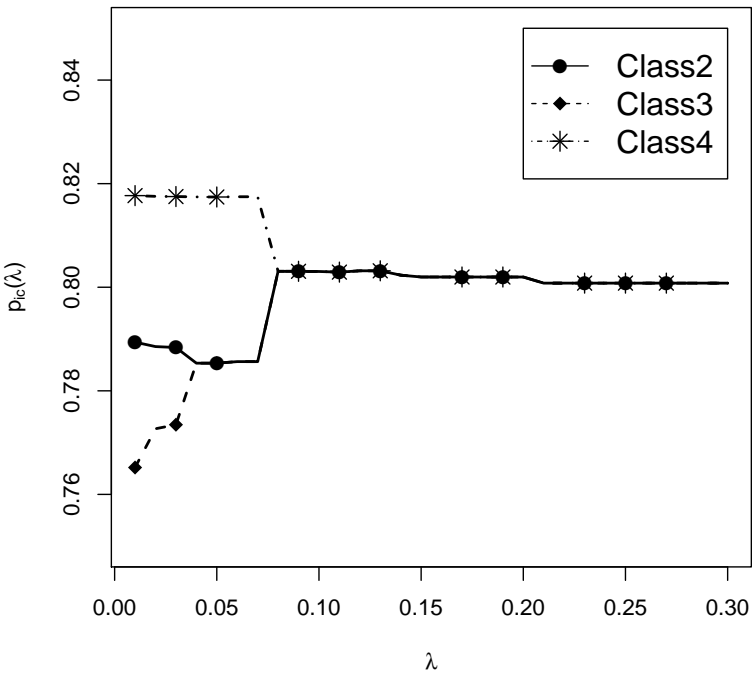


Figure 3. Data illustration dichotomous data: Regularization path for estimated item response probabilities for Item 1 for Classes 2, 3, 4 for the four-class solution

4.2. Polytomous Item Responses

4.2.1. Data Generation

Now, we simulate illustrated data with polytomous item responses with 12 items, each item possessing four categories (i.e., $K_i = 3$). The first category (i.e., Category 0) refers to the correct category, while categories 1, 2, and 3 refer to distractors of the item. As in the case of dichotomous data, item response probabilities were .30, .20, .10, and .40, and $N = 1000$ subjects were simulated. Again, we specified item parameters for the first six items and replicated the parameters for the remaining six items. In Table 5, true item response probabilities are shown that were used for generating the dataset. It is evident that the item response probabilities are strongly structured. All distractors of Items 1 and 5 function precisely the same. For Item 2, Category 1, and Category 2 show the same behavior. Category 3 only shows a differential behavior in Classes 2 and 4. At the other extreme, all item response probabilities differ for Item 6 among classes and categories. It can be expected that an RLCM will result in a substantial model improvement compared to an exploratory LCM without equality constraints.

Table 5. Data Illustration Polytomous Data: True Item Response Probabilities p_{ic}

Item	Cat	Class				Item	Cat	Class			
		1	2	3	4			1	2	3	4
1,7	0	.10	.82	.82	.82	4,10	0	.25	.85	.25	.85
1,7	1	.30	.06	.06	.06	4,10	1	.35	.03	.35	.03
1,7	2	.30	.06	.06	.06	4,10	2	.20	.03	.20	.03
1,7	3	.30	.06	.06	.06	4,10	3	.20	.09	.20	.09
2,8	0	.22	.88	.88	.88	5,11	0	.10	.10	.46	.91
2,8	1	.26	.05	.04	.06	5,11	1	.30	.30	.18	.03
2,8	2	.26	.05	.04	.06	5,11	2	.30	.30	.18	.03
2,8	3	.26	.02	.04	.00	5,11	3	.30	.30	.18	.03
3,9	0	.16	.79	.16	.79	6,12	0	.22	.22	.22	.79
3,9	1	.28	.11	.28	.11	6,12	1	.24	.23	.22	.06
3,9	2	.33	.05	.33	.05	6,12	2	.20	.17	.12	.04
3,9	3	.23	.05	.23	.05	6,12	3	.34	.38	.44	.11

4.2.2. Results

At first, we fit exploratory LCMs with 2, 3, 4, 5, and 6 classes. Based on the information criteria presented in Table 6, the correct model with $C = 4$ latent classes was selected. However, the difference in model improvement by moving from 3 to 4 classes would be considered as negligible (i.e., a BIC difference of 3) in practice. Estimated latent class probabilities in the model with 4 latent classes were estimated as .29, .21, .12, and .38. Estimated item response probabilities are shown in Table A1 in Appendix A.

Table 6. Data Illustration Polytomous Data: Model Comparison for Exploratory LCMs

C	#np	AIC	BIC
2	72	25082	25440
3	107	24616	25151
4	143	24431	25148
5	179	24439	25337
6	215	24444	25524

Note. C = number of classes; #np = number of estimated parameters.

In the next step, different RLCMs for polytomous data were specified. As explained in Subsection 3.2, one can regularize differences in item parameters among classes (using a regularization parameter λ_1), among categories (using a regularization parameter λ_2), or both (using both regularization parameters or applying fused group regularization). We fitted the five approaches (Approaches R1, ..., R5) that were introduced in Subsection 3.2 to the simulated data using unidimensional and two-dimensional grids of regularization parameters. For each of the regularization approaches, we selected the model with minimal BIC.

Table 7. Data Illustration Polytomous Data: Model Comparison for Different RLCMs with 4 Classes

Appr.	Fused	Eq.	C	λ_1	λ_2	#np	#nreg	BIC
R1	Class	10	4	.31	—	84	63	24982
R2	Cat	11	4	—	.18	68	79	24689
R3	Cat & Class	12	4	.40	.15	44	103	24836
R4	Grouped Cat	13	4	.45	—	82	65	24777
R5	Grouped Class	14	4	.65	—	79	67	24776

Note. Appr. = approach; Cat = category; Eq. = equation for regularization penalty in Subsection 3.2; C = number of classes; #np = number of estimated parameters; #nreg = number of regularized item parameters.

It can be seen that the model with the fused penalty on item categories fitted the model best (Approach R2: BIC = 24689). In this model, 79 item parameters are regularized. The decrease in BIC compared to an exploratory LCM is substantial. From these findings, it follows for this dataset that it is important to fuse item parameters among categories instead of among classes. The best-fitting model when using a simultaneous penalty for classes and categories (Approach R3: BIC = 24836) outperformed the model in which only parameters were fused among classes (Approach R1: BIC = 24932). However, it was inferior to the model with fusing among categories. Notably, the largest number of regularized parameters (#nreg = 103) was obtained for Approach R3. The fused grouped regularization approaches (Approaches R4 and R5) also improved fit compared to an unrestricted exploratory LCM but were also inferior to R2. The reason might be that applying group regularization results in the extreme decision that either all item parameters are equal or all are different. In contrast, fused regularization Approaches R1, R2, and R3 allow the situation in which only some of the item parameters are estimated to be equal to each other.

For the best-fitting model of Approach R2 (i.e., fusing among categories), estimated class probabilities were .29, .22, .10, and .39, respectively. Estimated item response probabilities from this model are shown in Table 8. It can be seen that model estimation was quite successful in identifying parameters that were equal in the data generating model.

Table 8. Data Illustration Polytomous Data: Estimated Item Response Probabilities in the RLCM with $C = 4$ Classes and Fused Regularization Among Classes Based on the Minimal BIC

Item	Cat	Class				Item	Cat	Class				Item	Cat	Class			
		1	2	3	4			1	2	3	4			1	2	3	4
1	0	.07	.79	.82	.82	5	0	.10	.13	.46	.92	9	0	.13	.82	.10	.82
1	1	.31	.07	.06	.06	5	1	.30	.29	.18	.02	9	1	.29	.06	.30	.06
1	2	.31	.07	.06	.06	5	2	.30	.29	.18	.02	9	2	.29	.06	.30	.06
1	3	.31	.07	.06	.06	5	3	.30	.29	.18	.04	9	3	.29	.06	.30	.06
2	0	.22	.85	.87	.91	6	0	.22	.24	.30	.76	10	0	.20	.89	.37	.82
2	1	.26	.05	.01	.06	6	1	.26	.18	.32	.07	10	1	.42	.05	.25	.05
2	2	.26	.05	.01	.03	6	2	.26	.18	.06	.03	10	2	.19	.01	.13	.03
2	3	.26	.05	.11	.00	6	3	.26	.40	.32	.14	10	3	.19	.05	.25	.10
3	0	.16	.74	.19	.80	7	0	.07	.79	.85	.85	11	0	.10	.16	.55	.91
3	1	.28	.16	.27	.10	7	1	.31	.09	.05	.05	11	1	.30	.28	.15	.03
3	2	.28	.05	.27	.05	7	2	.31	.09	.05	.05	11	2	.30	.28	.15	.03
3	3	.28	.05	.27	.05	7	3	.31	.03	.05	.05	11	3	.30	.28	.15	.03
4	0	.26	.88	.28	.85	8	0	.25	.86	.91	.88	12	0	.24	.20	.20	.76
4	1	.36	.05	.24	.04	8	1	.25	.06	.03	.06	12	1	.21	.21	.35	.10
4	2	.19	.02	.24	.02	8	2	.25	.06	.03	.06	12	2	.21	.21	.10	.04
4	3	.19	.05	.24	.09	8	3	.25	.02	.03	.00	12	3	.34	.38	.35	.10

Note. Cat = category.

5. Application of the SPM-LS Data

In this section, we illustrate the use of RLCM to the SPM-LS dataset.

5.1. Method

According to the topic of this special issue, the publicly available dataset from the [1] study is reanalyzed. The original study compared various parametric item response models (i.e., 1PL, 2PL, 3PL, 4PL, and nested logit model) performed on a dataset comprised of $N = 499$ students (214 males and 285 females) aged between 19 and 24. The analyzed data consist of responses on the 12 most difficult SPM items and are made freely available at <https://data.mendeley.com/datasets/h3yhs5gy3w/1>. For details regarding the data gathering procedure, we refer to [1].

Each of the $I = 12$ items has one correct category and $K_i = 7$ distractors. To be consistent with the notation introduced in the paper and to ease interpretation of the results, we recoded the original dataset when using polytomous item responses. First, we scored the correct response as Category 0. Second, we recoded the order of distractors according to their frequency. In more detail, Category 1 in our rescored dataset is the most attractive distractor (i.e., most frequent distractor), while Category 7 is the least attractive distractor. The relative frequencies and references to categories of the original dataset are shown in Table 9. It could be supposed that there some especially attractive distractors for each item. However, many item category frequencies turned out to be relatively similar such that it could be that they would also function homogeneously among latent classes. We also analyze the SPM-LS dataset in its dichotomous version in which Category 1 is scored as correct, and Category 0 summarizes responses of all distractors.

Table 9. SPM-LS Polytomous Data: Percentage Frequencies and Recoding Table

Item	Cat0	Cat1	Cat2	Cat3	Cat4	Cat5	Cat6	Cat7
SPM1	76.0 (7)	13.6 (3)	3.0 (1)	2.4 (4)	2.2 (6)	2.0 (2)	0.8 (5)	—
SPM2	91.0 (6)	3.0 (3)	2.4 (4)	2.2 (1)	0.8 (5)	0.4 (7)	0.2 (2)	—
SPM3	80.4 (8)	8.0 (2)	4.2 (6)	2.0 (4)	1.8 (3)	1.6 (5)	1.2 (7)	0.8 (1)
SPM4	82.4 (2)	5.6 (3)	3.2 (5)	2.6 (1)	2.2 (8)	1.8 (6)	1.2 (7)	1.0 (4)
SPM5	85.6 (1)	3.8 (2)	3.0 (3)	2.6 (7)	1.8 (6)	1.6 (5)	1.0 (4)	0.6 (8)
SPM6	76.4 (5)	7.0 (4)	5.2 (6)	3.0 (3)	2.8 (7)	2.6 (8)	2.0 (2)	1.0 (1)
SPM7	70.1 (1)	6.6 (4)	5.8 (5)	5.4 (3)	4.4 (8)	3.4 (6)	2.4 (7)	1.8 (2)
SPM8	58.1 (6)	7.6 (1)	7.0 (3)	6.6 (8)	6.4 (2)	6.2 (5)	5.8 (7)	2.2 (4)
SPM9	57.3 (3)	12.0 (5)	9.0 (1)	7.2 (4)	6.6 (8)	4.0 (7)	3.0 (2)	0.8 (6)
SPM10	39.5 (2)	17.2 (6)	11.2 (7)	8.0 (3)	7.8 (8)	7.4 (5)	6.0 (4)	2.8 (1)
SPM11	35.7 (4)	14.0 (1)	13.8 (7)	9.8 (5)	9.4 (6)	8.0 (3)	6.6 (2)	2.6 (8)
SPM12	32.5 (5)	15.4 (2)	14.2 (3)	10.4 (1)	8.2 (4)	8.2 (7)	7.4 (6)	3.6 (8)

Note. Numbers in parentheses denote the original item category.

For the dataset with dichotomous items, the exploratory LCM and the RLCM is fitted for 2 to 6 latent classes. Model selection is conducted based on the BIC. For the dataset with rescored polytomous items, we use the same number of classes for estimating the exploratory LCM. For the RLCM, we applied the fused regularization approach with respect to classes (Subsection 3.2.1), categories (Subsection 3.2.2), and to classes and categories in a simultaneous manner (Subsection 3.2.3).

5.2. Results

5.2.1. Results for Dichotomous Item Responses

For the SPM-LS dataset with dichotomous items, a series of exploratory LCMs and RLCMs with 2 to 6 classes was fitted. According to the BIC, an exploratory LCM with 4 classes would be selected. When RLCMs were fitted, a model with 5 classes would be selected that has 19 regularized item parameters.

Table 10. SPM-LS Dichotomous Data: Model Comparison for Exploratory LCMs and RLCM

	C	λ	#np	#nreg	BIC
LCM	2	0	25	0	5973
	3	0	38	0	5721
	4	0	51	0	5680
	5	0	64	0	5696
	6	0	77	0	5694
RLCM	2	.01	25	0	5973
	3	.33	35	3	5715
	4	.38	39	12	5643
	5	.29	45	19	5621
	6	.53	45	32	5620

Note. C = number of classes; λ = regularization parameter of selected model with minimal BIC; #np = number of estimated parameters; #nreg = number of regularized parameters.

In Table 11, item response probabilities and skill class probabilities for the RLCM with $C = 5$ classes are shown. By considering the average item response probabilities per skill class $\bar{p}_{\bullet c} = (\sum_{i=1}^I p_{ic})/I$, Class C1 (12% frequency) is the least performing and Class C5 (37% frequency) the best performing class. Class C3 (40% frequency) can be seen as an intermediate class. Classes C2 and C4 are relatively rare. Compared to the medium Class C3, students in Class C2 had a particularly bad performance at Items 3, 6, and 11, but outperformed them on Items 7, 8, and 12. Students in Class C4 showed perfect performance on Items 8 and 9, but notably worse performance on Items 10 and

11. Interestingly, one can define a partial order on the classes if we allow at most two violations of inequality conditions. In Figure 4, this partial order is depicted. The arrow from Class C1 to Class C3 means that C1 is smaller than C3. There are arrows with particular labels that indicate violations of the partial order. For example, C1 is approximately smaller than C2, and Items 3 and 11 violate the ordering property.

Table 11. SPM-LS Dichotomous Data: Estimated Item Probabilities and Latent Class Probabilities for Best Fitting RLCM with $C = 5$ Latent Classes

	Class				
	C1	C2	C3	C4	C5
p_c	.12	.04	.40	.07	.37
Item					
SPM1	.39	.39	.83	.83	.83
SPM2	.57	.57	.99	.86	.99
SPM3	.33	.00	.86	.96	.96
SPM4	.05	1.00	.91	.60	1.00
SPM5	.08	1.00	.96	.77	1.00
SPM6	.07	.07	.85	.85	.97
SPM7	.20	.83	.58	.83	.95
SPM8	.06	.69	.36	1.00	.90
SPM9	.16	.34	.34	1.00	.90
SPM10	.00	.23	.23	.00	.79
SPM11	.14	.00	.14	.00	.77
SPM12	.11	.62	.11	.11	.62
$\bar{p}_{\bullet c}$.18	.48	.60	.65	.89

Note. p_c = skill class probability; $\bar{p}_{\bullet c}$ = average of item probabilities within class c .

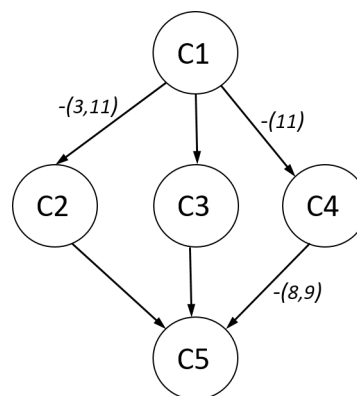


Figure 4. SPM-LS dichotomous data: Partial order for latent class from RLCM

5.2.2. Results for Polytomous Item Responses

We now only briefly discuss the findings for the analysis of the SPM-LS dataset based on polytomous item responses. For the exploratory latent class models, the model with just 2 latent classes would be selected according to the BIC. However, the model with 6 latent classes would be selected according to the AIC. Given a large number of estimated item parameters, applying the RLCM seems to be required for obtaining a parsimonious model. The best-fitting model was obtained with $C = 3$ classes by fusing categories with a regularization parameter of $\lambda_2 = .24$. Classes C1 (28% frequency) and C2 (5% frequency) have low performance, while Class C3 is the high-performing class (67% frequency). As an illustration, we provide in Table 12 estimated item probabilities for the last three items. It can be seen that some of the categories are fused such that they have equal item response probabilities within a latent class. All item parameters are shown in Table A2 in Appendix B.

Table 12. SPM-LS Polytomous Data: Estimated Item Response Probabilities and Latent Class Probabilities for Best-Fitting RLCM with $C = 3$ Latent Classes for Items SPM10, SPM11 and SPM12

Item	Cat	C1	C2	C3	Item	Cat	C1	C2	C3	Item	Cat	C1	C2	C3
SPM10	0	.08	.00	.56	SPM11	0	.11	.26	.48	SPM12	0	.03	.24	.45
SPM10	1	.14	.12	.19	SPM11	1	.18	.43	.10	SPM12	1	.14	.09	.17
SPM10	2	.26	.40	.03	SPM11	2	.21	.00	.12	SPM12	2	.19	.44	.10
SPM10	3	.10	.08	.07	SPM11	3	.15	.12	.07	SPM12	3	.23	.03	.06
SPM10	4	.13	.00	.06	SPM11	4	.14	.00	.08	SPM12	4	.12	.00	.07
SPM10	5	.10	.08	.06	SPM11	5	.08	.19	.07	SPM12	5	.14	.00	.06
SPM10	6	.10	.32	.03	SPM11	6	.07	.00	.07	SPM12	6	.07	.20	.07
SPM10	7	.09	.00	.00	SPM11	7	.06	.00	.01	SPM12	7	.08	.00	.02

Note. Cat = category.

At the time of writing, results for polytomous data for the SPM-LS does not seem to be very consistent with those for dichotomous data. It could be the large number of parameters to be estimated (several hundred depending on the number of classes) for the relatively small sample size of $N = 499$ is critical. Other research has also shown that regularization methods for LCMs need sample sizes of at least 1000 or even more for performing satisfactorily [58].

6. Discussion

In this article, we proposed an extension of regularized latent class analysis to polytomous item responses. We have shown using the simulated data illustration and the SPM-LS dataset that fusing among classes or categories can be beneficial in terms of model parsimony and interpretation. Often, conceptualizing substantive questions as latent classes led researchers to easier to think in types of persons. This interpretation is not apparent in latent variables with continuous latent variables.

In our regularization approach to polytomous data, we based regularization penalties on distractors of items. Hence, the correct item response serves as a reference category. In LCA applications in which the definition of a reference category cannot be done, the regularization approach has certainly to be modified. Note that for $K_i + 1$ categories, only K_i item parameters per class can be independently estimated. Alternatively, a sum constraint $\sum_{k=0}^{K_i} \gamma_{ikc} = 0$ could be posed if γ_{ikc} ($k = 0, 1, \dots, K_i$ denote the item parameters of item i of category k in class c). Such constraints can be replaced by adding ridge-type penalties of the form $\lambda_3 \sum_{k=0}^{K_i} \gamma_{ikc}^2$ to the fused regularization penalty, where λ_3 is another regularization parameter. By squaring item parameters in the penalty function, they are uniformly shrunk to zero in the estimation.

It should be noted that similar regularization approaches have been studied for cognitive diagnostic models [58–62]. These kinds of models pose measurement models on D dichotomous latent variables. These D latent variables constitute 2^D latent classes. Also, in this model class, the modeling of violations of the local independence assumption in LCA has been of interest [63,64].

Previous articles on the SPM-LS dataset also used distractor information by employing the nested logit model (NLM; [1]). The NLM is also very data-hungry, given the low sample size of the dataset. It has been argued that reliability can be increased by using distractor information [1,51]. It should be noted that this is only true to the extent that item parameters can be reliably estimated. For $N = 499$ in the SPM-LS dataset, this will probably be not the case. Regularized estimation approaches could circumvent estimation issues (see [56] for a similar approach in the nominal response model).

Finally, we would like to emphasize that the regularization approaches can be interpreted as empirical Bayesian approaches that employ hierarchical prior distributions on item parameters [65]. It can be expected that Bayesian variants of RLCMs are competitive to EM-based estimation, especially for small(er) samples.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AIC	Akaike information criterion
BIC	Bayesian information criterion
EM	expectation maximization
LASSO	least absolute shrinkage and selection operator
LCA	latent class analysis
LCM	latent class model
NLM	nested logit model
RLCA	regularized latent class analysis
RLCM	restricted latent class model
SPM-LS	last series of Raven’s standard progressive matrices

Appendix A. Additional Results for Simulated Data Illustration with Polytomous Item Responses

In Table A1, estimated item response probabilities for the exploratory LCM with 4 latent classes are shown.

Table A1. Data Illustration Polytomous Data: Estimated Item Response Probabilities in the Exploratory LCM with C = 4 Classes

Class						Class						Class					
Item	Cat	1	2	3	4	Item	Cat	1	2	3	4	Item	Cat	1	2	3	4
1	0	.08	.78	.80	.82	5	0	.10	.11	.45	.92	9	0	.14	.82	.14	.81
1	1	.33	.08	.07	.07	5	1	.34	.27	.23	.02	9	1	.27	.08	.26	.08
1	2	.31	.05	.05	.06	5	2	.32	.31	.17	.02	9	2	.32	.05	.33	.05
1	3	.29	.08	.08	.06	5	3	.24	.31	.15	.04	9	3	.27	.05	.27	.05
2	0	.20	.84	.89	.91	6	0	.22	.24	.31	.77	10	0	.19	.89	.40	.83
2	1	.30	.06	.00	.06	6	1	.23	.17	.16	.07	10	1	.42	.05	.28	.05
2	2	.28	.06	.01	.03	6	2	.23	.21	.06	.03	10	2	.19	.01	.12	.03
2	3	.23	.03	.10	.00	6	3	.31	.39	.47	.13	10	3	.20	.05	.20	.10
3	0	.15	.76	.20	.80	7	0	.07	.79	.83	.84	11	0	.10	.13	.55	.90
3	1	.25	.15	.34	.10	7	1	.34	.10	.06	.05	11	1	.32	.28	.10	.03
3	2	.36	.06	.18	.05	7	2	.31	.07	.06	.06	11	2	.26	.34	.17	.03
3	3	.24	.03	.28	.06	7	3	.28	.03	.06	.05	11	3	.32	.25	.18	.04
4	0	.27	.89	.30	.86	8	0	.24	.86	.91	.88	12	0	.25	.19	.21	.77
4	1	.35	.04	.28	.04	8	1	.23	.06	.05	.06	12	1	.23	.24	.27	.08
4	2	.19	.02	.21	.01	8	2	.24	.06	.02	.06	12	2	.19	.20	.10	.04
4	3	.19	.05	.22	.09	8	3	.28	.02	.03	.00	12	3	.34	.38	.42	.12

Note. Cat = category.

Appendix B. Additional Results for SPM-LS Dataset with Polytomous Item Responses

Table A2 shows all estimated item response probabilities for the SPM-LS dataset with polytomous items for the best fitting RLCM with C = 3 classes.

Table A2. SPM-LS Polytomous Data: Estimated Item Response Probabilities and Latent Class Probabilities for Best-Fitting RLCM with C = 3 Latent Classes

Item	Cat	C1	C2	C3	Item	Cat	C1	C2	C3	Item	Cat	C1	C2	C3
SPM1	0	.73	.15	.82	SPM5	0	.72	.04	.98	SPM9	0	.25	.22	.73
SPM1	1	.11	.48	.12	SPM5	1	.05	.48	.00	SPM9	1	.14	.03	.12
SPM1	2	.06	.01	.02	SPM5	2	.03	.32	.01	SPM9	2	.17	.00	.07
SPM1	3	.07	.04	.01	SPM5	3	.08	.00	.00	SPM9	3	.12	.32	.03
SPM1	4	.00	.24	.01	SPM5	4	.06	.00	.00	SPM9	4	.19	.15	.01
SPM1	5	.01	.08	.02	SPM5	5	.02	.08	.01	SPM9	5	.06	.00	.03
SPM1	6	.02	.00	.00	SPM5	6	.02	.08	.00	SPM9	6	.06	.16	.01
SPM1	7	.00	.00	.00	SPM5	7	.02	.00	.00	SPM9	7	.01	.12	.00
SPM2	0	.87	.32	.97	SPM6	0	.51	.08	.92	SPM10	0	.08	.00	.56
SPM2	1	.02	.12	.03	SPM6	1	.10	.36	.04	SPM10	1	.14	.12	.19
SPM2	2	.02	.36	.00	SPM6	2	.11	.00	.03	SPM10	2	.26	.40	.03
SPM2	3	.07	.00	.00	SPM6	3	.09	.00	.01	SPM10	3	.10	.08	.07
SPM2	4	.01	.12	.00	SPM6	4	.06	.24	.00	SPM10	4	.13	.00	.06
SPM2	5	.00	.08	.00	SPM6	5	.06	.20	.00	SPM10	5	.10	.08	.06
SPM2	6	.01	.00	.00	SPM6	6	.05	.08	.00	SPM10	6	.10	.32	.03
SPM2	7	.00	.00	.00	SPM6	7	.02	.04	.00	SPM10	7	.09	.00	.00
SPM3	0	.67	.04	.92	SPM7	0	.38	.20	.88	SPM11	0	.11	.26	.48
SPM3	1	.17	.00	.05	SPM7	1	.06	.12	.06	SPM11	1	.18	.43	.10
SPM3	2	.08	.40	.00	SPM7	2	.13	.28	.01	SPM11	2	.21	.00	.12
SPM3	3	.01	.32	.00	SPM7	3	.12	.28	.01	SPM11	3	.15	.12	.07
SPM3	4	.01	.04	.02	SPM7	4	.11	.00	.02	SPM11	4	.14	.00	.08
SPM3	5	.00	.12	.01	SPM7	5	.07	.00	.02	SPM11	5	.08	.19	.07
SPM3	6	.04	.04	.00	SPM7	6	.09	.00	.00	SPM11	6	.07	.00	.07
SPM3	7	.02	.04	.00	SPM7	7	.04	.12	.00	SPM11	7	.06	.00	.01
SPM4	0	.62	.00	.98	SPM8	0	.18	.00	.81	SPM12	0	.03	.24	.45
SPM4	1	.11	.40	.01	SPM8	1	.24	.04	.01	SPM12	1	.14	.09	.17
SPM4	2	.04	.36	.00	SPM8	2	.08	.00	.07	SPM12	2	.19	.44	.10
SPM4	3	.07	.08	.00	SPM8	3	.14	.08	.03	SPM12	3	.23	.03	.06
SPM4	4	.05	.00	.01	SPM8	4	.11	.20	.03	SPM12	4	.12	.00	.07
SPM4	5	.04	.12	.00	SPM8	5	.07	.28	.04	SPM12	5	.14	.00	.06
SPM4	6	.04	.00	.00	SPM8	6	.11	.40	.01	SPM12	6	.07	.20	.07
SPM4	7	.03	.04	.00	SPM8	7	.07	.00	.00	SPM12	7	.08	.00	.02

Note. Cat = category.

References

1. Myszkowski, N.; Storme, M. A snapshot of g. Binary and polytomous item-response theory investigations of the last series of the standard progressive matrices (SPM-LS). *Intelligence* **2018**, *68*, 109–116. doi:10.1016/j.intell.2018.03.010.
2. Forthmann, B.; Förster, N.; Schütze, B.; Hebbeker, K.; Flessner, J.; Peters, M.T.; Souvignier, E. How much g is in the distractor? Re-thinking item-analysis of multiple-choice items. *J. Intell.* **2020**, *8*, 11. doi:10.3390/jintelligence8010011.
3. Garcia-Garzon, E.; Abad, F.J.; Garrido, L.E. Searching for g: A new evaluation of SPM-LS dimensionality. *J. Intell.* **2019**, *7*, 14. doi:10.3390/jintelligence7030014.
4. Partchev, I. Diagnosing a 12-item dataset of Raven matrices: With Dexter. *J. Intell.* **2020**, *8*, 21. doi:10.3390/jintelligence8020021.
5. Myszkowski, N. A Mokken scale analysis of the last series of the standard progressive matrices (SPM-LS). *J. Intell.* **2020**, *8*, 22. doi:10.3390/jintelligence8020022.
6. Bürkner, P.C. Analysing standard progressive matrices (SPM-LS) with Bayesian item response models. *J. Intell.* **2020**, *8*, 5. doi:10.3390/jintelligence8010005.
7. von Davier, M.; Naemi, B.; Roberts, R.D. Factorial versus typological models: A comparison of methods for personality data. *Meas. Interdiscip. Res. Persp.* **2012**, *10*, 185–208. doi:10.1080/15366367.2012.732798.

8. Borsboom, D.; Rhemtulla, M.; Cramer, A.O.J.; van der Maas, H.L.J.; Scheffer, M.; Dolan, C.V. Kinds versus continua: A review of psychometric approaches to uncover the structure of psychiatric constructs. *Psychol. Med.* **2016**, *46*, 1567–1579. doi:10.1017/S0033291715001944.
9. Jacobucci, R.; Brandmaier, A.M.; Kievit, R.A. A practical guide to variable selection in structural equation modeling by using regularized multiple-indicators, multiple-causes models. *Adv. Methods Pract. Psychol. Sc.* **2019**, *2*, 55–76. doi:10.1177/2515245919826527.
10. Agresti, A.; Kateri, M. Some remarks on latent variable models in categorical data analysis. *Commun. Stat. – Theory Methods* **2014**, *43*, 801–814. doi:10.1080/03610926.2013.814783.
11. Collins, L.M.; Lanza, S.T. *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*; Wiley: New York, 2009. doi:10.1002/9780470567333.
12. Langeheine, R.; Rost, J. *Latent trait and latent class models*; Plenum Press: New York, 1988. doi:10.1007/978-1-4757-5644-9.
13. Lazarsfeld, P.F.; Henry, N.W. *Latent structure analysis*; Houghton Mifflin, 1968.
14. Formann, A.K. Linear logistic latent class analysis. *Biometrical J.* **1982**, *24*, 171–190. doi:10.1002/bimj.4710240209.
15. Janssen, A.B.; Geiser, C. On the relationship between solution strategies in two mental rotation tasks. *Learn. Individ. Differ.* **2010**, *20*, 473–478. doi:10.1016/j.lindif.2010.03.002.
16. Finch, W.H.; Bronk, K.C. Conducting confirmatory latent class analysis using Mplus. *Struct. Equ. Modeling* **2011**, *18*, 132–151. doi:10.1080/10705511.2011.532732.
17. Nussbeck, F.W.; Eid, M. Multimethod latent class analysis. *Front. Psychol.* **2015**, *6*, 1332. doi:10.3389/fpsyg.2015.01332.
18. Oberski, D.L.; Hagenaars, J.A.P.; Saris, W.E. The latent class multitrait-multimethod model. *Psychol. Methods* **2015**, *20*, 422–443. doi:10.1037/a0039783.
19. Schmiede, S.J.; Masyn, K.E.; Bryan, A.D. Confirmatory latent class analysis: Illustrations of empirically driven and theoretically driven model constraints. *Organ. Res. Methods* **2018**, *21*, 983–1001. doi:10.1177/1094428117747689.
20. von Davier, M.; Lee, Y.S., Eds. *Handbook of diagnostic classification models*; Springer: Cham, 2019. doi:10.1007/978-3-030-05584-4.
21. Formann, A.K.; Kohlmann, T. Structural latent class models. *Sociol. Methods Res.* **1998**, *26*, 530–565. doi:10.1177/0049124198026004005.
22. Formann, A.K. (Almost) equivalence between conditional and mixture maximum likelihood estimates for some models of the Rasch type. In *Multivariate and mixture distribution Rasch models*; von Davier, M.; Carstensen, C.H., Eds.; Springer: New York, 2007; pp. 177–189. doi:10.1007/978-0-387-49839-3_11.
23. von Davier, M. A general diagnostic model applied to language testing data. *Brit. J. Math. Stat. Psychol.* **2008**, *61*, 287–307. doi:10.1348/000711007X193957.
24. von Davier, M. Hierarchical mixtures of diagnostic models. *Psych. Test Assess. Model.* **2010**, *52*, 8–28.
25. Formann, A.K. Linear logistic latent class analysis for polytomous data. *J. Am. Stat. Assoc.* **1992**, *87*, 476–486.
26. Chen, Y.; Li, X.; Liu, J.; Ying, Z. Regularized latent class analysis with application in cognitive diagnosis. *Psychometrika* **2017**, *82*, 660–692. doi:10.1007/s11336-016-9545-6.
27. Hastie, T.; Tibshirani, R.; Wainwright, M. *Statistical learning with sparsity: The lasso and generalizations*; CRC Press: Boca Raton, 2015. doi:10.1201/b18401.
28. Wu, B. Sparse cluster analysis of large-scale discrete variables with application to single nucleotide polymorphism data. *J. Appl. Stat.* **2013**, *40*, 358–367. doi:10.1080/02664763.2012.743977.
29. Yamamoto, M.; Hayashi, K. Clustering of multivariate binary data with dimension reduction via L_1 -regularized likelihood maximization. *Pattern Recognit.* **2015**, *48*, 3959–3968. doi:10.1016/j.patcog.2015.05.026.
30. Huang, P.H.; Chen, H.; Weng, L.J. A penalized likelihood method for structural equation modeling. *Psychometrika* **2017**, *82*, 329–354. doi:10.1007/s11336-017-9566-9.
31. Jacobucci, R.; Grimm, K.J.; McArdle, J.J. Regularized structural equation modeling. *Struct. Equ. Modeling* **2016**, *23*, 555–566. doi:10.1080/10705511.2016.1154793.
32. Chen, Y.; Li, X.; Liu, J.; Ying, Z. Robust measurement via a fused latent and graphical item response theory model. *Psychometrika* **2018**, *83*, 538–562. doi:10.1007/s11336-018-9610-4.

33. Sun, J.; Chen, Y.; Liu, J.; Ying, Z.; Xin, T. Latent variable selection for multidimensional item response theory models via L_1 regularization. *Psychometrika* **2016**, *81*, 921–939. doi:10.1007/s11336-016-9529-6.
34. Scharf, F.; Nestler, S. Should regularization replace simple structure rotation in exploratory factor analysis? *Struct. Equ. Modeling* **2019**, *26*, 576–590. doi:10.1080/10705511.2018.1558060.
35. Bhattacharya, S.; McNicholas, P.D. A LASSO-penalized BIC for mixture model selection. *Adv. Data Anal. Classi.* **2014**, *8*, 45–61. doi:10.1007/s11634-013-0155-1.
36. Ruan, L.; Yuan, M.; Zou, H. Regularized parameter estimation in high-dimensional Gaussian mixture models. *Neural Comput.* **2011**, *23*, 1605–1622. doi:10.1162/NECO_a_00128.
37. DeSantis, S.M.; Houseman, E.A.; Coull, B.A.; Stemmer-Rachamimov, A.; Betensky, R.A. A penalized latent class model for ordinal data. *Biostatistics* **2008**, *9*, 249–262. doi:10.1093/biostatistics/kxm026.
38. DeSantis, S.M.; Andrés Houseman, E.; Coull, B.A.; Nutt, C.L.; Betensky, R.A. Supervised Bayesian latent class models for high-dimensional data. *Stat. Med.* **2012**, *31*, 1342–1360. doi:10.1002/sim.4448.
39. Houseman, E.A.; Coull, B.A.; Betensky, R.A. Feature-specific penalized latent class analysis for genomic data. *Biometrics* **2006**, *62*, 1062–1070. doi:10.1111/j.1541-0420.2006.00566.x.
40. Leoutsakos, J.S.; Bandeen-Roche, K.; Garrett-Mayer, E.; Zandi, P.P. Incorporating scientific knowledge into phenotype development: Penalized latent class regression. *Stat. Med.* **2011**, *30*, 784–798. doi:10.1002/sim.4137.
41. Sun, J.; Herazo-Maya, J.D.; Molyneaux, P.L.; Maher, T.M.; Kaminski, N.; Zhao, H. Regularized latent class model for joint analysis of high-dimensional longitudinal biomarkers and a time-to-event outcome. *Biometrics* **2019**, *75*, 69–77. doi:10.1111/biom.12964.
42. Wu, Zhenke, L.C.R.; Rosen, A.; Zeger, S.L. A Bayesian approach to restricted latent class models for scientifically-structured clustering of multivariate binary outcomes. *arXiv Preprint* **2018**, p. arXiv:1808.08326.
43. Fop, M.; Murphy, T.B. Variable selection methods for model-based clustering. *Stat. Surv.* **2018**, *12*, 18–65. doi:10.1214/18-SS119.
44. Fan, J.; Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* **2001**, *96*, 1348–1360. doi:10.1198/016214501753382273.
45. George, A.C.; Robitzsch, A.; Kiefer, T.; Groß, J.; Ünlü, A. The R package CDM for cognitive diagnosis models. *J. Stat. Softw.* **2016**, *74*, 1–24. doi:10.18637/jss.v074.i02.
46. Robitzsch, A.; George, A.C. The R package CDM for diagnostic modeling. In *Handbook of diagnostic classification models*; von Davier, M.; Lee, Y.S., Eds.; Springer: Cham, 2019; pp. 549–572. doi:10.1007/978-3-030-05584-4_26.
47. Tibshirani, R.; Saunders, M.; Rosset, S.; Zhu, J.; Knight, K. Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Series B Stat. Methodol.* **2005**, *67*, 91–108. doi:10.1111/j.1467-9868.2005.00490.x.
48. Tutz, G.; Gertheiss, J. Regularized regression for categorical data. *Stat. Model.* **2016**, *16*, 161–200. doi:10.1177/1471082X16642560.
49. Zhang, C.H. Nearly unbiased variable selection under minimax concave penalty. *Ann. Stat.* **2010**, *38*, 894–942.
50. Wang, C.; Lu, J. Learning attribute hierarchies from data: Two exploratory approaches. *J. Educ. Behav. Stat.* **2020**. Advance online publication, doi:10.3102/1076998620931094.
51. Storme, M.; Myszkowski, N.; Baron, S.; Bernard, D. Same test, better scores: Boosting the reliability of short online intelligence recruitment tests with nested logit item response theory models. *J. Intell.* **2019**, *7*, 17. doi:10.3390/jintelligence7030017.
52. Huang, J.; Breheny, P.; Ma, S. A selective review of group selection in high-dimensional models. *Stat. Sci.* **2012**, *27*, 481–499. doi:10.1214/12-STS392.
53. Cao, P.; Liu, X.; Liu, H.; Yang, J.; Zhao, D.; Huang, M.; Zaiane, O. Generalized fused group lasso regularized multi-task feature learning for predicting cognitive outcomes in Alzheimers disease. *Comput. Methods Programs Biomed.* **2018**, *162*, 19–45. doi:10.1016/j.cmpb.2018.04.028.
54. Liu, X.; Cao, P.; Wang, J.; Kong, J.; Zhao, D. Fused group lasso regularized multi-task feature learning and its application to the cognitive performance prediction of Alzheimer's disease. *Neuroinformatics* **2019**, *17*, 271–294. doi:10.1007/s12021-018-9398-5.
55. Oelker, M.R.; Tutz, G. A uniform framework for the combination of penalties in generalized structured models. *Adv. Data Anal. Classi.* **2017**, *11*, 97–120. doi:10.1007/s11634-015-0205-y.

56. Battauz, M. Regularized estimation of the nominal response model. *Multivar. Behav. Res.* **2019**. Advance online publication, doi:10.1080/00273171.2019.1681252.
57. Robitzsch, A. *sirt: Supplementary item response theory models*, 2020. R package version 3.9-4. <https://CRAN.R-project.org/package=sirt>.
58. Chen, Y.; Liu, J.; Xu, G.; Ying, Z. Statistical analysis of Q-matrix based diagnostic classification models. *J. Am. Stat. Assoc.* **2015**, *110*, 850–866. doi:10.1080/01621459.2014.934827.
59. Chen, Y.; Culpepper, S.; Liang, F. A sparse latent class model for cognitive diagnosis. *Psychometrika* **2020**, *85*, 121–153. doi:10.1007/s11336-019-09693-2.
60. Gu, Y.; Xu, G. Learning attribute patterns in high-dimensional structured latent attribute models. *J. Mach. Learn. Res.* **2019**, *20*, 115.
61. Liu, J.; Kang, H.A. Q-matrix learning via latent variable selection and identifiability. In *Handbook of diagnostic classification models*; von Davier, M.; Lee, Y.S., Eds.; Springer: Cham, 2019; pp. 247–263. doi:10.1007/978-3-030-05584-4_12.
62. Xu, G.; Shang, Z. Identifying latent structures in restricted latent class models. *J. Am. Stat. Assoc.* **2018**, *113*, 1284–1295. doi:10.1080/01621459.2017.1340889.
63. Kang, H.A.; Liu, J.; Ying, Z. A graphical diagnostic classification model. *arXiv Preprint* **2017**, p. arXiv:1707.06318.
64. Tamhane, A.C.; Qiu, D.; Ankenman, B.E. A parametric mixture model for clustering multivariate binary data. *Stat. Anal. Data Min.* **2010**, *3*, 3–19. doi:10.1002/sam.10063.
65. van Erp, S.; Oberski, D.L.; Mulder, J. Shrinkage priors for Bayesian penalized regression. *J. Math. Psychol.* **2019**, *89*, 31–50. doi:10.1016/j.jmp.2018.12.004.